Molecular Phylogenetics and Historical Biogeography of Basal Angiosperms

– A Case Study in Nymphaeales –

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch–Naturwissenschaftlichen Fakultät

der Rheinischen Friedrich–Wilhelms–Universität Bonn

vorgelegt von

Cornelia Löhne

aus

Leipzig

Bonn, März 2006

Angefertigt mit Genehmigung der Mathematisch–Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhems-Universität Bonn.

Erstgutachter: Prof. Dr. Wilhelm Barthlott
Zweitgutachter: PD Dr. Thomas Borsch

Tag der mündlichen Prüfung: 30.5.2006

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn
http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

Erscheinungsjahr 2007

Preface and Acknowledgements

The present thesis was carried out in the framework of the Graduiertenkolleg "Evolution and Biodiversity in Time and Space" and supported by a graduate student fellowship from the Deutsche Forschungsgemeinschaft (DFG–Graduiertenkolleg 721), and supervised by Prof. Wilhelm Barthlott and Dr. habil. Thomas Borsch.

The Graduiertenkolleg "Evolution and Biodiversity in Time and Space" united researchers and graduate students from the fields of organismic zoology, botany, developmental biology as well as palaeontology. In the course of several seminars and workshops all participants could exchange their specific knowledge and experience in fruitful discussions, broaden their horizons and get a wider picture on the diversity of organisms and their evolution. This innovative approach was extremely inspiring and provided a solid basis and backbone for my own interdisciplinary study.

Many persons have been of great importance for this PhD project. First of all, I would like to express my gratitude to Prof. Wilhelm Barthlott. He gave me the chance to step into this integrative field of botany by inviting me to start this PhD project under his supervision in 2002. His stimulating fascination for the diversity of plants, together with his encouragement and support throughout the years were of great help. Also, I am indebted to Dr. Thomas Borsch, who introduced me into the world of phylogenetic systematics and molecular evolution. Without his continuous support and encouragement the present project would not have been possible. His valuable suggestions and comments on all parts of this thesis as well as his experience with the water lilies are greatly appreciated.

Furthermore, I would like to express my gratitude to John Wiersema (US Department of Agriculture, Beltsville / USA), who was of great help for large parts of this thesis. My sincere thanks are for letting me benefit from his outstanding knowledge on the biology and diversity of water lilies. I am grateful to Prof. Klaus Peter Sauer and Prof. Thomas Litt (both University of Bonn), as well as to all other members of the Graduiertenkolleg for sharing their experience and for many stimulating discussions. Many thanks are due to Prof. Khidir W. Hilu (Virginia Tech, Blacksburg / USA) for valuable suggestions on the outline of this project.

The molecular data sets in this thesis were produced in the Molecular Systematics Lab at the Nees Institute for Biodiversity of Plants in Bonn, which is led by Dr. Thomas Borsch. I wish to express my warmest thanks to all my colleagues: Dr. Kai Müller, Andreas Worberg, Dr. Dietmar Quandt, Dr. Rolf Blöcher, Kim Govers and Kassahun Tesfaye – thank you for all the assistance in the lab and at the PC, for interesting discussions and for much sympathy. The comments and suggestions of Kai Müller and Dietmar Quandt with regard to molecular

evolution were especially helpful for the first part of my thesis. The respective chapter (2) profited also a lot from discussions with Prof. Volker Knoop (University of Bonn), Dr. Scot Kelchner (Idaho State University, Pocatello / USA) and Dr. Helga Ochoterena Booth (Universidad Nacional Autónoma de México).

With regard to chapters 3 and 4, I am indebted to Dr. John Wiersema, Dr. Barre Hellquist (Massachusetts College of Liberal Arts, North Adams / USA) and Dr. Surrey Jacobs (Royal Botanical Gardens Sydney / Australia) for providing DNA material of several specimens of *Nymphaea* and *Ondinea* and for extremely useful comments on the results of the phylogenetic analyses. At this point I would also like to express my thanks to the Botanic Gardens of Bonn University, especially to Dr. Wolfram Lobin and to Michael Neumann, for curating and cultivating the living collection of water lilies, which was mainly established through collections of Dr. Thomas Borsch.

The most integrative section of the present thesis is the last chapter, drawing a bow from phylogeny and extant distribution to the fossil record and biogeographic history. Therefore, I am indebted to many people who contributed to the completion of this part of the study. First, I would like to mention the BIOMAPS group of the Nees Institute, namely Dr. Jens Mutke, Holger Kreft, Dr. Wolfgang Küper and Henning Sommer. I could benefit a lot from their practical know-how in geographic information systems and from the long–standing experience and theoretical background in biodiversity mapping in this working group. I am also very grateful to PD Dr. Volker Wilde (Forschungsinstitut Senckenberg, Frankfurt / Germany), who shared his knowledge on earth history and the fossil record with me and provided valuable suggestions for the interpretation of the data. This last chapter of my thesis is an outcome of a close cooperation with Mi-Jeong Yoo (Soltis Lab, Florida Museum of Natural History / USA). Mi-Jeong Yoo has conducted most of the work on molecular dating of the Nymphaeales data set. I am very grateful to her for all the help, for the stimulating discussions and the introduction into cloning methods – but also for the good time I had during my stay at the Soltis Lab in Gainesville. I also sincerely thank Dr. Pamela Soltis and Prof. Douglas Soltis for their valuable suggestions, the encouraging support and their warm hospitality. Many thanks to Nils Köster, Holger Kreft, Nadja Korotkova and Andreas Worberg (all University of Bonn) for proof–reading.

Last but not least I would like to express my deep gratitude to Dr. Tom Motzer for his untiring personal encouragement and loving support. Thanks for being there and for standing behind me. I very much enjoyed the time with you.

Contents

Chapter 1

General Introduction

Contents

1.1 Why work on basal angiosperms?

The angiosperms are the largest, morphologically and ecologically most diverse group of land plants. Angiosperms are still an "abominable mystery" (Darwin 1879) with respect to the circumstances of their origin and their rapid diversification in the Early Cretaceous. Today they occur with over 260,000 species[1] and a great variety of life forms in almost all habitats on earth, except the highest mountaintops, the polar regions and the deep sea; and mostly they are dominating the habitats. Over two thirds of angiosperms (>200,000 species) belong to the "eudicots", a monophyletic group of taxa that are characterized by tricolpate pollen or derivations of that. The "monocotyledons" (or short "monocots"), a natural group of taxa share several specific traits such as the presence of only one cotyledon, constitute another considerable portion of the angiosperms (>50,000 species). The rest (~10,000 species) is a highly diverse, paraphyletic group of taxa, that constitute the basal branches of the angiosperm tree. These taxa are generally referred to as the "basal angiosperms" although this is not a formal taxonomic rank (see Chapter 1.2.3). In these basal angiosperms almost all life forms are represented: from trees and lianas, over herbaceous plants and epiphytes to submerged aquatics.

The magnoliids are the largest and most diversified clade among extant basal angiosperms, consisting of the four orders Magnoliales (2850 species in 155 genera and 6 families), Laurales (3000 species in 95 genera and 8 families), Piperales (3450 species in 22 genera and 4 families) and Canellales (81 species in 10 genera and 2 families). All other clades of basal angiosperms are considerably smaller. There are the exclusively woody members of the order Austrobaileyales (100 species in 5 genera and 4 families), the herbaceous to shrubby Chloranthales (75 species in 4 genera in a single family), and the aquatic order Nymphaeales (70 species in 2 families and 8 genera). *Ceratophyllum* (6 species), another aquatic member of basal angiosperms is rather isolated from other lineages, therefore is has been treated as a separate order (Ceratophyllales). Similarly the monotypic, New Caledonian genus *Amborella*, which does not show closer morphological affinity to any other angiosperm lineage, constitutes its own order Amborellales. Actually, the scandent understorey shrub *Amborella trichopoda* has lately been inferred to be the sister to all other angiosperms, thus being the "most basal angiosperm" (see also page 9 ff.)

The most attractive and best–known representatives of basal angiosperms are undoubtedly the magnolias (Magnoliaceae) and water lilies (Nymphaeaceae), both with large, showy flowers, for which they are commonly grown as ornamentals. Some less important

[1] The total number of species is estimated from the number of accepted species names. Thereby the percentage of synonymous names existing in the literature has to be taken into account. Due to different ways of approximating synonymy rates, recent estimates of the number of species in seed plants vary considerably from ~220,000 (Scotland & Wortley 2003), over 260,000 (Thorne 2002) and 300,000-320,000 (Prance *et al.* 2000) to ~420,000 (Govaerts 2001, Bramwell 2002). Currently, Thorne's estimation of 260,000 of species number seems to be most reasonable based on current data. It is widely accepted, as evident from several recent treatments of angiosperms (e.g., Judd *et al.* 2002, Angiosperm Phylogeny Group (APG II) 2003, Soltis & Soltis 2004b, Bell *et al.* 2005) and will be adopted here, too. However, also the approach of Prance *et al.* (2000) deserves consideration since his estimation also regards the number of species that are not yet described (i.e. that don't have names yet). Generally, the number of seed plants can be equated with the number of angiosperms, since the percentage of non-angiosperm seed plants, i.e. gymnosperms, is neglectably small (~ 1000, Thorne 2002).

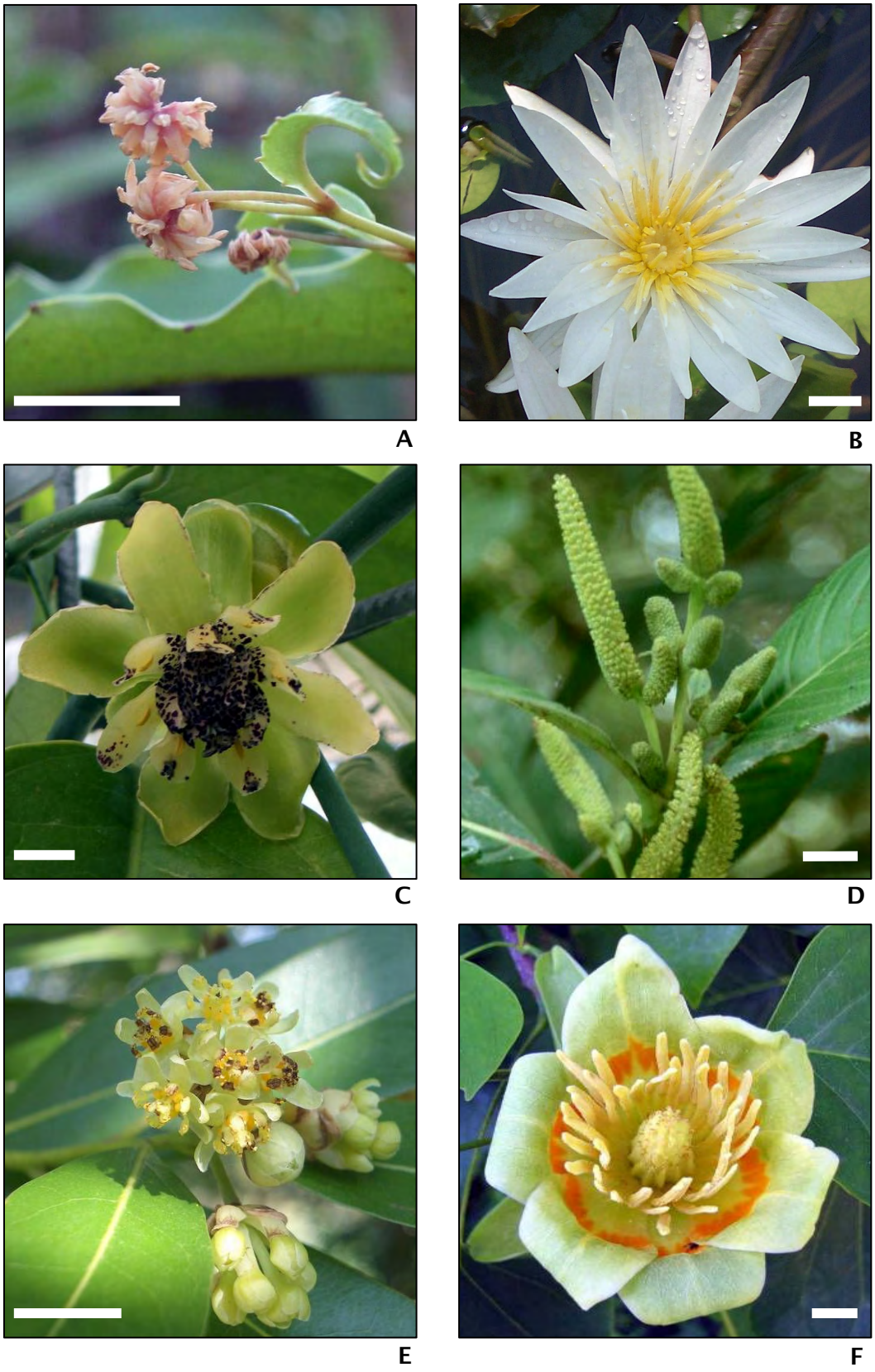


Figure 1.1 — Floral diversity in basal angiosperms. A) *Amborella trichopoda* (Amborellaceae), **B)** *Nymphaea micrantha* (Nymphaeaceae), **C)** *Austrobaileya scandens* (Austrobaileyaceae), **D)** *Hedyosmum arborescens* (Chloranthaceae), **E)** *Umbellularia californica* (Lauraceae), **F)** *Liriodendron tulipifera* (Magnoliaceae). The scale bar indicates 1 cm. Photos A-C, E: C. Löhne; D: N. Köster (with permission); F: N. Korotkova (with permission).

ornamental plants can be found in other lineages such as Austrobaileyales (e.g. *Illicium*, *Schisandra*), Piperales (e.g.*Aristolochia*), or Magnoliales (e.g. *Cananga*, Annonaceae). A few representatives of the magnoliids are cultivated because of their edible fruits, e.g. avocado (*Persea americana*, Lauraceae), cherimoya (*Annona cherimola*, Annonaceae) or the custard apple (*Annona squamosa*, Annonaceae). The economic importance of basal angiosperms is founded mainly on the diverse spectrum of secondary compounds – above all essential oils – that are characteristic for the majority of families. Several members of basal angiosperms provide exquisite spices, e.g. star anise (*Illicium verum*, Illiciaceae), cinnamon (*Cinnamomum verum*, Lauraceae), laurel (*Laurus nobilis*, Lauraceae), nutmeg (*Myristica fragrans*, Myristicaeae) and black pepper (*Piper nigrum*, Piperaceae), or fragrances, e.g. camphor (*Cinnamomum camphora*, Lauraceae) and ylang-ylang (*Cananga odorata*, Annonaceae). Typically for basal angiosperms, the essential oils are produced in excretory cells (Frohne & Jensen 1998). Because of the essential oils and other secondary compounds such as benzylisoquinoline–alkaloids, many species play important roles in traditional medicine, e.g. *Schisandra chinensis* in Chinese traditional medicine, but also *Kadsura* (Schisandraceae), *Illicium* (Illiciaceae), several water lilies (*Nymphaea*, Nymphaeaceae), several members of Chloranthaceae, *Sassafras* (Lauraceae), and many others (Vogel 1970, Emboden 1978, 1979, Tang & Eisenbrand 1992).

Besides their relevance in applied fields of botany, basal angiosperms are crucial for understanding the history of flowering plants and for analysing the crucial factors in plant evolution. Studying basal angiosperms provides insights into early evolution of morphological traits in flowering plants and allows the identification of typical evolutionary trends. Basal angiosperms have therefore been studied in several works on morphology and development of flowers (e.g., Endress 2001, 2004, Friedman & Williams 2004, Smyth 2005) and embryos (Forbis *et al.* 2002). Other researchers have tried to draw conclusions from extant basal angiosperms about the life form and ecology of the earliest flowering plants (e.g., Thien *et al.* 2000, Feild *et al.* 2003, Feild *et al.* 2004, Kim *et al.* 2004a, Feild & Arens 2005, but see Crisp & Cook 2005). Finding the root of angiosperms has been the focus of several studies because it provides a direction and temporal scale for plant evolution, not only on the morphological but also on the molecular level. A profound idea of the phylogenetic evolution of plants allows the establishment of explicit hypotheses of how molecular traits such as genome size (Soltis *et al.* 2003, Zonneveld *et al.* 2005) and genome structure (Kim *et al.* 2004b, De Bodt *et al.* 2005) have changed during the past 125 million years (see also Savolainen & Chase 2003). Furthermore, recent studies have shown that a detailed picture about the early radiation of flowering plants is crucial for understanding the evolution of other groups of plants, such as ferns (Schneider *et al.* 2004), or animals, e.g. insects (Berenbaum 1983, Labandeira *et al.* 1994, Farrell 1998, Labandeira 2002, Percy *et al.* 2004).

1.2 Hypotheses on the evolution and phylogeny of basal angiosperms

1.2.1 The origin and early diversification of angiosperms

Today angiosperms are the dominating plant group on earth, but compared to other seed plants they represent a relatively young radiation as evident from their sudden and explosive appearance in the fossil record of the Early Cretaceous (see Crane *et al.* 1995 for review). Darwin's "abominable mystery" of the origin of angiosperms has triggered innumerable studies, hypotheses and speculations, but still remains one the most persistent puzzles in understanding the evolution of plants. Clarifying the circumstances of the origin of angiosperms goes hand in hand with the search for their closest relatives among other extant or fossil seed plants. However, the fossil record does not provide satisfying answers to this question, since no fossils with intermediate characters between angiosperms and other seed plants that would help to elucidate the phylogenetic history. On the contrary, the fossil record reveals a huge diversity of different plant lineages that are not closely related to any other extant lineage und rather underscores the uniqueness of angiosperms. Nevertheless, the closest relatives of angiosperms were sought, e.g., in Mesozoic seed ferns such as *Caytonia* (see, e.g., Judd *et al.* 2002 for overview).

Until the middle of the 20th century Bennettitales or Gnetales, both sometimes united with angiosperms as "anthophyta" because of an aggregation of sporophylls and perianth–like structures resembling angiosperm flowers, were believed to be the closest relatives of flowering plants (e.g., Arber & Parkin 1907). Inspired by the idea of a close relationship of angiosperms and Gnetales, the development of the angiosperm flower was explained with a transformation of male and female flowers of Gnetales along with their subtending leaves into the floral organs of the angiosperm flower, similar to the organisation of pseudanthia in *Euphorbia* (***"Pseudanthial theory"***, Wettstein 1924, Zimmermann 1930). However, interpreting the angiosperm flower as an inflorescence has met with much opposition from botanists. Instead, it was widely believed that the angiosperm flower was derived from a large compound bisexual strobiloid flowerlike organ with enrolled megasporophylls and stamina and perianth organs similar to those of Bennettitales. This interpretation is based on the ***"Euanthial theory"*** of Arber & Parkin (1907). There has been a lot of debate about these opposing theories in the early 20th century, but over time the Euanthial theory gained wider acceptance. However, neither of the two theories could be proved due to the lack of any intermediate states in the fossil record. A newly discovered early angiosperm fossil from the Lower Cretaceous, *Archaefructus* (Sun *et al.* 1998, Sun *et al.* 2002, see also Chapter 1.3.1), has refreshed this debate, since its carpels and stamens, which are loosely arranged on a prolonged axis, were interpreted either as an inflorescence (Friis *et al.* 2003) or as a single flower (Sun *et al.* 2002).

With the introduction of cladistic methods into evolutionary research from the mid 80s on, new evidence could be provided to substantiate the close relationship of Gnetales and angiosperms. In fact, all cladistic analyses of morphological characters support this so–called ***"anthophyte hypothesis"*** (Crane 1985, Doyle & Donoghue 1986, 1992, Nixon *et al.* 1994, Doyle 1996, 1998). However, Donoghue & Doyle (2000) conceded that many of the morphological features that were thought to support the anthophyte hypothesis differ in

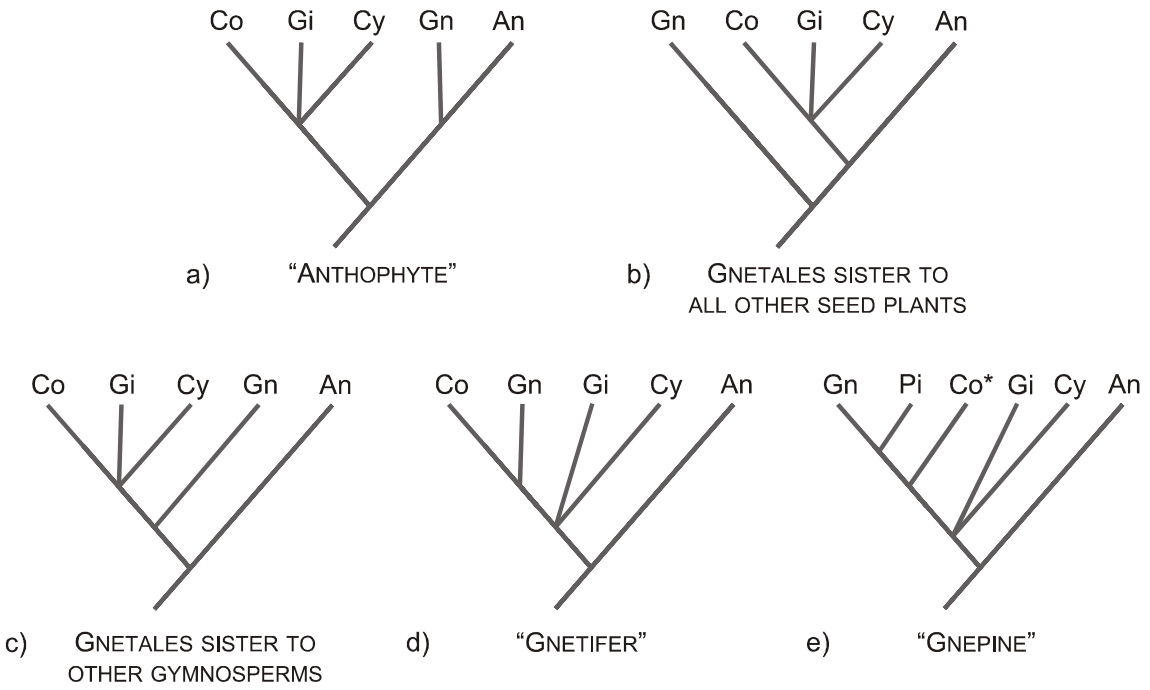


Figure 1.2 — Alternative hypotheses on seed plant relationships, modified from Burleigh & Mathews (2004). The anthophyte hypothesis (a) is mainly based on morphological characters, where as hypotheses (b) to (e) are inferred from molecular data. Currently, hypotheses implying a monophyly of extant gymnosperms (c–e) are favoured. See text for references and further explanation. An = angiosperms, Co = conifers, Co* = all conifers except Pinaceae, Cy = Cycadales, Gi = Ginkgoales, Gn = Gnetales, Pi = Pinaceae.

details, such as the tunica, which consists of two cell layers in angiosperms and one in Gnetales, or the megaspore wall, which is thin in Gnetales and absent in angiosperms. The "demise of the anthophyte hypothesis" (Donoghue & Doyle 2000) was initiated by the first results of DNA analyses that placed Gnetales sister to all other seed plants (e.g., Hamby & Zimmer 1992, see Figure 1.2, b) or the rest of gymnosperms (Hasebe *et al.* 1992, see Figure 1.2, c, Goremykin *et al.* 1996). The latter topology (Figure 1.2, c) has recently been found again from chloroplast (Samigullin *et al.* 1999) and several nuclear markers (Winter *et al.* 1999, Frohlich & Parker 2000, Schmidt & Schneider-Poetsch 2002, Becker & Theissen 2003). Topologies consistent with the anthophyte hypotheses could only be retrieved from partial sequences of 28S rDNA by Stefanovic *et al.* (1998) and in analyses of some data subsets by Rydin *et al.* (2002), but in both cases without any statistical support. Evidence for Gnetales being sister to all other seed plants (Figure 1.2, b) was also provided from a multi–locus chloroplast dataset of Rai *et al.* (2003) and a two–genome dataset of Rydin *et al.* (2002), and mor recently from the chloroplast *trnL* intron (Quandt *et al.* 2004). Bowe et al. (2000) argued that the Gnetales–sister topology (Figure 1.2, b) might be an artefact of long branch attraction between Gnetales and strongly divergent outgroup taxa. In fact, most current studies point towards the monophyly of extant gymnosperms by placing Gnetales either as sister to conifers (Figure 1.2, d, ***"Gnetifer–hypothesis"***, e.g. Chaw *et al.* 1997, Hansen *et al.* 1999, Nickerson & Drouin 2004) or by placing them even within extant conifers, sister to the Pinaceae (Figure 1.2, e, ***"Gnepine–hypothesis"***, e.g. Qiu *et al.* 1999, Bowe *et al.* 2000, Chaw *et al.* 2000, Nickrent *et al.* 2000, Gugerli *et al.* 2001, Soltis *et al.* 2002a, Burleigh & Mathews 2004). However, a closer examination of the underlying data ets reveals some conflict: depending on analysis methods (parsimony vs. Bayesian), data

partitions (first and second vs. third codon partitions), or genomes (nuclear, plastid or mitochondrial) either Gnetales–sister or Gnepine trees were recovered (e.g., Sanderson *et al.* 2000, Magallón & Sanderson 2002, Rydin *et al.* 2002, Soltis *et al.* 2002a, Burleigh & Mathews 2004). Rydin & Källersjö (2002) stress the influence of taxon sampling on the inferred topology. The main cause for the persisting difficulties and conflicts probably lies in the very short branches at the base of the seed plant tree (Sanderson *et al.* 2000), which are thought to be a result of an early rapid radiation into the five extant lineages of seed plants. The question about the closest relative of flowering plants cannot be answered yet, but discovery of new fossils and careful evaluation of existing data may eventually shed light on the ancestors of angiosperm (Crepet 2000).

The sudden appearance and the concomitant rapid diversification of angiosperms in the Early Cretaceous are well documented in the fossil record (see Chapter 1.3.1 on page 12). Due to the diversity of fossil forms it is not possible to draw unequivocal conclusions on the appearance and life form of the first flowering plants. Thus, the hypotheses on the habit of early flowering plants have been diverse, ranging from large tropical rain forest trees and weedy drought–adapted colonizing shrubs to disturbance–tolerant and light–demanding herbs or, more recently, to aquatic herbs. One of the earliest and most influential ideas about early angiosperms is the ***"woody magnoliid hypothesis"*** (e.g., Takhtajan 1980, Cronquist 1988, Donoghue & Doyle 1989). According to this hypothesis the ancestral angiosperm was a woody shrub or tree with large flowers with numerous free perianth organs, stamens, and carpels – features that are characteristic of living Magnoliaceae and Winteraceae. The "woody magnoliids" inhabited tropical lowland rainforests or cloud forests and their seedlings established in undisturbed soil in wet, low-light environments under a closed canopy. This hypothesis was supported by magnolia–like fossils, such as *Archaeanthus* (Dilcher & Crane 1984) from the Cenomanian of North America.

An alternative hypothesis suggested that seasonal drought could be the major catalyst for the development of angiospermy and the extensive diversification (e.g., Axelrod 1972, Stebbins 1974). Accordingly, the earliest angiosperms were described as fast–growing, drought–adapted shrubs from disturbed habitats (***"weedy xeric shrub hypothesis"***, Feild & Arens 2005), and rainforest habitats were thought to be secondarily invaded by angiosperms. Stebbins (1974) correlates the invasion of new environments with polyploidisation events and points out that modern "woody magnoliids" such as Magnoliaceae and Winteraceae consist mainly or entirely of polyploid genera. However, the fossil record does not provide any evidence for or against this hypothesis, since angiosperm pollen was equally diverse and abundant in wet environments as well as in more arid areas (Feild & Arens 2005).

Another idea, related to the "weedy xeric shrub hypothesis", emphasized disturbance as the driving force for early angiosperm diversification. But, opposite to the before–mentioned hypothesis, the first angiosperms were thought to be rhizomatous herbs with rapid life cycles and high photosynthetic rates inhabiting sunny, unstable stream–sides and flood plains (Feild & Arens 2005). This so–called ***"paleoherb hypothesis"*** was introduced and strongly promoted by Taylor & Hickey (1990, 1992). Similar ideas were brought up earlier by Burger (1981) and Meeuse (1987). The term "paleoherb" was introduced by Doyle & Hickey (1976) and referred to the herbaceous representatives of the former class Magnoliidae, i.e. Nymphaeales, Piperaceae, Saururaceae, Aristolochiaceae, Ceratophyllaceae, and members

of Chloranthaceae. Therefore, the earliest angiosperm was searched among these groups. The "paleoherb hypothesis" gained support from some cladistic analyses of morphological characters that showed Nymphaeales to be basal in angiosperms (Doyle & Donoghue 1986), as well as from the first large–scale molecular study based on *rbcL* (Chase *et al.* 1993), where Ceratophyllum was inferred to be basal. The hypothesis was extended to include aquatic herbs after the discovery of a new angiosperm fossil from China, *Archaefructus* (Sun *et al.* 1998, Sun *et al.* 2002, Ji *et al.* 2004, see also Chapter 1.3.1). However, Feild & Arens (2005) provide a variety of reasons from ecology and developmental biology that would reject the idea of an aquatic origin of angiosperms. Instead, they suggest that the angiosperms first evolved inconspicuously under a "dark and disturbed" environment (see also Feild *et al.* 2004) and than diversified explosively parallel to the invasion of sunnier and dryer habitats.

1.2.2 Pre–cladistic hypotheses on angiosperm diversification

Prior to the advent of cladistic methods, biologists tried to group organisms following characteristics of their external appearance, such as life forms (trees vs. herbs), reproductive organs (cryptogams vs. phanerogams) or the complexity of flowers (e.g. lacking, simple, chori– or sympetalous perianth). Plants were arranged in "morphological series" by differentially weighting morphological characters from primitive to derived. The early systems of plants differed remarkably depending on the author because they were generally based on only few characters and those characters were given different weights by different authors (see e.g. Lawrence 1951, or Judd *et al.* 2002 page 41–53 for overview). John Ray (1682, 1696), for instance, was the first who clearly described the distinct features of monocotyledons, although the division into plants with one or two cotyledons was subordinate to his primary separation between trees and herbs. Antoine Laurent de Jussieu (1748–1836) subdivided phanerogams into monocotyledons and dicotyledons, but placed the gymnosperms at the end of his series of dicotyledon classes, probably because some genera have many cotyledons. The unique characters of gymnosperms (only conifers and cycads by that time) were shown by Robert Brown already in 1826 , but botanists did not adopt this distinction until the end of the 19th century. Instead, in his famous classification from 1813 Augustin Pyramus de Candolle (1778–1841) constructed series of classes, according to the complexity of flowers, in which monocots followed dicots, and the gymnosperms (not named as such) were placed somehow intermediate between monocots and dicots.

A clear distinction between gymnosperms and angiosperms eventually became accepted with the system of plant families of Adolf Engler (1844–1930). His ideas about the evolution of plants (e.g., Engler 1886, 1892, 1897) were most influential on the development of systematic botany at the beginning of the 20th century, and his system was the most important one in use for decades. Engler separated gymnosperms from angiosperms and monocots from dicots; and he arranged all plants within these groups along what he called "morphological progressions" following the general assumption that complex structures evolve from simple ones. According to his assumption, plants with small, perianthless flowers were at the base of monocots and dicots. For the development of systematic botany in North America the system of Charles Bessey (1845–1915) was most influential. Similarly to Engler he described specific evolutionary trends in flowering plants as a basis for the

reconstruction of phylogenies (Bessey 1915). The main difference to Engler's system is that according to Bessey small, apetalous and wind–pollinated flowers are derived from primitive perianth–bearing polymerous flowers.

In the course of the 20th century many different plant systems have been established (see Judd *et al*. 2002), but the most important of those are basically variants of Engler's and Bessey's ideas: Cronquist (1981, 1988), Takhtajan (1980, 1997), Dahlgren (1979-1980, 1983, 1985) and Thorne (1976, 1992). However, all the major classification systems of the 20th century had in common the division of angiosperms into the two classes Monocotyledonae and Dicotyledonae. Monocotyledons (or short "monocots") were characterized by their embryos with a single cotyledon, adventitious root systems, parallel leaf venation (with a few exceptions), trimerous flowers, scattered vascular bundles in the stems, sieve-tube plastids with several cuneate protein crystals (Behnke 1969), and the lack of vascular–cambium–produced secondary phloem and secondary xylem. Dicotyledons (or short "dicots") on the other hand were characterized by the presence of two cotyledons, a main root system developing from the radicule, reticulate leaf venation forming a pinnate to palmate pattern, 4– to 5–merous flowers, ring–like arrangement of vascular bundles in the stem and secondary growth.

1.2.3 Cladistic analyses and the changing view on angiosperm relationships

An important innovation in biological thinking in the 20th century was the introduction of phylogenetic systematics or cladistics by Willi Hennig (1950, 1965, 1966). The main principle of cladistics is that all observed morphological or other biological features of organisms have to be translated into discrete characters that can have alternative manifestations (= character states, e.g. present or absent, small or large, and so on). Cladistic methods aim to reconstruct phylogenies by establishing hierarchical sister–group relationships based on shared derived character states (= synapomorphies, Kitching *et al*. 1998). Hennig suggested that only monophyletic groups or "clades", i.e. groups that comprise all descendants of a common ancestor, should be regarded as valid units for classification whereas paraphyletic or polyphyletic groupings[2] should be avoided (Hennig 1966, 1974). The first cladistic analyses of angiosperms based on morphological characters (and later also on molecular data) revealed the paraphyletic nature of the dicots (e.g., Donoghue & Doyle 1989). Instead, features such as the presence of two cotyledons or reticulate leaf venation were inferred to be ancestral character states for the angiosperms as a whole. The monophyly of monocots, however, was clearly demonstrated by the first cladistic analyses and got further substantiated by molecular data (e.g., Chase *et al*. 1993, Duvall *et al*. 1993, see Chase 2004 for review). Donoghue & Doyle (1989) were the first who recognized that a large number of taxa of the paraphyletic dicots constitute a well–supported clade that is characterized by the presence of tricolpate pollen and modifications thereof. Doyle & Hotton (1991) introduced the name ***"eudicots"*** for this clade, which has later

[2] *Monophyletic group* (=clade): includes a most recent common ancestor plus all of its descendants and no other lineages (e.g. monocots or eudicots). *Paraphyletic group*: includes a most recent common ancestor and only some (not all) of its decendants, i.e. a group that remains of a monophyletic group if one or more components are excluded (e.g. basal angiosperms). *Polyphyletic group*: does not include the most recent common ancestor of all its members (e.g. Dilleniidae). (Definitions based on Kitching *et al*. 1998).

become more popular in the scientific community than the name ***"tricolpates"*** as suggested by Donoghue & Doyle (1989, but see Judd & Olmstead 2004). Those angiosperms, which are neither monocots nor eudicots, are often referred to as the ***"basal angiosperms"***. However, this term does not imply a monophyletic origin of the basal angiosperms but indicates that at least a part of that group constitutes the base of the angiosperm tree, i.e. representing sister lineages to all other angiosperms. Although the term "basal angiosperms" refers to a paraphyletic group and, therefore, is not valid as a formal taxonomic rank, it is widely used in the scientific community today.

Together with the general acceptance of the idea that monocots and eudicots originated from a stock of basal angiosperms, the search for the "most basal angiosperm" began. The first large-scale molecular study based on the *rbcL* gene (Chase *et al.* 1993) depicted the aquatic herb *Ceratophyllum* as the sister lineage to all other angiosperms. Besides the fact that no statistical tests were run to evaluate this analysis (not possible due to computational problems), the "*Ceratophyllum* basal" hypothesis was widely discussed and triggered other hypothesis on the origin of angiosperms, such as the "paleoherb theory" (Taylor & Hickey 1996). However, later Savolainen *et al.* (2000a) could show that *rbcL* does not provide any statistical support for the inferred position of *Ceratophyllum*. The picture started to change with presentations of several research groups at the 16th International Botanical Congress, taking place in St. Louis / Missouri in 1999. Surprisingly, the enigmatic New Caledonian understorey shrub *Amborella trichopoda* has been depicted as the sister to all other angiosperms in several independent studies based on molecular evidence from all three plant genomes (Parkinson *et al.* 1999, Qiu *et al.* 1999, Soltis *et al.* 1999a, Soltis *et al.* 1999b).

1.2.4 Current understanding of phylogenetic relationships

Phylogenetic relationships among basal angiosperm lineages are far from being resolved. Instead, a large polytomy characterizes the current consensus tree of basal angiosperms (Figure 1.3). However, due to the accumulation of data during the last decade, some relationships could be elucidated: It has become evident that three lineages of the basal angiosperms form a basal grade: *Amborella*, Nymphaeales, Austrobaileyales, which are sometimes referred to as the "ANITA"-grade (*Amborella*, Nymphaeales, Illiciaceae, Trimeniaceae, Austrobaileyaceae, Qiu *et al.* 2000). In most phylogenetic analysis of basal angiosperms *Amborella* is inferred to branch first, thus being sister to all other angiosperms (e.g., Parkinson *et al.* 1999, Mathews & Donoghue 2000, Soltis *et al.* 2000, Borsch *et al.* 2003, Hilu *et al.* 2003). However, Barkman *et al.* (2000) found an alternative topology where *Amborella* and Nymphaeales together form the basal branch of angiosperms. Although most studies point on the first hypothesis, i.e. *Amborella* basal, the alternative hypothesis could not be rejected in most recent analyses employing extensive character or taxon sampling (e.g., Zanis *et al.* 2002, Stefanovic *et al.* 2004, Leebens-Mack *et al.* 2005, Qiu *et al.* 2005). A recently proposed hypothesis by Goremykin *et al.* (2003), saying that not *Amborella* but the monocots are sister to the rest of angiosperms, could be rejected by several follow–up studies (Soltis & Soltis 2004a, Stefanovic *et al.* 2004, Leebens-Mack *et al.* 2005). Instead the "monocots basal" hypothesis was shown to be an artefact of long branch attraction in combination with limited taxon sampling.

More and more information on the relationships at the base of the angiosperm tree is accumulating from analyses of molecular data and careful re–examination of morphological characters. Kim *et al.* (2004b), for instance, could provide evidence for the monophyly of all angiosperms above Nymphaeales and *Amborella* from MADS–box genes: a 12 bp deletion in exon 5 of *PI*–homologues. The basal grade, including *Amborella*, Nymphaeales and Austrobaileyales is also characterized by some symplesiomorphic morphological characters, such as ascidiate carpels that are sealed by secretion (in contrast to plicate carpels, sealed by postgenital fusion of epidermal layers in higher angiosperms, see Endress & Igersheim 2000). The four orders Laurales, Magnoliales, Canellales and Piperales have been inferred to constitute a monophyletic group in several (Qiu *et al.* 1999, 2000, Zanis *et al.* 2002, Hilu *et al.* 2003, Zanis *et al.* 2003) but not all (Doyle & Endress 2000, Savolainen *et al.* 2000b, Soltis *et al.* 2000) phylogenetic studies. A very recent study employing 9 genes from all three genomes has added further evidence to themonophyly of this clade, called the magnoliids (Qiu *et al.* 2005, see also Figure 1.3). The *petD* intron — a new marker for phylogenetic inference at higher taxonomic levels (described in this thesis, Chapter 2) — substantiated the magnoliid clade by a synapomorphic length mutation. Furthermore, there is converging evidence, that within the magnoliids Piperales are sister to Canellales and Laurales are sister to Magnoliales (APG II 2003, Zanis *et al.* 2003, Borsch *et al.* 2005, Qiu *et al.* 2005). Besides the existence of a basal grade and the proposed monophyly of magnoliids, there are still several unsettled issues concerning the topology at the backbone of the tree of angiosperms. Resolving the relationships among Chloranthales, Ceratophyllales, magnoliids, eudicots and

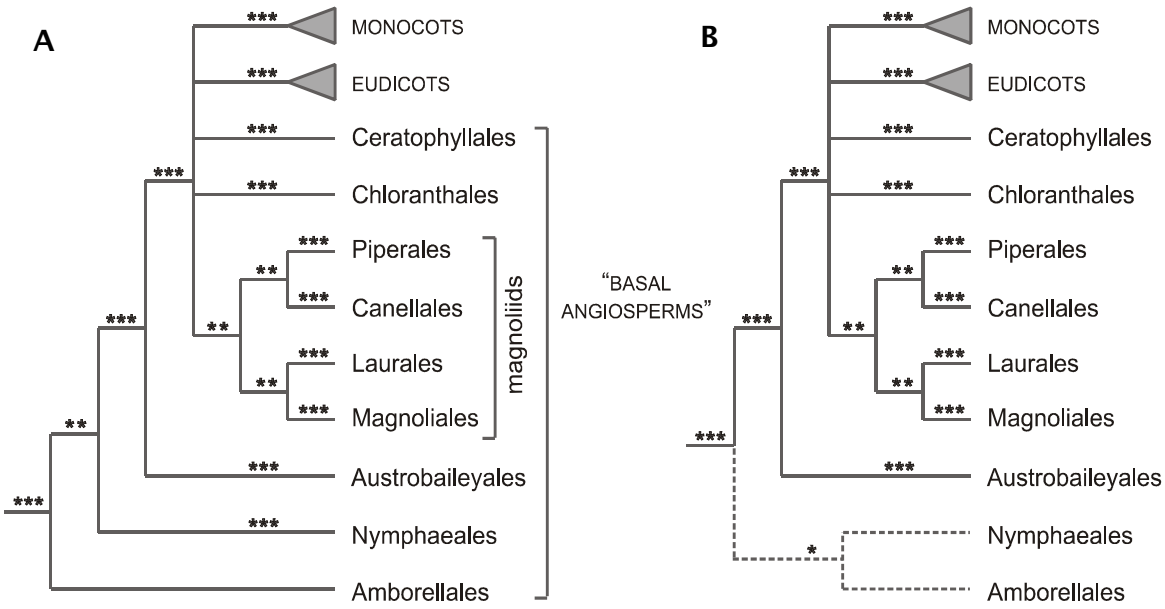


Figure 1.3 — Current understanding of phylogenetic relationships among basal angiosperms. Two alternative hypotheses exist for relationships within the basal grade. The number of stars above branches (1 to 3) symbolizes weak to strong statistical support for the respective node as derived from surveying the recent literature. **A)** A topology with *Amborella*, Nymphaeales and Austrobaileyales as successive sister lineages to the remaining angiosperms has been inferred in most of the recent phylogenetic studies based on molecular data (e.g., Parkinson *et al.* 1999, Mathews & Donoghue 2000, Soltis *et al.* 2000, Borsch *et al.* 2003, Hilu *et al.* 2003, Borsch *et al.* 2005, Löhne & Borsch 2005, Qiu *et al.* 2005). **B)** In an alternative hypothesis a clade consisting of *Amborella* plus Nymphaeales is sister to the remaining angiosperms. This hypothesis has been inferred by Barkman *et al.*(2000) and Zanis *et al.* (2002) and can not conclusively be rejected based on currently available data (e.g., Stefanovic *et al.* 2004, Leebens-Mack *et al.* 2005, Qiu *et al.* 2005).

monocots remains a challenge in flowering plant phylogenetics. All analyses of the last years, using different taxon and character sampling and employing different tree inference method, have produced conflicting topologies that are generally only weakly supported (e.g., Barkman *et al.* 2000, Qiu *et al.* 2000, Soltis *et al.* 2000, Soltis *et al.* 2002a, Zanis *et al.* 2002, Borsch *et al.* 2003, Hilu *et al.* 2003, Zanis *et al.* 2003). Regarding *Ceratophyllum*, sister–relationships to either the monocots (Zanis *et al.* 2002, Borsch *et al.* 2003, Davies *et al.* 2004) or the eudicots (Soltis *et al.* 2000, Hilu *et al.* 2003, Borsch *et al.* 2005) have been proposed. Although its position still has to await final clarification, a tendency for placing *Ceratophyllum* sister to eudicots can be observed in the most recent literature (96% posterior probability in Bayesian analyses of combined sequences of *trnT–trnF*, *matK* and *petD* in Borsch *et al.* 2005).

The genomic regions employed so far as markers in angiosperm phylogenetics come from the chloroplast genome, e.g. *rbcL* (Chase *et al.* 1993, Källersjö *et al.* 1998, Soltis *et al.* 2000), *atpB* (Soltis *et al.* 2000, Qiu *et al.* 2001, Soltis *et al.* 2002a, Qiu *et al.* 2005) and the *matK* gene (Hilu *et al.* 2003, Qiu *et al.* 2005), the mitochondrial genome, e.g. *atp1*, *matR*, *mtLSU mtSSU*, (Qiu *et al.* 1999, 2000, Qiu *et al.* 2005), *cox1* and *rps2* (Parkinson *et al.* 1999), as well as from the nuclear genome, e.g. 18S and 26S rDNA, (Soltis *et al.* 1997, Soltis *et al.* 2000, Qiu *et al.* 2005), phytochrome A and C (Mathews & Donoghue 2000), and the MADS–Box genes *Apetala3* and *Pistillata* (Aoki *et al.* 2004, Kim *et al.* 2004b). It has long been argued that only rather conserved coding regions can provide suitable markers at deeper phylogenetic levels, because saturation and homoplasy in more rapidly evolving genomic regions would blur the phylogenetic signal and cause wrong inferences. However, Källersjö *et al.* (1999) could prove that homoplasy does not necessarily lead to decreased phylogenetic accuracy and Müller *et al.* (in prep.) showed that average phylogenetic signal per informative character is significantly higher in rapidly evolving DNA than in the more slowly evolving *rbcL* gene. In fact, the utility of rapidly evolving cp DNA, such as the *matK* gene or introns and spacers, has been proven in recent studies (e.g., Borsch *et al.* 2003, Hilu *et al.* 2003, Borsch *et al.* 2005). Adding more evidence to these findings and developing another informative marker for deep level phylogenetics was the main motive behind the study described in Chapter 2 of this thesis.

1.3 The age of angiosperms

1.3.1 Evidence from the fossil record

The oldest fossils that have been assigned to angiosperms are dispersed tectate–columellate, monoaperturate pollen grains from the Hauterivian (132–127 Ma, see timescale in Figure 1.4) of England (Hughes *et al.* 1991) and Israel (Brenner & Bickoff 1992). However, since these pollen grains can not be conclusively differentiated from pollen of other mesozoic seed plants, Hughes suggests the more conservative approach of regarding only triaperturate pollen, which is characteristic for the eudicot clade, as a fossil marker for angiosperms. The earliest records for triaperturate pollen are from the Late Barremian–Early Aptian (~125 Ma) of Gabon (Doyle & Hotton 1991, Doyle 1992a) and England (Hughes 1994). From the same period the first angiosperm megafossils have been reported, i.e. stamens, carpels and fruits

similar to the extant taxon *Hedyosmum* (Chloranthaceae) from the Barremian–Aptian of Portugal (Friis *et al.* 1994b, 1999). Some years ago, reports of a whole–plant angiosperm fossil *Archaefructus* from the Late Jurassic (144 Ma) of China (Sun *et al.* 1998) caused a little sensation among palaeontologists and botanists, but the age of this fossil has recently been corrected to the Late Barremian (125 Ma, Swisher *et al.* 1999), thus being in line with other fossil angiosperm records. Up to now, three species of this aquatic early angiosperm family Archaefructaceae (Sun *et al.* 2002, Ji *et al.* 2004) have been described. Sun *et al.* (2002) placed Archaefructaceae sister to all extant angiosperms, but Friis *et al.* (2003) argued that these fossils rather represent derived members of the Nymphaeales. However, the latter inference is questionable since it was mainly biased by misinterpretation of morphology and erroneous coding of characters, i.e. ignoring the peltate leaves in *Cabomba* and coding only the dissected leaves (see Crepet *et al.* 2004).

Shortly after their first appearance in the fossil record, a drastic increase of number and diversity of angiosperm fossils can be observed, sometimes referred to as the "cretaceous explosion of angiosperm diversity" (Brenner 1962, Muller 1970, Crane *et al.* 1995, Wing & Boucher 1998, Feild & Arens 2005). From the Aptian on, angiosperms began to dominate the species composition of some low–latitude floras (e.g., Lidgard & Crane 1990, Wing & Boucher 1998) and expanded towards both poles (Retallack & Dilcher 1986, Kennedy 2003, Eklund 2004). Many extant flowering plants appeared in the Early Cretaceous, e.g. Platanaceae (*Platanocarpus*, Crane *et al.* 1993), or in the Late Cretaceous, e.g. Clusiaceae (*Paleoclusia*, Crepet & Nixon 1998a), Betulaceae (Bedellia, Sims *et al.* 1999) and Hydrangeaceae (*Tylerianthus*, Gandolfo *et al.* 1998a). Basal monocot lineages were present in the Early Cretaceous, as evident from Araceae–pollen from the Barremian–Aptian of Portugal (Friis *et al.* 2004), and more derived monocot lineages appeared in the Turonian (e.g. Triuridaceae, Gandolfo *et al.* 1998b). Wing & Boucher (1998) stated that almost half of all extant angiosperm orders already occur in the Cretaceous fossil record.

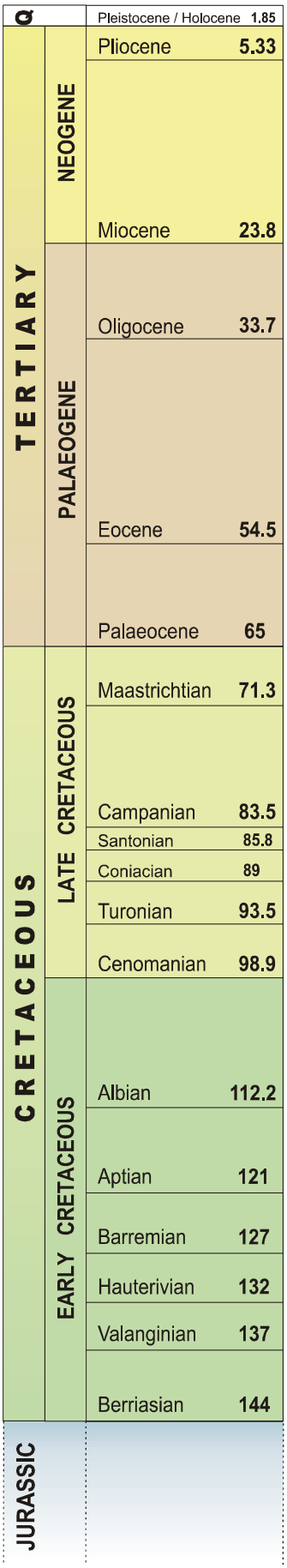


Figure 1.4 — A simplified geological timescale. Nomenclature and time frames of geological epochs follow Berggren et al. (1995) for the Tertiary and Gradstein et al. (1994) for the Cretaceous.

Oldest reports from the so–called basal angiosperm lineages are the above–mentioned chloranthaceous flowers and fruits from Barremian–Aptian of Portugal (Friis *et al*. 1994b, 1999). Several Early Cretaceous pollen types have been assigned to Chloranthaceae, e.g. *Clavatipollenites*, *Asteropollis* or *Stephanocolpites*, but none of these assignments is unequivocal (see Eklund *et al*. 2004 for review). Lauralean fossils are also known from the Cretaceous, e.g. Calycanthaceae from the Albian (*Virginianthus*, Friis *et al*. 1994a) and Lauraceae from the Cenomanian (*Mauldina*, Drinnan *et al*. 1990). Magnoliales as well as Austrobaileyales appear in the Turonian (e.g. Cronquistiflora, Crepet & Nixon 1998b, Illiciaceae, Frumin & Friis 1999). Fruits of Ceratophyllum have been reported from the Palaeocene (Herendeen *et al*. 1990) as well as from the Aptian (Dilcher 1989), but the latter fossil is regarded as equivocal (Herendeen *et al*. 1990). The fossil record of Nymphaeales, which appear in the Tertiary (but see Friis *et al*. 2001, Gandolfo *et al*. 2004), is described in detail in Chapter 5 of this thesis. Up to now, no fossils have been found that could be unambiguously assigned to the proposed "most basal angiosperm" *Amborella* (Crepet *et al*. 2004).

1.3.2 Insights from molecular studies

The existence of a molecular clock was first hypothesized by Zuckerkandl & Pauling (1962), postulating that the amount of difference between the DNA of two species is a function of the time since their evolutionary separation. Soon its potential for estimating the age of lineages was realised (Bromham & Penny 2003), but early estimations of divergence times varied markedly and showed a huge disparity to the fossil record. The age of angiosperms was calculated to 420–350 Ma (Ramshaw *et al*. 1972), 354–300 Ma (Martin *et al*. 1989, Martin *et al*. 1993) or 200 Ma (e.g. Wolfe *et al*. 1989). These early studies were mainly biased by limited character or taxon sampling (Magallón & Sanderson 2005) and by the fact that – at least in some studies – angiosperms were represented only by taxa with exceptionally high substitution rates, such as Poaceae, thereby introducing significant bias in divergence time estimates (Sanderson & Doyle 2001). Furthermore, a strict molecular clock was generally applied, i.e. assuming constant rates of molecular evolution among lineages. Recent studies revealed the stochastic nature of molecular evolution and the presence of different substitution rates in different lineages (e.g. Sanderson 1997, 1998, Sanderson & Doyle 2001, Rodriguez-Trelles *et al*. 2002, Soltis *et al*. 2002b, Magallón 2004, Sanderson *et al*. 2004). As a consequence several new methods have been developed that relax the assumption of a strict molecular clock, e.g. nonparametric rate smoothing (Sanderson 1997), local clocks (Yoder & Yang 2000), penalised likelihood (Sanderson 2002), and Bayesian approaches (e.g., Thorne *et al*. 1998, Huelsenbeck *et al*. 2000, Thorne & Kishino 2002). An overview about methods for dating phylogenies in the absence of a molecular clock is given by Welch & Bromham (2005), but see also Magallón (2005) and Renner (2005).

Additionally, more attention has been paid to the fossils that are used for calibrating the phylogenies (Crane *et al*. 2004, Crepet *et al*. 2004, Hedges & Kumar 2004, Near & Sanderson 2004, Reisz & Müller 2004, Sanderson *et al*. 2004, Heads 2005). Heads (2005) stresses possible sources of error from using fossils as calibration points, i.e. ambiguities in the taxonomic affinity or the age of the fossil itself as well as the general incompleteness of the fossil record. The fact that certain taxa are absent from certain strata of the fossil record,

does not necessarily imply that this taxa did not exist by that time. There are several examples where newly found fossils were drastically older than previous estimations from molecular data had suggested for the specific clade (e.g., Kress *et al.* 2001, Sanderson *et al.* 2004). Therefore, fossil data can only provide a minimum age of a clade, and for a correct interpretation of estimated ages it is essential to consider the confidence intervals of estimations (Graur & Martin 2004, Heads 2005).

Most recent studies on divergence times point toward a Jurassic age of the angiosperm crown group, thereby predating the occurrence of the first fossil by 60 to 5 Ma depending on the used markers and time constraints (Wikström *et al.* 2001, Davies *et al.* 2004, Sanderson *et al.* 2004, Bell *et al.* 2005, Magallón & Sanderson 2005). The divergence of the monocot lineage from the basal angiosperms is estimated to Late Jurassic – Early Cretaceous based from 61 genes of the cp genome (Chaw *et al.* 2004). Molecular based estimates for the age of eudicots are timely consistent with the fossil record, i.e. 125 Ma (Bell *et al.* 2005), or slightly predate it (Wikström *et al.* 2001). A proposed age of the asterid lineage of 128 Ma (Bremer *et al.* 2004) goes in line with the assumption of a rapid diversification into all major angiosperm lineages in the Early Cretaceous. However, molecular analyses of extant taxa also suggest that at least some clades of angiosperms are characterized by a rather recent diversification of the crown group compared to a comparatively old stem lineage. Among basal angiosperms this was suggested for Chloranthaceae (Zhang & Renner 2003), Illiciaceae (Oh *et al.* 2003), and recently also for Nymphaeaceae (Yoo *et al.* 2005, but see also Chapter 5 of this thesis).

1.4 Extant and historical biogeography of basal angiosperms

Distributional patterns and diversity of extant taxa of basal angiosperm lineages vary markedly. There are only a few species–rich families of the magnoliid clade with an almost worldwide distribution but most diversity in tropical and subtropical regions: Annonaceae (Magnoliales), Lauraceae (Laurales), Piperaceae and Aristolochiaceae (both Piperales), and Myristicaceae (Magnoliales). Some other extant lineages are less diverse but also very widespread, e.g. Ceratophyllaceae, Nymphaeales, and the three Laurales families Hernandiaceae, Monimiaceae and Atherospermaceae. Smaller families of basal angiosperms are often rather restricted in their geographic extension, e.g. Austrobaileyaceae (Australia), Amborellaceae (New Caledonia), Eupomatiaceae and Himantandraceae (both Magnoliales, Australia and New Guinea), Degeneriaceae (Magnoliales, Fiji Archipelago), Lactoridaceae (Piperales, Juan Fernández Archipelago) and the Gomortegaceae (Laurales) with only 1 species in Chile.

Several families of basal angiosperms show characteristic patterns of disjunct distribution in East to South Asia and temperate to tropical America (Illiciaceae, Schisandraceae, Calycanthaceae, Chloranthaceae, Magnoliaceae, and Saururaceae) or disjunctions between tropical America and Africa (Canellaceae). However, all these families appear in the fossil record of Europe, either in the Tertiary (Illiciaceae, Schisandraceae, Calycanthaceae, Magnoliaceae, and Saururaceae, Mai 1995) or in the Cretaceous (Canellaceae, see Feild *et al.* 2002 and references therein, and Chloranthaceae, Eklund *et al.* 2004). Thus, for most of

the geographically restricted lineages of basal angiosperms there is evidence that the extant distribution represents only relictual ranges. For instance, in the case of Winteraceae, which are nowadays scattered to small patches in South America, Australia, South East Asia and Madagascar, the fossil record suggests a widespread southern hemispheric distribution of the family during the Cretaceous (Doyle 2000, Feild *et al*. 2002, Eklund 2004). Even extreme local endemics, like the monotypic family Lactoridaceae from the Juan Fernández Archipelago west of Chile, probably had a wider distribution during earlier times, as suggested from fossil remains from the Turonian of southern Africa (Zavada & Benson 1987) and the Maastrichtian to Mid–Tertiary of Australia (Macphail *et al*. 1999). But also those families, which are today widespread in all tropical regions (e.g., Annonaceae, Piperaceae, Myristicaceae, Lauraceae), have been present in the Tertiary of Europe and North America (Mai 1995, Chanderbali *et al*. 2001, Doyle *et al*. 2004, Richardson *et al*. 2004), and in some cases also in Antarctica (several families of Laurales, see Renner 2004d). A wider distribution of mostly tropical lineages in the Tertiary is not surprising in view of the climatic conditions during that period. There is strong evidence for higher global temperatures and less pronounced latitudinal climatic gradients during the Cretaceous and the Tertiary (e.g., Lawver & Gahagan 1998, Hay *et al*. 1999, Zachos *et al*. 2001). Especially warm periods have been reported for the Palaeocene–Eocene boundary and for the Middle Miocene (Zachos *et al*. 2001). During these periods, a more or less continuous tropical vegetation covered large parts of the northern hemisphere ("boreotropical flora", Wolfe 1975).

Traditionally, a Gondwanan or Laurasian origin has been assumed for most of the large basal angiosperm lineages based on the extant distribution and on information from the fossil record. Extant disjunct patterns have generally been explained by vicariance due to continental drift and/or extinction due to climatic deterioration (see, e.g, Parenti & Humphries 2004 for review). However, new insights from increasing molecular data and improving methods in dating of phylogenies could show that the historical biogeographic pattern of lineages is often much more complex than previously thought (Donoghue & Moore 2003). In many cases, molecular based estimations of divergence times revealed that the age of a given lineage with disjunct distribution is much younger than the break–up of the respective continents. Azuma *et al*. (2001) could show that tropical and temperate disjunctions in Magnoliaceae can be attributed to separate speciation events and subsequent migrations through the northern hemisphere during distinct intervals of the Eocene. A very complex picture has also been drawn of the biogeographic history of Laurales: whereas the split into Laurales families and major clades within these happened contemporaneous to the break–up of Gondwana, there is strong evidence for more recent migrations of younger subclades through the northern hemisphere as well as for several long–distance dispersal events (Renner *et al*. 2000, Chanderbali *et al*. 2001, Renner 2004d). Similar scenarios have been drawn for other lineages in basal angiosperms such as Annonaceae and Myristicaceae (Doyle *et al*. 2004, Richardson *et al*. 2004, Scharaschkin & Doyle 2005), as well as for several clades of the eudicots, e.g. Malpighiaceae (Davis *et al*. 2002b, Davis *et al*. 2004), Melastomataceae (Renner *et al*. 2001, Renner 2004a, b) or Fabaceae (Lavin *et al*. 2001, Lavin *et al*. 2005). New insights from molecular phylogenetics on the age of lineages have renewed the methods and approaches in historical biogeography. The relevance of changing climates for the migration of plants (Morley 2003), but also the importance of long–distance dispersal across oceanic barriers, mediated by wind or sea currents is gaining more attention

and a wider acceptance among the scientific community (Givnish & Renner 2004, Muñoz *et al.* 2004, Renner 2004c, McGlone 2005). De Queiros (2005) even regards this new trend as a "resurrection of oceanic dispersal in historical biogeography".

1.5 Nymphaeales — the first globally diverse clade of angiosperms

In the field of basal angiosperm research Nymphaeales are of special interest for several reasons. They represent, together with *Amborella* and the Austrobaileyales, one of the earliest branching lineages in the tree of angiosperms (see Chapter 1.2.4), and are, therefore, crucial for understanding the early evolution and diversification of the flowering plants. Considering their geographic distribution, they can be described as the "first globally diverse clade of angiosperms": whereas *Amborella* and most members of Austrobaileyales are characterized by rather restricted distributions (see Chapter 0), the water lilies and their relatives, i.e. the order Nymphaeales, are distributed almost all over the world (Figure 1.5).

The Nymphaeales comprise about 70 species of aquatic plants in eight genera: *Brasenia* Schreb. (water shield), *Cabomba* Aubl. (fanworts), *Barclaya* Wall., *Euryale* Salisb., *Nuphar* Sm. (pond lily), *Nymphaea* L. (water lily), *Ondinea* Hartog, and *Victoria* Lindl. (Amazon water lily). Contemporary treatments of Nymphaeales favour the recognition of two families, Cabombaceae (comprising *Brasenia* and *Cabomba*) and Nymphaeaceae (comprising the other six genera). Earlier classifications of water lilies and relatives (overview in Table 1.1) differed from the present system in grouping the genera into more families such as Barclayaceae (Li 1955, Cronquist 1988), Euryalaceae (Li 1955), or Nupharaceae (Kerner von Marilaun 1891, Nakai 1943, Takhtajan 1997). More remarkably, until the 1990's the genera *Ceratophyllum* and *Nelumbo* were assumed to be closely related to water lilies and were generally included in the circumscription of Nymphaeales. However, accumulating

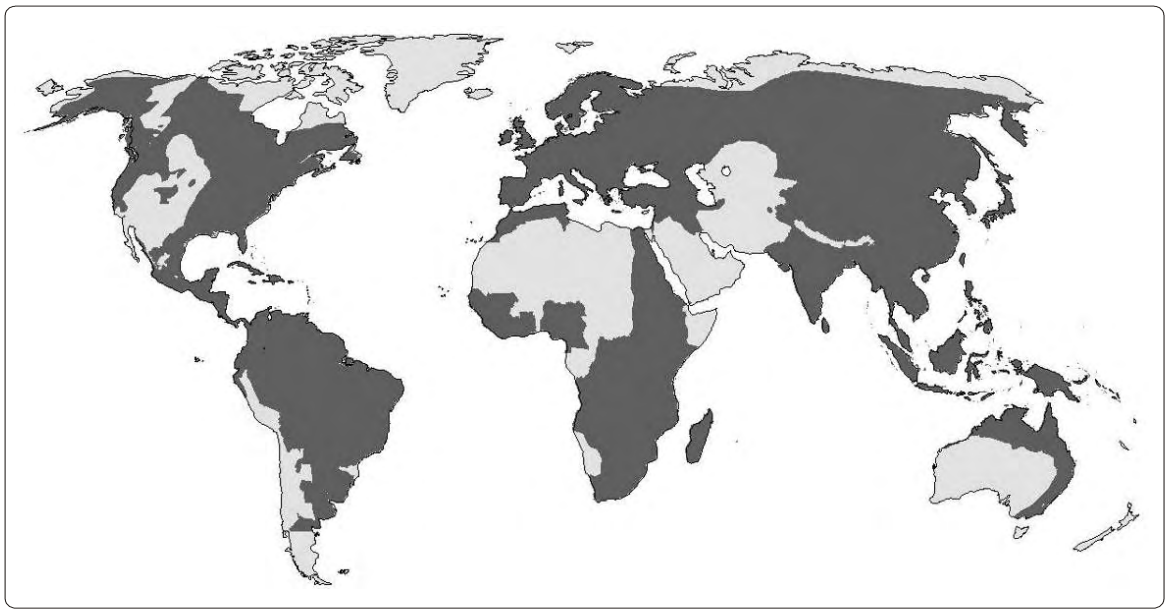


Figure 1.5 — Distribution of Nymphaeales, the first globally diverse clade. This map is based on a detailed examination of species ranges which has been carried out in the course of the present PhD–project. An analysis of extant and historical biogeographic patterns in Nymphaeales is presented in Chapter 5 of this thesis.

morphological (Simon 1970, Kak & Durani 1986, Les 1988, Donoghue & Doyle 1989, Moseley *et al*. 1993), palaeobotanical (Collinson 1980), and molecular evidence (e.g. Les *et al*. 1991, Hamby & Zimmer 1992, Chase *et al*. 1993) clearly showed that similarities among these taxa where due to parallel adaptations to aquatic life. Thus, the exclusion of *Nelumbo* and *Ceratophyllum* from Nymphaeales seemed justified and was adopted in more recent classifications (e.g., Thorne 1992, Takhtajan 1997, Thorne 2000).

In view of the uncertainties regarding the relationships within the basal grade of angiosperms, i.e. *Amborella* basal or *Amborella* + Nymphaeales basal (Chapter 1.2.4, page 10 ff.), the Angiosperm Phylogeny Group did not assign an ordinal rank to Nymphaeaceae and Cabombaceae in their most recent angiosperms classification (APG II 2003), but placed them together with Amborellaceae and Chloranthaceae at the beginning of their list of taxa. Furthermore, a broader circumscription Nymphaeaceae as including Cabombaceae was suggested emphasizing their common descent (APG II 2003). However, given the substantial differences between Cabombaceae and Nymphaeaceae (e.g. Schneider & Williamson 1993, Williamson & Schneider 1993) a broad circumscription of Nymphaeaceae s.l. did not gain wider acceptance. Instead, most recent studies (e.g., Yoo *et al*. 2005) follow Thorne's classification (1992, 2000), which is adopted in the present thesis as well.

Nymphaeales in the above described sense represent a well–defined clade of basal angiosperms since the monophyly of this group gains strong support in all recent molecular phylogenetic studies (e.g., Parkinson *et al*. 1999, Mathews & Donoghue 2000, Soltis *et al*. 2000, Borsch *et al*. 2003, Hilu *et al*. 2003, Löhne & Borsch 2005). Relationships within Nymphaeales are, however, less clear. Especially the monophyly of the family Nymphaeaceae, as comprising *Nuphar*, *Barclaya*, *Euryale*, *Victoria*, *Nymphaea* and *Ondinea*, is not convincingly supported (e.g. 50% bootstrap support from *rbcL*, Les *et al*. 1991, 77% jackknife support from *trnT–trnF*, Borsch *et al*. subm.). In the course of the present project (Chapter 2 and 3) it could be shown that this weakly supported node is mainly caused by an uncertain position of *Nuphar* at the base of the Nymphaeaceae. Furthermore, first molecular analyses employing a comprehensive taxon sampling (Borsch 2000, Borsch *et al*. subm.) revealed that also the genus *Nymphaea* is statistically not very well supported as a clade; and the monotypic genus *Ondinea* was inferred to be sister to some of the Australian water lilies (*Nymphaea* subg. *Anecphya*) by the *trnT–trnF* data set (Borsch *et al*. subm.). These previous results emphasize the need for a comprehensive approach to unravel the phylogenetic history of Nymphaeales.

The order Nymphaeales, with a generic and species–level diversity suitable for comprehensive phylogenetic and biogeographic studies, lends itself for a case study in historical biogeography. The long–standing tradition in morphological–anatomical and systematic studies in water lilies and relatives (e.g., Salisbury 1806, De Candolle 1819, Caspary 1865, 1891, Weberbauer 1894, Masters 1902, Conard 1905, Grüß 1927a, b) and the comprehensive analyses from the later 20th century (see several references for E. Schneider, P. Williamson, M. Moseley, J. Wiersema, and W. Weidlich between 1958 and 2003) provide a solid background for interpretation of results from molecular phylogenetic analyses. In view of a broad current knowledge on the biology of Nymphaeales, which is

Table 1.1 — Overview on classifications of water lilies and relatives (Nymphaeales). Until the 1990's *Ceratophyllum* and *Nelumbo* usually were subsumed under the order Nymphaeales. The views on phylogenetic relationships changed with the introduction of cladistic methods. Consequently *Nelumbo* and *Ceratophyllum* are now treated separately from Nymphaeales, which today comprise only Cabombaceae and Nymphaeaceae (sensu Thorne 2000). Suffixes of family names are abbreviated in order to save space.

Caspary/Engler (1891)	Li (1955)	Thorne (1976)	Tamura (1982)	Ito (1987)	Cronquist (1988)	Takhtajan (1997)	Thorne (1992, 2000)	APG (1998, 2003)
RANALES Ceratophyllac. - *Ceratophyllum* Nymphaeac. Nelumbonoideae - *Nelumbo* Cabomboideae - *Cabomba* - *Brasenia* Nymphaeoideae - *Nuphar* - *Barclaya* - *Nymphaea* - *Victoria* - *Euryale*	**NELUMBONALES** Nelumbonac. - *Nelumbo* **RANALES** Ceratophyllac. - *Ceratophyllum* Cabombac. - *Cabomba* - *Brasenia* Nymphaeac. - *Nuphar* - *Nymphaea* Barclayac. - *Barclaya* **EURYALES** Euryalac. - *Victoria* - *Euryale*	**NYMPHAEALES** Nelumbonac. - *Nelumbo* Ceratophyllac. - *Ceratophyllum* Cabombac. - *Cabomba* - *Brasenia* Nymphaeac. - *Nuphar* - *Barclaya* - *Nymphaea* - *Victoria* - *Euryale* - *Ondinea*	**NYMPHAEALES** Nelumbonac. - *Nelumbo* Ceratophyllac. - *Ceratophyllum* Nymphaeac. Hydropeltidoideae - *Cabomba* - *Brasenia* Nymphaeaoideae - *Nuphar* - *Barclaya* - *Nymphaea* - *Victoria* - *Euryale* - *Ondinea*	**NYMPHAEALES** Nelumbonac. - *Nelumbo* Ceratophyllac. - *Ceratophyllum* - *Cabomba* - *Brasenia* Nymphaeac. - *Nuphar* - *Barclaya* - *Nymphaea* - *Victoria* - *Euryale* - *Ondinea*	***MAGNOLIIDAE*** **NYMPHAEALES** Nelumbonac. - *Nelumbo* Ceratophyllac. - *Ceratophyllum* Cabombac. - *Cabomba* - *Brasenia* Barclayac. - *Barclaya* Nymphaeac. - *Nuphar* - *Nymphaea* - *Victoria* - *Euryale* - *Ondinea*	***NELUMBONIDAE*** **NELUMBONALES** Nelumbonac. - *Nelumbo* ***NYMPHAEIDAE*** **CERATOPHYLLALES** Ceratophyllac. - *Ceratophyllum* **HYDROPELTIDALES** Hydropeltidac. =Cabombaceae - *Cabomba* - *Brasenia* **NYMPHAEALES** Nupharac. - *Nuphar* Barclayac. - *Barclaya* Nymphaeac. - *Nymphaea* - *Victoria* - *Euryale* - *Ondinea*	***MAGNOLIIDAE*** **NELUMBONALES** Nelumbonc. - *Nelumbo* **CERATOPHYLLALES** Ceratophyllaceae - *Ceratophyllum* **NYMPHAEALES** Cabombac. - *Cabomba* - *Brasenia* Nymphaeac. Nymphaeoideae - *Nuphar* - *Nymphaea* - *Ondinea* Euryaloideae - *Victoria* - *Euryale* Barclayoideae - *Barclaya*	***"BASAL FAMILIES"*** not assigned to higher taxa: Nymphaeac. (optionally including Cabombac.) - *Cabomba* - *Brasenia* - *Nuphar* - *Barclaya* - *Nymphaea* - *Ondinea* - *Victoria* - *Euryale* ***EUDICOTS*** **PROTEALES** Nelumbonac. - *Nelumbo*

based on a variety of studies on physiology and secondary compounds (e.g., Schmucker 1932, Schmucker 1933, Emboden 1978, 1979, Fossen *et al*. 1998, Fossen & Andersen 1999), cytology (Langlet & Söderberg 1927, Bukowiecki *et al*. 1972, Gupta 1978, 1980, Okada & Tamura 1981), and ecology (e.g., Valla & Cirino 1972, Prance & Arias 1975, Prance & Anderson 1976, Prance 1980, Schneider 1982, Capperino & Schneider 1985, Cook 1988, Wiersema 1988, Williamson & Schneider 1994, Lippok *et al*. 2000, Ervik & Knudsen 2003, Hirthe & Porembski 2003), it seems to be the best time for an integrative study on the evolution of this group in space and time. The solid basis for the present study is complemented by an extensive fossil record for Nymphaeales (e.g., Collinson 1980, Mai 1988, Chen *et al*. 2004, see overview given in Chapter 5.1).

1.6 Aims of the study and project design

In line with the general idea of the Graduiertenkolleg "Evolution and biodiversity in space and time", the main objective of the present PhD–project is to study the history of a selected group of flowering plants, the Nymphaeales, and to identify radiations that led to the extant patterns of diversity. In order to outline a plausible scenario for the evolution of Nymphaeales in space and time, a combination of approaches from the fields of phylogenetics, biogeography, and palaeobotany is necessary (Figure 1.6). Above all, a profound hypothesis on phylogenetic relationships within the study group is essential. In view of the poor resolution and support for crucial relationships within the study group (as outlined in previous paragraph), a main focus of this project is on reconstructing a well–supported phylogeny of Nymphaeales.

The success of phylogenetic studies in terms of finding a well–supported tree is strongly dependent on the quality and information content of the utilised genomic markers. It is very important to choose a suitable marker, depending on the kind of relationships that are to be resolved (e.g., interspecific, interfamilial, or between higher taxonomic levels). In a first part of the project, therefore, a new marker – the group II intron of the chloroplast *petD* gene – is developed and its utility for inferring relationships among basal angiosperms is being evaluated (Chapter 2). Together with several other non–coding and rapidly evolving cp genomic regions, this marker is then utilised to infer the phylogeny of the study group Nymphaeales (Chapter 3). A very comprehensive approach, i.e. sampling different genomic regions (introns, spacers, and a gene) for a broad taxon sampling (covering all genera of Nymphaeales and subgenera of *Nymphaea*), was necessary to resolve intriguing phylogenetic problems in this order. The molecular phylogenetics part of this project is complemented by a detailed examination of relationships within a specific subgenus of *Nymphaea* that is endemic to the wet tropics of Australia and New Guinea and that represents a rather young radiation in the water lily family (Chapter 4). The establishment of large molecular data sets based on a comprehensive DNA sample base was facilitated due to the integration of the present project into existing collaborations between Thomas Borsch (University of Bonn), John Wiersema (US Department of Agriculture, Beltsville / MD), Surrey Jacobs (Royal Botanical Gardens, Sydney) and Barre Hellquist (Massachusetts College of Liberal Arts, North Adams / MA).

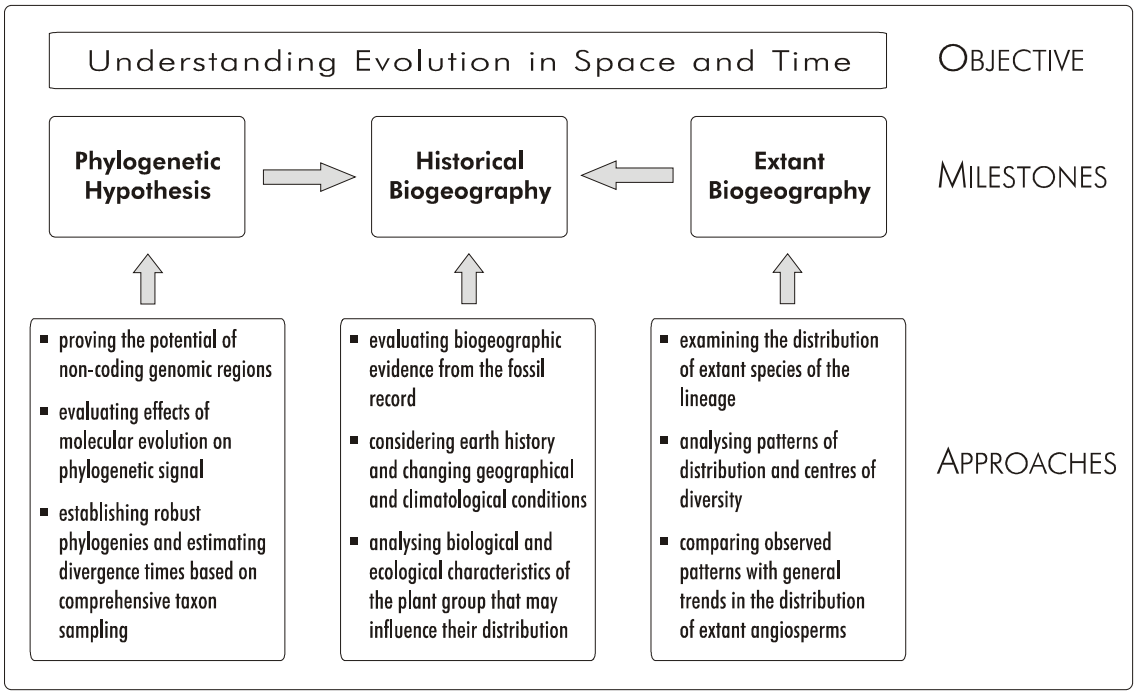


Figure 1.6 — Design and approaches of the present study on historical biogeography of Nymphaeales. A main focus of this PhD–project will be on establishing robust phylogenetic hypotheses. Extant and historical biogeography of Nymphaeales are examined in the integrative final chapter of this thesis.

Additionally to the molecular phylogenetics study, data on the distribution of all extant species in Nymphaeales (currently 69) had to be collected. Only with a profound idea on phylogenetic relationships and the extant distribution patterns in Nymphaeales, an effort could be made to track the biogeographic history of that lineage (Chapter 5). The integrative final section of the present thesis is in part the outcome of a close cooperation with Mi-Jeong Yoo from the "Soltis Lab" at the Florida Museum of Natural History (Gainesville, FL). Mi–Jeong Yoo has used the above–mentioned molecular data set on Nymphaeales to estimate the divergence times in this group as a part of her dissertation on floral development in Nymphaeales. An idea about the age of the lineage and its subclades is one prerequisite for studying historical biogeography. However, an overview on important geological and climatological events in earth history as well as a critical evaluation of the important factors inherent to the plants themselves (life form, dispersability, and other ecological factors) is also essential.

A contribution towards understanding the evolution of basal angiosperms in space and time by shedding light on the phylogenetic history of Nymphaeales and their past distribution is the main objective of the present study. On the basis of this case study the potential of integrating findings from molecular phylogenetics, biogeography, and palaeobotany for drawing a comprehensive picture of the evolution of lineages shall be demonstrated.

Chapter 2

Phylogeny of basal angiosperms based on the *petD* group II intron

Contents

2.1 Introduction

Phylogenetic studies based on DNA sequences from all three plant genomes have greatly improved our understanding of the early evolution of angiosperms. Combining three (Soltis *et al.* 2000), five (Qiu *et al.* 1999), and six to 11 genes (Zanis *et al.* 2002) from two or three genomes or up to 17 genes from the chloroplast genome (Graham & Olmstead 2000b) led to the identification of *Amborella* as the first branch in angiosperms followed by Nymphaeales and Austrobaileyales and of major clades such as the magnoliids consisting of Piperales + Canellales and Magnoliales + Laurales (Qiu *et al.* 1999, Graham & Olmstead 2000b, Mathews & Donoghue 2000, Zanis *et al.* 2002). Still, phylogenetic affinities of the monocots, eudicots, Ceratophyllales and Chloranthales remain unclear. Qiu *et al.* (1999) argued that the inclusion of even more genes might help to resolve these questions. Recent analyses of whole chloroplast genome sequences refreshed the discussion on the root of angiosperms by inferring Poaceae basal (Goremykin *et al.* 2003). However, given the effort of sequencing numerous genes, the number of taxa that can be included usually is a limiting factor. Finding markers offering strong historic signal for the amount of sequence generated, therefore, becomes crucial. This signal might not only be provided by substitutions but also provided by length mutations and genomic rearrangements.

To infer phylogenetic relationships at deeper levels, rather conserved genes such as *rbcL* have been sequenced, whereas rapidly evolving non–coding regions served to infer relationships among species and genera. Examples of such rapidly evolving parts in the chloroplast genome are the intergenic spacers between *atpB* and *rbcL*, *trnT* and *trnL*, *trnL* and *trnF*, the group I intron in *trnL*, and the group II introns in *rpl16*, *rps16*, *rpoC1*, *trnV*, and *ndhA* (see Soltis & Soltis 1998, and Kelchner 2000, Kelchner 2002 for reviews), all of which are located in the large and small single-copy regions. The group II introns in *rpl2*, *rps12*, and *ndhB* are situated in the highly conserved inverted repeat and were sequenced for a set of basal angiosperms by Graham & Olmstead (2000a) and Graham *et al.* (2000). Introns in particular possess a mosaic of highly conserved core elements that are responsible for their function alternating with sequence stretches that might be more or less freely evolving. Nevertheless, the overall variability of spacers and introns is higher than in most coding DNA. Therefore, it has been assumed that high substitution rates are present, leading to saturation and homoplasy, and that frequent length mutations cause homology assessment (alignment) of non–coding sequences to be difficult or even impossible in data sets covering a broad taxonomic spectrum.

Recently, Borsch *et al.* (2003) employed the non-coding parts of the *trnT-trnF* region, consisting of two spacers and the *trnL* group I intron, for deep-level phylogenetic analysis of angiosperms. Their study revealed that high length variability in trnT-trnF was confined to mutational hotspots in the intron and that these corresponded to certain stem-loop elements of the proposed secondary structure. Considering length mutations as single evolutionary events involving one or more nucleotides allowed for reliable alignment of *trnT-trnF* sequences. This finding and others suggest that microstructural changes in chloroplast genomes follow certain patterns and that understanding the nature of these patterns is essential for phylogenetic interpretation of length variability of sequences (Graham *et al.* 2000, Kelchner 2000). Analysis of the *trnT-trnF* data revealed a tree for basal angiosperms

largely congruent with multi–gene, multi–genome studies (Qiu *et al*. 1999, Soltis *et al*. 2000, Zanis *et al*. 2002) with most nodes gaining high statistical support. Several questions arise concerning the general significance of these findings. Does extreme variability in other non-coding cpDNA regions also correspond to particular structural elements (resulting in mutational hotspots), and can they be confidently aligned similar to *trnT-trnF*? Are other non-coding regions effective for deep-level phylogenetic studies as well, or is *trnT-trnF* an exception? Are there differences in the phylogenetic utility of different types of non-coding regions, such as spacers and group I and group II introns?

Variation in intron sequences is to a large extent correlated with the secondary structure of their RNA, which is essential for the self-splicing function of the intron (Learn *et al*. 1992). Based on differing RNA folding patterns, organelle introns are classified into group I and II (Michel *et al*. 1989, Michel & Ferat 1995). Because the *trnL* intron employed by Borsch *et al*. (2003) is a group I intron, we searched for an omnipresent and alignable group II intron in the chloroplast genome to compare information content of group I and II introns. We restricted our analyses to the chloroplast genome because it is inherited as a single linkage unit. Consequently, differences between group I and group II introns could be more clearly linked to evolutionary processes operating at structurally and functionally different loci, without having to worry about effects of recombination, hybridization, or lineage sorting (Doyle 1992b).

Structure and function of group II introns have been studied in detail by several authors (Michel *et al*. 1989, Knoop & Brennicke 1993, Michel & Ferat 1995, Bonen & Vogel 2001, Federova *et al*. 2003). The secondary structure model (Michel *et al*. 1989, Michel & Ferat 1995) has largely been validated in numerous experiments (see Kelchner 2002 for review). Utilizing group II introns as phylogenetic tools, their presence or absence was found to provide valuable information among land plant lineages (Qiu *et al*. 1998, Pruchner *et al*. 2002). In angiosperms, intron losses in different plant lineages have been reported from several chloroplast genes, such as *rps16*, *rpoC1*, and *rpl16* (see Kelchner 2002 for review). At the sequence level, sound knowledge of the secondary structure allows recognition of structure-linked mutation patterns and, subsequently, their evaluation in a phylogenetic context (Kelchner 2002). So far, chloroplast group II intron sequences yielded well-resolved phylogenies at the genus level (Kelchner & Clark 1997, Asmussen & Chase 2001, Clausing & Renner 2001). The mitochondrial *nad5* intron proved useful among ferns and allies (Vangerow *et al*. 1999). To examine their broader applicability, we generated a data set covering the range of seed plants.

The intron in *petD* was selected for this study because (1) no losses of this locus have been reported for angiosperms and gymnosperms, (2) initial alignment using sequences of available chloroplast genomes was successful, and (3) the presence of highly conserved sequences in the flanking regions suitable for designing universal primers for amplification. To our knowledge, the *petD* intron or the *petB-petD* intergenic spacer so far had never been used in phylogenetic studies. The *petD* gene is part of the *psbB* operon (Westhoff & Herrmann 1988). The pentacistronic primary transcript of this operon is processed into monocistronic and dicistronic mRNAs, with *petB* and *petD* normally staying connected as a dicistronic mRNA (Rock *et al*. 1987, Tanaka *et al*. 1987, Dixit *et al*. 1999, Monde *et al*. 2000). The present study has involved characterization of patterns of variability and

homoplasy in the *petD* intron of flowering plants. We discuss the impact of structural (and functional) constraints on substitutions and microstructural changes in *petD* and test hypotheses concerning the phylogenetic relationships of basal angiosperms. This second aspect of our work has allowed us to evaluate the potential of group II introns as molecular markers for deeper–level phylogenetic problems.

2.2 Materials and Methods

2.2.1 Taxon sampling and plant material

Sequences of the *petD* intron and the *petB–petD* intergenic spacer were obtained from 47 angiosperms (representing 30 families) and three gymnosperms. The *Pinus* sequence was taken from GenBank (Tsudzuki *et al*. 1992). All taxa included in this study are listed in Table 2.1 on page 27, along with their respective families, the origin of material, and GenBank accession numbers. To compare the results of this study with the analysis based on *trnT-trnF* (Borsch *et al*. 2003), we also carried out analyses of a reduced dataset congruent with the 42 taxa in the *trnT-trnF* study (*Chloranthus officinalis*, *Impatiens noli-tangere*, *Magnolia sieboldii*, *Nymphaea micrantha*, *Nymphaea nouchali*, *Piper crocatum*, *Pseudowintera colorata* and *Tasmannia insipida* were excluded).

2.2.2 DNA isolation

Total genomic DNA was isolated from fresh or silica gel–dried leaf tissue. To gain an optimal quantity of high quality DNA, a modified CTAB method with triple extractions was used (Borsch *et al*. 2003). After chloroform extraction, DNA was precipitated with isopropanol, resuspended in TE, and further purified by ammonium acetate and sodium acetate washing steps followed by ethanol precipitation.

2.2.3 Primer design

Universal primers to amplify the *petD* region (consisting of the *petB-petD* intergenic spacer, the *petD* 5′-exon, and the *petD* intron) in seed plants were designed based on the completely sequenced chloroplast genomes of *Arabidopsis thaliana* (GenBank accession number NC001284), *Lotus japonicus* (GenBank accession number NC002694), *Nicotiana tabacu*m (GenBank accession number NC001879), *Pinus thunbergii* (GenBank accession number NC001631), and *Spinacia oleracea* (Gen-Bank accession number NC001631). The *petD* region was amplified in one fragment with the forward primers PIpetB1411F (5′-GCCGTMTTTATGTTAATGC-3′) or PIpetB1365F (5′-TTGACYCGTTTTTATAGTTTAC-3′) that anneal to the 3´-exon of *petB* and the reverse primer PIpetD738R (5′-AATTTAGCYCTTAATACAGG-3′) that anneals to the 3′-exon of *petD* (Fig. 1). PIpetD346R (5′-TCTTCCTYAGATCCC-3′) was designed as an additional internal sequencing primer located in domain I of the *petD* intron because pherograms were not readable after homonucleotide strings in the *petB-petD* spacer of *Aristolochia* and *Ginkgo*.

Table 2.1 — Taxa used in the study, their respective families, source of material, location of voucher specimens, and GenBank accession numbers of deposited sequences. "BG" = Botanical Garden.

Genus / Species	Family	Garden / Field Origin	Voucher	GenBank accession number
Acorus calamus L.	Acoraceae	Bonn BG.	C. Löhne 51 (BONN)	AY590840
Aextoxicon punctatum Ruiz & Pav.	Aextoxicaceae	Bonn BG	T. Borsch 3459 (BONN)	AY590831
Amborella trichopoda Baill.	Amborellaceae	Univ. of California, Sta. Catarina BG	T. Borsch 3480 (VPI)	AY590876
Annona muricata L.	Annonaceae	Bonn BG	T. Borsch 3460 (BONN)	AY590843
Asimina triloba Dun.	Annonaceae	Bonn BG	T. Borsch 3461 (BONN)	AY590842
Orontium aquaticum L.	Araceae	Bonn BG	T. Borsch 3457 (BONN)	AY590839
Araucaria araucariana C. Koch	Araucariaceae	Bonn BG	T. Borsch 3462 (BONN)	AY590878
Nypa fruticans Wurmb.	Arecaceae	Bonn BG	T. Borsch 3463 (BONN)	AY590837
Aristolochia pistolochia L.	Aristolochiaceae	France, Herault	T. Borsch 3257 (FR)	AY590862
Saruma henryi Oliv.	Aristolochiaceae	Bonn BG	T. Borsch 3456 (BONN)	AY590861
Austrobaileya scandens C. White	Austrobaileyaceae	Bonn BG	T. Borsch 3464 (BONN)	AY590867
Impatiens noli-tangere L.	Balsaminaceae	Germany	T. Borsch 3485 (BONN)	AY590830
Buxus sempervirens L.	Buxaceae	Bonn BG	T. Borsch 3465 (BONN)	AY590832
Brasenia schreberi Gmelin	Cabombaceae	USA, Virginia	T. Borsch & T. Wieboldt 3298 (VPI, FR)	AY590869
Cabomba caroliniana Gray	Cabombaceae	USA, Virginia	J.C. Ludwig, s.n. (VPI)	AY590868
Calycanthus floridus L. var. *laevigatus* (Willd.) T. & G.	Calycanthaceae	Bonn BG	T. Borsch 3455 (BONN)	AY590849
Canella winterana Gaertn.	Canellaceae	Bonn BG	T. Borsch 3466 (BONN)	AY590851
Ceratophyllum demersum L.	Ceratophyllaceae	Bonn BG	C. Löhne 52 (BONN)	AY590841
Chloranthus brachystachys Bl.	Chloranthaceae	Bonn BG	T. Borsch 3467 (BONN)	AY590864
Chloranthus officinalis Blume	Chloranthaceae	Bonn BG	T. Borsch & C. Löhne 3491 (BONN)	AY590863
Dicentra eximia (Ker Gawl.) Torr.	Fumariaceae	Bonn BG	T. Borsch 3468 (BONN)	AY590835
Ginkgo biloba L.	Ginkgoaceae	Virginia Tech BG	T. Borsch 3469 (BONN)	AY590877
Illicium floridanum Ellis	Illiciaceae	Bonn BG	T. Borsch 3552 (BONN)	AY590865
Lactoris fernandeziana Phil.	Lactoridaceae	DNA from T.F. Stuessy	T.F. Stuessy & D. Crawford s.n.	AY590858
Umbellularia californica Nutt.	Lauraceae	Bonn BG	T. Borsch 3471 (BONN)	AY590850
Liriodendron tulipifera L.	Magnoliaceae	USA, Virginia	T. Slotta s.n. (VPI)	AY590844

Table 2.1 (continued)

Genus / Species	Family	Garden / Field Origin	Voucher	GenBank accession number
Magnolia officinalis Rehder & E.H. Wilson	Magnoliaceae	Bonn BG	C. Löhne 53 (BONN)	AY590846
Magnolia sieboldii K. Koch	Magnoliaceae	Bonn BG	C. Löhne 54 (BONN)	AY590845
Michelia champaca L.	Magnoliaceae	Bonn BG	T.Borsch 3472 (BONN)	AY590847
Myristica fragrans Houtt.	Myristicaceae	Bonn Agr. BG	T.Borsch 3473 (BONN)	AY590848
Nelumbo nucifera ssp. *lutea* (Willd.) Borsch & Barthlott	Nelumbonaceae	USA, Missouri	T. Borsch & B. Summers 3220 (FR)	AY590836
Nuphar advena (Aiton) W.T. Aiton	Nymphaeaceae	USA, Florida	T. Borsch & V. Wilde 3093 (FR)	AY590871
Nuphar lutea (L.) Sm.	Nymphaeaceae	Germany, Hesse	T. Borsch 3337 (FR)	AY590872
Nymphaea micrantha Guill. & Perr.	Nymphaeaceae	Bonn BG 5830 [Zimbabwe]	T. Borsch s.n. (BONN)	AY590874
Nymphaea cf. nouchali Burm. f.	Nymphaeaceae	Bonn BG 14244 [Rwanda]	E. Fischer s.n. (BONN)	AY590875
Nymphaea odorata Ait. ssp. *tuberosa* (Paine) Wiersema & Hellquist	Nymphaeaceae	Canada	T. Borsch, B. Hellquist & J. Wiersema 3389 (BONN)	AY590873
Victoria cruziana Orbign.	Nymphaeaceae	Bonn BG	C. Löhne 55 (BONN)	AY590870
Pinus thunbergii Parl.	Pinaceae	see Tsudzuki *et al.* 1992 (1992)		NC001631
Piper crocatum Ruiz & Pav.	Piperaceae	Bonn Bot. Gard	T. Borsch 3553 (BONN)	AY590857
Piper ornatum N. E. Br.	Piperaceae	Bonn BG	C. Löhne 56 (BONN)	AY590856
Piper sp.	Piperaceae	Bonn BG	T. Borsch 3475 (BONN)	AY590855
Platanus occidentalis L.	Platanaceae	USA, Virginia	T. Slotta s.n. (VPI)	AY590834
Houttuynia cordata Thunb.	Saururaceae	Bonn BG	T.Borsch 3481 (BONN)	AY590860
Saururus chinensis Hort. ex Loud.	Saururaceae	Bonn BG	C. Löhne 57 (BONN)	AY590859
Schisandra chinensis (Turcz.) Baill.	Schisandraceae	Bonn BG	T. Borsch & C. Löhne 3492 (BONN)	AY590866
Tofieldia glutinosa (Michx.) Pers.	Tofieldiaceae	USA	T. Borsch, B. Hellquist & J. Wiersema 3397 (VPI, BONN)	AY590838
Trochodendron aralioides Siebold & Zucc.	Trocho-dendraceae	Bonn BG	T. Borsch 3478 (BONN)	AY590833
Drimys winteri J.R. Forst. & G. Forst.	Winteraceae	Bonn BG	T. Borsch 3479 (BONN)	AY590852
Pseudowintera colorata (Raoul) Dandy	Winteraceae	Bonn BG	T. Borsch 3490 (BONN)	AY590853
Tasmannia insipida R.Br.ex DC	Winteraceae	Bonn BG	C. Löhne 58 (BONN)	AY590854

2.2.4 Amplification and sequencing

PCR was performed in a T3 Thermocycler (Biometra, Göttingen/Germany) with initial denaturation (1.5 min at 96°C), 34 cycles of denaturation (0.5 min at 95°C), annealing (1 min at 50°C) and extension (1.5 min at 72° C), and a final extension step (20 min at 72°C). Reaction mixtures (total 50 µl) contained 4 µl DNA template (1:10), 26.7 µl H_20, 5 µl *Taq*-buffer S (Peqlab, Erlangen/Germany; including 15 mM $MgCl_2$), 2 µl of each primer, 10 µl dNTPs (each 1.25 mM), and 1.5 units of *Taq*-DNA-polymerase (Peqlab).

PCR products were purified using the QIAquick gel extraction kit (QIAGEN, Hilden/Germany) and labelled using ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kits V 1.0 or 1.1 (Applied Biosystems, Foster City, California/USA) in 30 cycles of denaturation (5 s at 96°C), annealing (15 s at 50°C) and elongation (4 min at 50°C). Extension products were run on ABI Prism 310 and 373XL automated sequencers. Sequences were edited manually with EditView 1.0.1 (Hall 1999). In case of single base length mutations, positions were cross-checked again after their alignment with the original pherograms to ensure correct reads. Gene-spacer and intron-exon boundaries are based on annotations in completely sequenced chloroplast genomes (Tsudzuki *et al*. 1992, Schmitz-Linneweber *et al*. 2001).

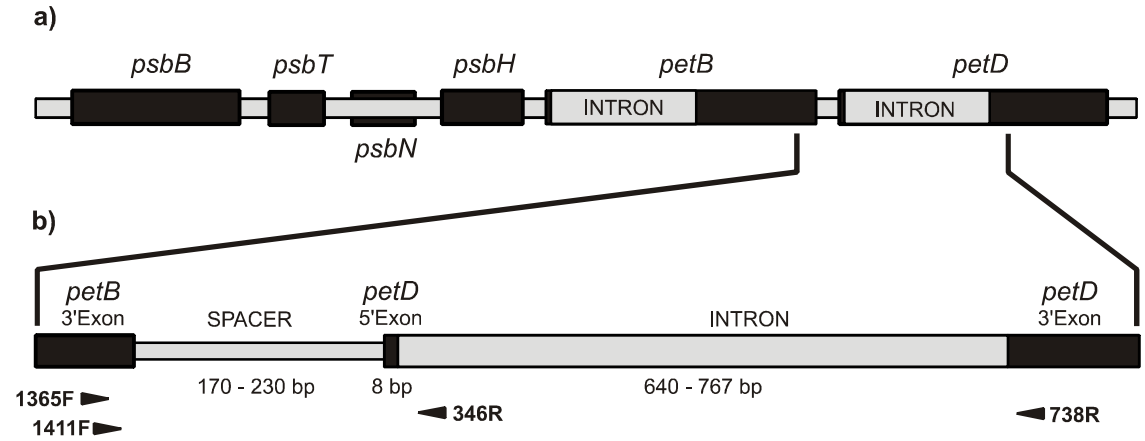


Figure 2.1 — Schematic of the *psbB* operon and the *petD* region studied in basal angiosperms and gymnosperms. Coding stretches are marked by black boxes; non-coding regions are grey. (a) *psbB* operon: the *psbN* gene is not part of this operon but is located on the counterstrand between *psbH* and psbT. (b) Enlarged section of the *psbB* operon including the *petB*-*petD* intergenic spacer, the *petD* 5'-exon (8bp: ATGGGAGT) and the *petD* intron. Minimum and maximum size of spacers and intron among the taxa sequenced are indicated below the bar. Arrows mark the positions of primers.

2.2.5 Sequence alignment

Non-coding regions are characterized by relatively high numbers of length mutations in addition to substitutions. Levinson & Gutman (1987) suggested slipped strand mispairing (SSM) as a mechanism that generates length mutations. SSM might be the underlying process for simple sequence repeats, whereas hairpin structures have been shown to favour inversions (Kelchner & Wendel 1996). For correct primary homology assessment (De Pinna

1991) the molecular processes leading to microstructural changes have to be considered, first pointed out by (Gu & Li 1995). Unfortunately our understanding about the exact mechanisms leading to microstructural changes is still scarce, but the resulting sequence motifs can be observed. Although recent years show considerable progress in automated multiple sequence alignment methods, leading from global Needleman-Wunsch scoring schemes to local procedures (e.g., Dialign2, Morgenstern 1999), these new programs still do not allow recognition of motives such as repeats or inversions. To determine such motifs sequences were carefully examined by eye and aligned manually using QuickAlign 1.5.6 (Müller & Müller 2003). Rules for alignment were proposed on the basis of known mechanisms of sequence evolution and similarity-based criteria for homology assessment (Golenberg *et al*. 1993, Kelchner & Clark 1997, Hoot & Douglas 1998, Graham *et al*. 2000, Kelchner 2000, Simmons & Ochoterena 2000, Borsch *et al*. 2003). In the following we make a few additions to the rules as presented by Borsch *et al*. (2003), which became apparent considering sequence variability patterns observed in this dataset. The alignment can be obtained from author upon request.

(1) **Gap insertion**. In order to minimize length mutational events globally in the affected partition a gap was inserted only if it prevented the inclusion of more than two substitutions among closely adjacent nucleotides.

(2) For **gap placement** recognition of sequence motifs was given priority (Figure 2.2, a). Based on the sequence variability found in this dataset, motif recognition is extended to multiple repeats. These are understood as multiple events. A prerequisite of course is that primary sequences are sufficiently complex to allow unambiguous motif recognition (Figure 2.2, b).

(3) **Homonucleotide strings** strictly involving only one kind of nucleotide (microsatellites) occurred in different positions. Formally, homonucleotide strings can be considered as stepwise indels (overlapping indels), and can be aligned following a parsimony principle of individual steps as outlined in rule 6. However, the probability of reversal or parallel length mutations may be quite high depending on the number and size of repeat units. (Borsch *et al*. 2003) preferred not to align such microsatellites because prevalence of single nucleotides hinders motif recognition. Since microsatellites were frequent in this *petD* dataset, and no substitutions were involved leading to false phylogenetic signal, we treated them as overlapping indels for practical reasons.

(4) **Entire indels**, i.e. indels of the same positional extension occurring in several taxa, were placed in the same column(s). In cases where sequence composition adjacent to an entire indel was not sufficient to restrict the placement of this indel to only a single position we followed the argumentation of (Simmons & Ochoterena 2000) and placed the gap in the same position (same column) in all sequences, since it would be most parsimonious to assume a single event in all taxa (Figure 2.2, c).

(5) **Substitutions in indels** occurring only in one copy, template or repeat were excluded from phylogenetic analysis by introduction of ambiguity codes (Figure 2.2, d).

(6) In the case of **overlapping indels** a parsimony principle was employed to arrange gaps in a way that globally requires the least number of length mutational events.

(7) **Regions of uncertain homology** were excluded from analysis (hotspots sensu Borsch *et al.* 2003).

(8) **Inversions** in the *petB-petD* spacer were reverse complemented in the alignment whenever recognized (Figure 2.2, d). Left unchanged, this would result in substitutions giving false signal (Kelchner 2000, Quandt *et al.* 2003).

2.2.6 Coding of length mutational events

Several workers recently have developed methods to code information from length mutational events and to utilize them in phylogeny reconstruction (Graham & Olmstead 2000b, Graham *et al.* 2000, Kelchner 2000, Simmons & Ochoterena 2000). Simmons & Ochoterena (2000) proposed two kinds of formalized coding strategies, both recognizing gaps as characters. We basically used the simple indel coding approach, thereby assigning character state 1 when sequence was present in the respective taxon, and character state 0 if there was a gap. However, we found patterns of microstructural changes present in our dataset that apparently were not covered by the existing simple indel coding rules. These limitations stem from strictly focusing on gaps as characters, rather than considering any microstructural change as a character. The following additional rules were thus employed:

(1) **Inversions** were coded as a single binary character (state 1 = present, 0 = lacking; Figure 2.5).

(2) **Independent adjacent gaps** need to be distinguished as different characters. As a consequence of alignment principle 2 (recognition of sequence motifs has priority) situations of adjacent but different indels occur (see indels 64/65 and 66/67 in Figure 2.5, b). Such situations seem to be frequent in non-coding cpDNA and probably intron sequences in general, where structural conditions facilitate the insertion of short repeats. These SSRs appear in the same DNA region but are independent events involving different motifs. Therefore, gaps may appear adjacent although they are not resulting from a single event and thus cannot be interpreted as one character. This fact can be illustrated because repeat events result in palindromic motifs, so that there is no way to distinguish which of the nucleotides actually are the template and which the repeat. Therefore, the columns could be placed freely on either side of the motif, placing gaps adjacent or non-adjacent, respectively. To include an addition in the alignment rules that governs placement of columns does not remedy the problem, because SSRs might be frequent leading to adjacent gaps regardless of their placement.

(3) **Multiple repeats** within a given sequence were coded as separate indels. This is a consequence of extended alignment principle 2.

(4) **Length mutations within homonucleotide strings** were not coded. For microsatellites increased rates of length mutational events can be expected that might lead to high levels of homoplasy, in particular when including sequences from a broad spectrum of plants. (Lutzoni *et al.* 2000) suggested using multiple states for length variable homonucleotide strands. We omitted these indels in a more conservative approach because current understanding of the evolutionary processes involved at these sites is still very limited.

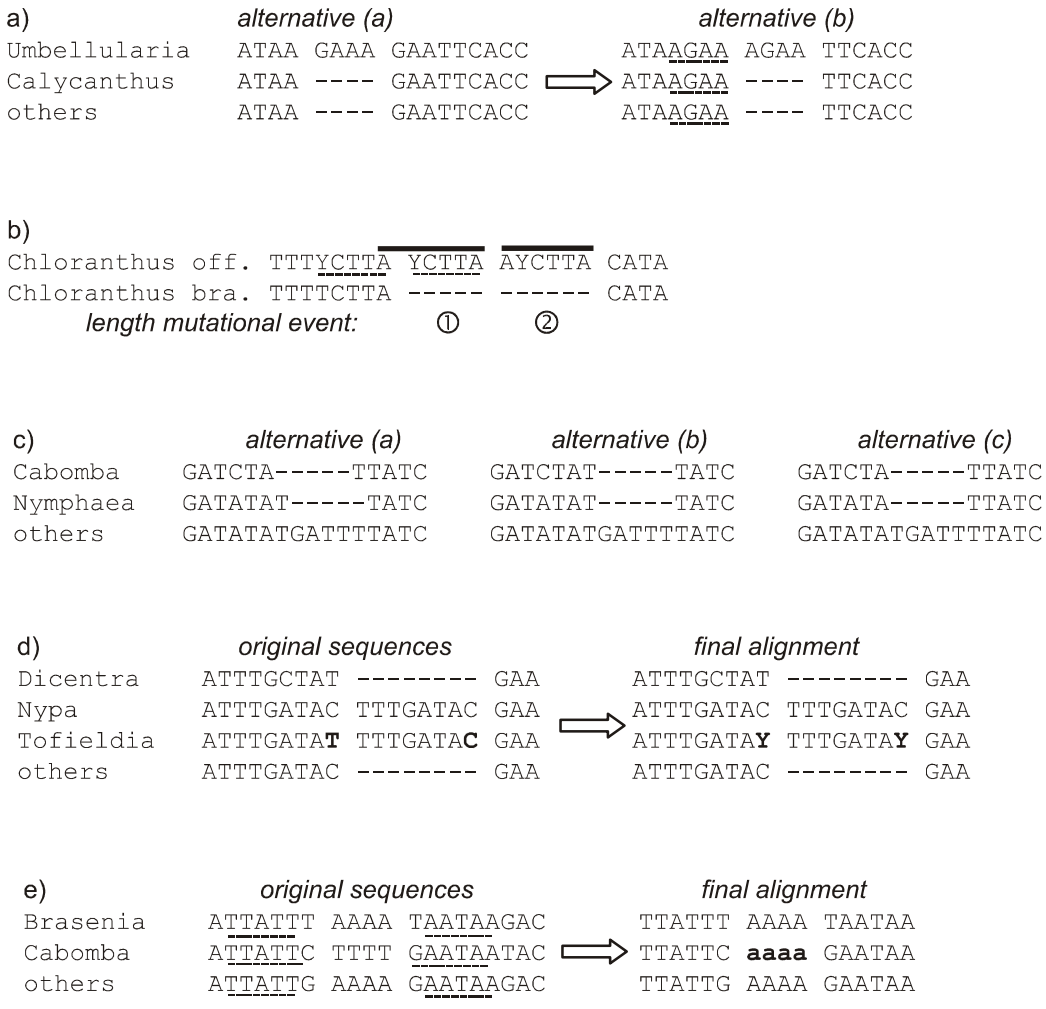


Figure 2.2 — Illustration of the alignment principles applied in this study. (a) *Gap placement* (principle 2). Primary homology assessment would allow alternative positions of the gap (a and b), but according to principle 2 alternative (b) was given priority because it reflects a possible simple sequence repeat (SSR; indel 241, positions 1792 to 1795). **(b)** *Multiple repeats* (principle 2) were recognized as independent events. Repeat motifs are marked by dashed (1st indel) or full lines (2nd indel) (indels 190 and 191, positions 1505 to 1515). **(c)** *Entire indels* (principle 4): If sequence composition adjacent to an entire indel was not sufficient to determine the position of the indel, gaps were placed arbitrarily. If such gaps were shared by two or more sequences, they were placed in the same column. In this example, one of alternative (b) or (c) would be chosen arbitrarily (indel 77, positions 564 to 568). **(d)** *Substitution in indels* (principle 5): If substitutions in either repeat or template occurred, they were replaced by ambiguity codes (indel 64, positions 424 to 431). **(e)** *Inversions* (principle 8) were reverse complemented (indel 41, positions 282 to 285).

2.2.7 Secondary structure

The large-scale study on group II introns by (Michel *et al.* 1989) has provided a secondary structure model that is nowadays widely accepted. For the purpose of this study the calculation of secondary structure appeared to be unnecessary because group II intron core structures are highly conserved, and visual examination of sequences allowed the recognition of reverse complement regions and the demarcation of domain boundaries and structural elements. This was facilitated by the *petD* intron sequences of maize, tobacco, spinach and the liverwort *Marchantia polymorpha*, already examined by (Michel *et al.* 1989). Classification of elements such as stems, loops, bulges and interhelical sequences followed (Vawter & Brown 1993), modified by (Kelchner 2002). *Stems* are helices formed by complementarily pairing nucleotides (including G-U wobble-pairs where they were not terminating a helix). Single stranded nucleotide stretches terminating a helix are termed *loops*, whereas unpaired nucleotides within stems are *bulges*. *Interhelical sequences* are those single-stranded nucleotides connecting helices of adjacent domains and subdomains.

2.2.8 Phylogenetic analysis

For phylogeny reconstruction following data partitions were analyzed: intron sequences alone (= intron matrix), intron and spacer sequences combined (= intron + spacer matrix), all indels alone (= indel matrix), intron sequences plus respective indels (= intron + indel matrix), intron and spacer sequences plus all indels (= combined matrix). Furthermore, all characters of the intron matrix that were assigned to stems and non-pairing elements were analyzed as separate partitions. All characters were equally weighted and gaps were treated as missing characters. Before combining individual matrices incongruence-length difference tests were performed in 1000 random addition replicates using PAUP* 4.0b10 (Swofford 2002). Parsimony analysis (MP) with PAUP* 4.0b10 employed heuristic searches with 1000 replicates of random addition and TBR branch swapping. For small matrices the limit of trees saved was set to 10,000. Measures of support for each node were obtained through bootstrapping (BS) 500 replicates (each with 10 random addition replicates) using PAUP* 4.0b10, and Bremer support (BrS) analysis using PRAP (Müller 2004, 10 random addition replicates per constraint tree, parsimony ratchet not employed).

Bayesian Inference (BI) of the substitution based matrices (intron matrix, intron + spacer matrix) was performed using MrBayes 2.01 (Huelsenbeck & Ronquist 2001). Following the Akaike Information Criterion in Modeltest 3.06 (Posada & Crandall 1998) a GTR+I+G model of molecular evolution was implemented. We conducted 4 runs of Metropolis-coupled Markov Chain Monte Carlo analysis, each with 4 chains of saving one tree every 100 generations for 1,000,000 generations starting with a random tree. The temperature for heating was set at 0.2. After 50,000 generations in the first two runs and 70,000 generations in the third and forth run likelihood scores appeared to be stationary. Consequently, the burn in was set at this generation, sampling only the trees obtained thereafter. GC content and transition : transversion (ti:tv) ratios were calculated using MEGA 2.1 (Kumar *et al.* 2001). Indel and substitution characters were optimized on one of the shortest trees found in the combined dataset using Winclada 1.00.08 (Nixon 2002) assuming accelerated transformation (ACCTRAN).

Table 2.2 — Characteristics of *petB-petD* spacer and *petD* intron sequences in 47 angiosperms and 3 gymnospermous outgroup taxa. Note that high numbers of length mutations in non–coding regions expand the alignment, causing an underestimation of variability. For a better approximation the amount of variability is also calculated on the basis of mean actual length of sequences [corrected]. Similarly, frequency of indels (here calculated as ratio of indel number to alignment length, in %) is calculated also on the basis of mean sequences length [corrected].

	Spacer	Intron	Domain I	Domain II	Domain III	Domain IV	Domain V	Domain VI
position in the alignment	1 - 508	517 - 1810	522 - 1137	1140 - 1272	1279 - 1347	1351 - 1733	1734 - 1767	1772 - 1805
mean sequence length, bp (SD)	200 (11)	708 (37)	365 (10)	60 (4)	46 (2)	147 (33)	34 (0.2)	30 (0.6)
mean sequence length, excluding hot spots (SD)	200 (11)	604 (29)	294 (9)	60 (4)	46 (2)	114 (21)	34 (0.2)	30 (0.6)
number of characters, excluding hot spots	508	1055	449	133	68	311	34	34
number of variable characters	152	406	264	51	26	109	11	17
% variable characters [corrected]	30 [76]	38 [67]	43 [90]	38 [85]	38 [57]	29 [96]	32 [32]	50 [56]
number of parsimony informative characters	117	277	181	32	18	76	6	11
% informative characters [corrected]	23 [59]	26 [46]	29 [62]	24 [53]	27 [39]	20 [67]	18 [18]	32 [37]
number of indels	73	153	57	30	6	73	1	1
% frequency of indels [corrected]	14 [37]	15 [22]	13 [19]	23 [50]	9 [13]	23 [64]	3 [3]	3 [3]
number of parsimony informative indels	21	42	20	12	2	24	1	0
% frequency of informative indels [corrected]	4 [11]	4 [6]	4 [7]	9 [20]	3 [4]	8 [21]	3 [3]	0 [0]
G/C content (SD)	30 (1.8)	39 (1.3)	40 (1.4)	36 (3.7)	39 (2.5)	33 (2.3)	45 (2.5)	45 (3.1)
Ti:Tv ratio (SD)	1.9 (0.6)	2.8 (0.5)	3.0 (0.79)	3.2 (3.8)	3.9 (3.1)	1.5 (1.4)	1.0 (1.6)	2.4 (1.7)

2.3 Results

2.3.1 Overall sequence variability

The *petD* region, including the *petB-petD* intergenic spacer, the *petD* 5′-exon and the *petD* intron, is present in all angiosperms and gymnosperms studied. The overall length ranges from 842 to 979 bp in the taxa sampled, with the spacer accounting for 167– 228 bp, the exon for 8 bp, and the intron for 634–784 bp. The spacer is shortest in the three genera of Austrobaileyales (173 bp, due to a 24 bp deletion) and longest in *Araucaria* (215 bp), while the intron is shortest in Nymphaeales (~ 643 bp, due to a 76 bp deletion in domain IV) and longest in *Impatiens* (785 bp). High numbers of length mutations lead to 1810 characters in the overall sequence alignment (spacer 508, 5′-exon 8, intron 1055) whereas the mean sequence length is 927 nucleotides. Stretches of high sequence variability are confined to 4 distinct mutational hotspots in the intron comprising 13–18% of its total length. These hotspots were mainly caused by strongly deviating sequences of the gymnosperms. The spacer could be aligned and analyzed as a whole without excluding any hotspots. After exclusion of hotspots total length of the aligned sequences was 1582 characters. At the 3´ terminus of the *petB* gene no length variability could be observed and all stop codons share homologous positions.

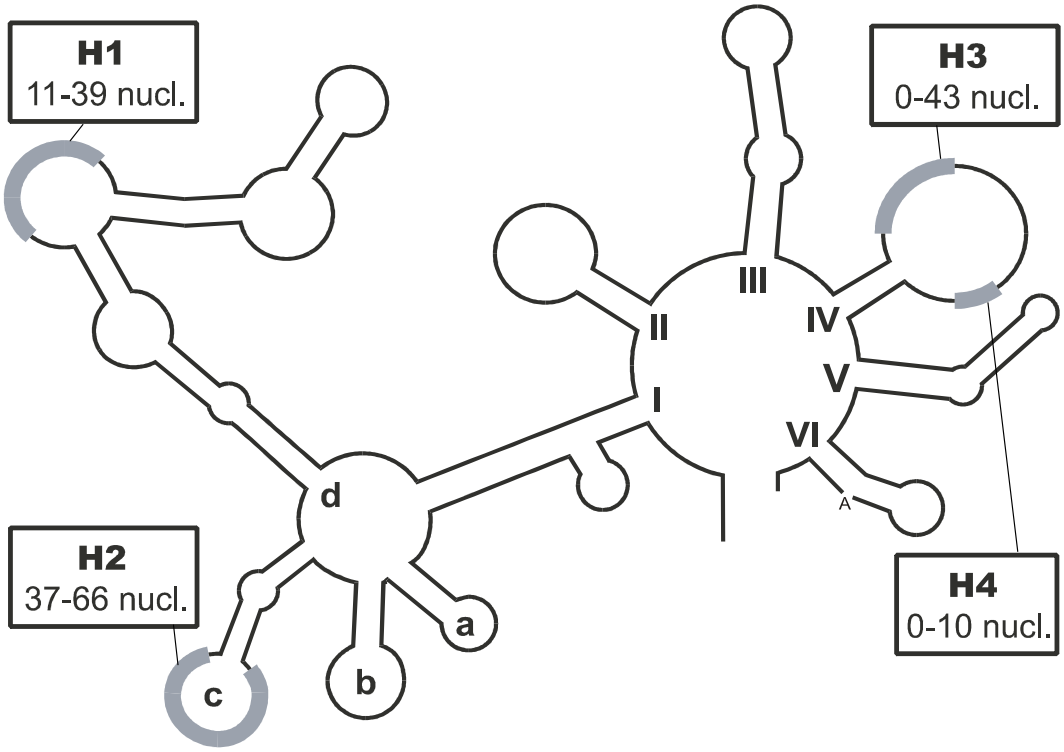


Figure 2.3 — Secondary structure model of group II introns (subgroup B1), modified after Michel *et al*. (1989). Visual identification of reverse complement stretches (stem regions) in the *petD* data set was based on this model. Positions of the four mutational hotspots in *petD* intron sequences are illustrated by grey stripes (HS1 positions 641 to 689, HS2 positions 862 to 974, HS3 positions 1433 to 1494, HS4 positions 1679 to 1688). Mean sequence lengths (and standard deviation, SD) are 19 in H1 (SD = 6), 46 in H2 (SD = 7), 28 in H3 (SD = 13), and 5 in H4 (SD = 2).

2.3.2 Microstructural changes

A total of 241 length mutations (see Appendix 1) were coded in a binary matrix, 73 of which belong to the spacer and 168 to the intron. Most indels are simple sequence repeats (SSRs; 58% in the spacer, 46% in the intron). In 33% of the SSRs, repeat and template differ by one or rarely more substitutions. All other indels are insertions of unknown origin or deletions. Indel length varies between 1 and 76 nucleotides. Indels longer than 10 nt are relatively rare (only 7% of all indels). Among the shorter indels, most of which are SSRs, single-base insertions or deletions are the most frequent size class (25%), followed by 4- and 5-nt-indels (16 and 17%, respectively). As mentioned above highest number of indels is found in domains I and IV, while in domain V and domain VI only one length mutational event could be detected (Table 2.2). Concerning structural partitions, indels are most frequent in loops and less frequent in stems (Table 2.3). Actually, only three indels were found in stem regions, all of them being single–base length mutations. Two inversions could be recognized in the spacer (Figure 2.5), a 4 to 6 bp inversion in *Cabomba* (accurate length cannot be detected due to a palindromic motif) and a 33 nucleotide inversion in *Impatiens*. In both cases the inversions are flanked by short (6 bp) inverted repeat stretches. No inversions were detected in the intron.

Table 2.3 — Characteristics of structural partitions of the *petD* intron sequences. Corrected values are calculated the basis of mean actual length of sequences (see Table 2.2 for further explanation). Loops, bulges, and interhelical nucleotides are summarized as non–pairing elements in the last column to compare with stem nucleotides.

	Stems S	Loops L	Bulges B	Interhelical I	Non–Pairing L + B + I
mean sequence length, bp (SD)	179 (1.9)	362 (34)	74 (4)	33 (0.1)	469 (35)
mean sequence length, excluding hot spots (SD)	179 (1.9)	266 (26)	74 (4)	33 (0.1)	373 (28)
number of characters, excluding hot spots	193	636	84	34	754
number of variable characters	64	232	51	17	411
% variable characters [corrected]	33 [36]	36 [87]	61 [69]	50 [52]	55 [80]
number of parsimony informative characters	40	154	42	12	289
% informative characters [corrected]	21 [22]	24 [58]	50 [57]	35 [36]	38 [56]
number of indels	6	122	22	1	145
frequency of indels	2 [2]	18	11 [12]	3 [3]	17
number of informative indels	0	37	3	0	40
frequency of informative indels	0 [0]	6	4 [4]	0 [0]	5
G/C content (SD)	48.7 (1.0)	31.0 (1.8)	37.6 (4.1)	42.9 (4.6)	33.9 (1.8)
Ti:Tv ratio (SD)	4.12 (1.0)	2.52 (0.9)	4.63 (3.3)	4.65 (6.4)	3.98 (2.9)

2.3.3 Secondary structure

The majority (89%) of all characters could be assigned to one of the four structural classes (stems, loops, bulges, interhelical sequences). Ambiguity in structure assignment was limited to some areas of domain I which is very complex and by far the largest domain, spanning more than half of the intron sequence length (Table 2.2). Domain I contains 33% of all indels in the intron. Domain II is quite simple and comprises a 7 bp stem that is terminated by a loop of 41-62 nucleotides. The 9 bp stem of domain III is interrupted by 2-3 nt bulges and terminated by an 18-34 nt loop. The second largest domain of the *petD* intron is domain IV, consisting of a short stem (5 bp) and a terminal loop varying from 72-197 nucleotides in this dataset. This loop also contains 44% of all indels in the intron. Domain V and VI are the shortest domains. Whereas domain V is built of a long stem of 14 bp (interrupted by a 2 nt bulge) and a small terminal loop of 3 or 4 nucleotides (AAA in *Araucaria* and *Pinus*, AAAA in all other taxa), domain VI consists of a 5 bp stem and a 19-23 nt loop. The branch point A in domain VI, that plays an essential role in the transesterification reaction during the self-splicing process, is present in all taxa.

The number of characters assigned to loop elements is highest in domain IV and lowest in domain V (Figure 2.4). On average, 51% of the nucleotides of the intron were assigned to loop regions (corresponding to 67% of all characters in the matrix), and 25% to stems (corresponding to 15% of all characters). Nucleotides assigned to bulges and interhelical sequences amount to 11% and 5%, respectively (Table 2.3). All four hotspots turned out to correspond to loop regions of the intron (Figure 2.3).

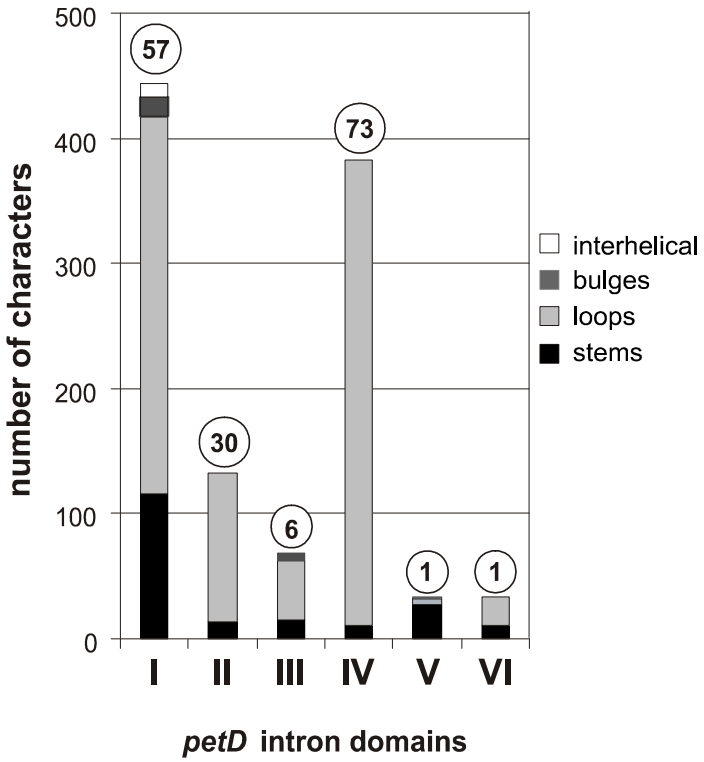


Figure 2.4 — Structure and size of the six group II intron domains in *petD*, illustrated as number of characters per structural element (stems, loops, bulges, and interhelical sequences) in each domain. Circles indicate numbers of length mutations that have been coded in a binary matrix.

Table 2.4 — Results of phylogenetic analyses using different approaches and different data sets. Maximum parsimony analysis (MP) has been applied to sequence data sets containing the intron partition alone or intron + spacer partitions, and to combination of these two data sets with the indel characters. Additionally, Bayesian analysis has been conducted with the intron + spacer data partition. Part a) of the table lists general statistic values of the trees obtained from analyses of the data partitions: intron matrix (substitutions only), intron + spacer matrix (substitutions only), indel matrix (microstructural changes only), intron + indel matrix (substitutions and indels) and the combined matrix (combination of all partitions), part b) summarizes bootstrap and Bremer (in brackets) support values for the maximum parsimony trees and posterior probability values for the trees revealed by Bayesian analysis. Italic numbers represent results of analyses of the *petD* data set reduced by 8 taxa for comparison with results of the trnT-trnF study (Borsch *et al.* 2003, see paragraph 2.2.1 for details).

	MP **intron matrix**	**MP** **intron + spacer matrix**	**MP** **indel matrix**	**MP** **intron + indel matrix**	**MP** **combined matrix**	**Bayesian** **intron + spacer matrix**
a) Tree statistics						
no. of char.	1048 *989*	1550 *1464*	241 *229*	1216 *1146*	1791 *1693*	1566
no. of trees	114 *309*	32 *8*	10,000[1] *10,000*[1]	6 *18*	2 *1*	37600
tree length	1125 *1037*	1565 *1441*	265 *242*	1309 *1204*	1831 *1684*	–
CI	0.5636 *0.5844*	0.5661 *0.5850*	0.8981 *0.9132*	0.6119 *0.6321*	0.6139 *0.6318*	–
RI	0.6475 *0.6152*	0.6498 *0.6179*	0.8866 *0.8833*	0.6710 *0.6407*	0.6752 *0.6447*	–
RC	0.3649 *0.3595*	0.3679 *0.3615*	0.7962 *0.8067*	0.4106 *0.4050*	0.4145 *0.4073*	–
b) Bootstrap values, Bremer support (in brackets), and posterior probabilities (right column)						
angiosperms	100 (49) *100 (49)*	100 (85) *100 (85)*	100 (8) *100 (8)*	100 (56) *100 (55)*	100 (93) *100 (93)*	100
Nymphaeales	100 (14) *100*	100 (16) *100 (16)*	100 (6) *99 (6)*	100 (19) *100*	100 (23) *100 (23)*	100
monocots	88 (5) *88 (5)*	95 (6) *95 (6)*	66 (1) *64 (1)*	94 (7) *95 (7)*	97 (7) *97(7)*	100
eudicots	56 (3) *61 (2)*	62 (6) *90 (5)*	57 (1) *60 (1)*	65 (3) *62 (2)*	96 (7) *96(6)*	100
magnoliids	– –	– –	< 50 (1) *< 50 (1)*	< 50 (1) *< 50 (1)*	59 (2) *63 (2)*	52
Laurales + Magnoliales	– –	– –	< 50 (1) *< 50 (1)*	< 50 (1) *< 50 (1)*	62 (1) *62 (1)*	43
Magnoliales	85 (3) *86 (3)*	90 (5) *95 (5)*	– –	90 (4) *89 (4)*	95 (5) *97 (5)*	100
Piperales + Canellales	– –	– *< 50 (1)*	– –	– –	< 50 (1) *54 (1)*	73
Piperales	< 50 (1) –	90 (5) *84 (5)*	– –	60 (3) *< 50 (2)*	95 (7) *92 (9)*	100

Note [1]: For parsimony analysis of the indel matrix the maximum number of trees retained was set to 10,000.

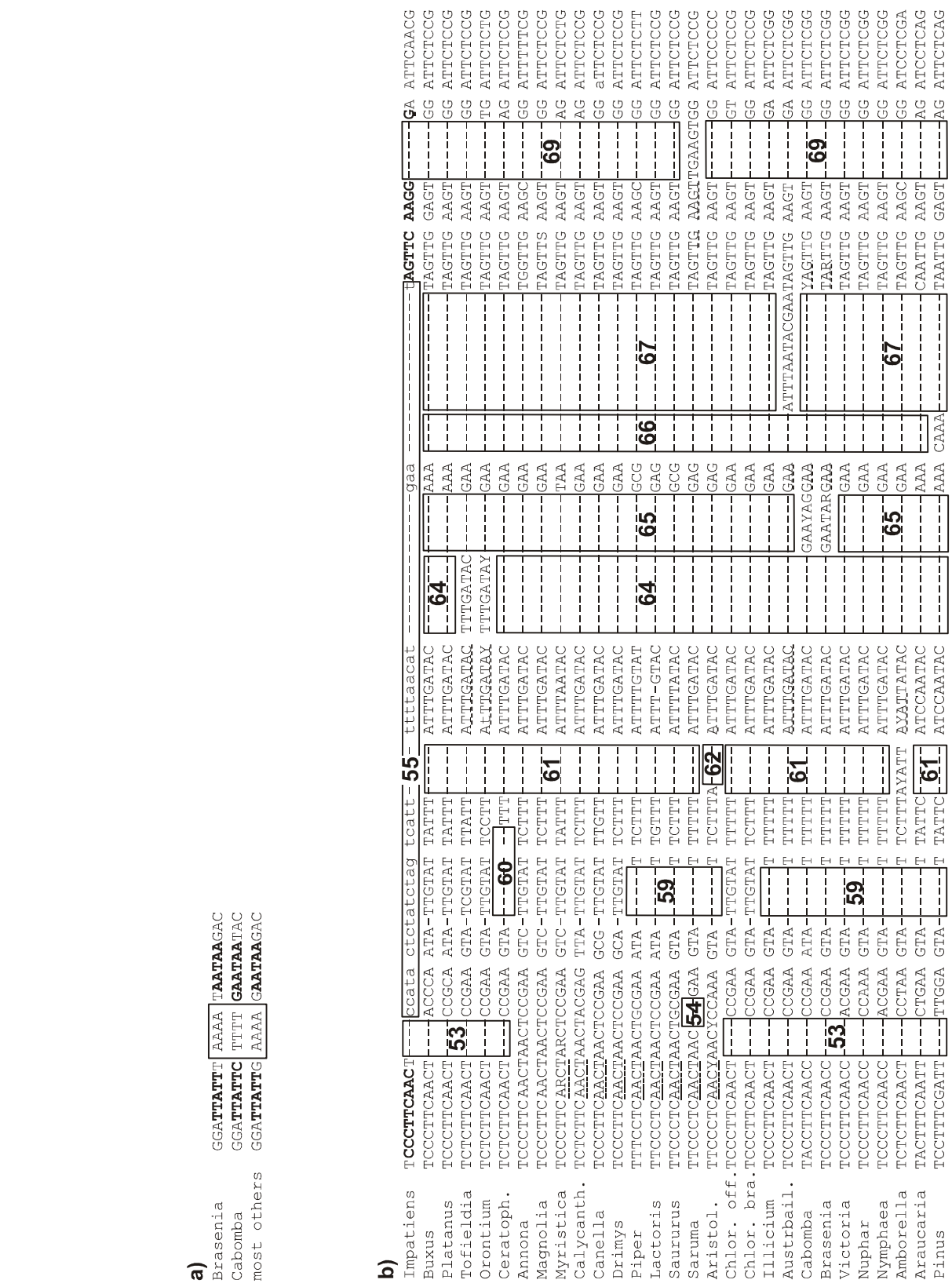


Figure 2.5 — Illustration of observed inversions. (a) Inversion in *Cabomba*, surrounded by a 7-bp inverted repeat (indel 41, positions 282 to 285). **(b)** Inversion in *Impatiens*, surrounded by a 10-bp inverted repeat (indel 55, positions 361 to 483). Indel number 53, a simple sequence repeat (AACT), is shared by all taxa that belong to the magnoliid clade. Substitutions within this indel element occur in *Aristolochia* (AACC), *Myristica* (AGCT), and *Tasmannia* (AATT). The most-parsimonious explanation for the sequence of *Saruma* is a subsequent loss of 3 nt (indel number 54) affecting 1 nt of indel number 53. Indel number 68 is not present in the sequences shown here.

2.3.4 Trees obtained from individual and combined data partitions

Incongruence-length difference tests (substitutions vs. indels, spacer vs. intron) indicated that data partitions are not significantly incongruent (P-values ranging from 0.09 to 1.0), and therefore can be combined for phylogenetic analysis. Table 2.4 gives an overview of the trees obtained from parsimony analyses. Analysis of the intron matrix (first column of Table 2.4) revealed 114 shortest trees with a CI of 0.564 and a RC of 0.365. By combining intron and spacer sequences (second column) the number of trees is reduced to 32, but CI and RC increase only slightly. MP analysis of the indel matrix revealed a lower homoplasy (CI = 0.898, RC = 0.796; see also Figure 2.8). In the combined analyses of indels and substitutions six or two most parsimonious trees were recovered (Figure 2.6; Table 2.3 columns 4,5). CI and RC values were considerably higher than in analyses of substitutions alone (Table 2.4, columns 1 and 2), but lower than in the indel matrix (column 3). The stem partition only comprises 192 characters (non–paring = 800) and only resolves few clades such as Nymphaeaceae and Piperaceae (see Appendix 4, b).

Part B of Table 2.4 summarizes details on topology and support of the trees recovered in different analyses. It becomes evident that major clades like eudicots and monocots are resolved by all data partitions, but other clades are only resolved by indels (e.g., the magnoliids) or by substitutions (e.g., Magnoliales and Piperales), respectively. The strict consensus tree of the combined dataset (column 5) is the best resolved tree (see also Figure 2.6). Furthermore, nodes in most parsimonious trees inferred from the combined matrix gained highest BS and decay values, whereas support was generally lower in trees based on substitutions alone. Bayesian inference of spacer and intron sequence data resulted in a total of 37600 trees sampled. The consensus is fully resolved but shows high posterior probabilities (≥ 95%) largely for terminal nodes at the family or order level, but also for monocots and eudicots (Figure 2.7). A similar trend can be observed for BS values in MP analyses (Figure 2.7).

BI of the intron + spacer matrix and MP of the combined matrix revealed largely congruent topologies. First branching is *Amborella* followed by Nymphaeales (Cabombaceae and Nymphaeaceae) (Figures 2.6, 2.7). Within Nymphaeales there is a clearly supported core Nymphaeaceae (*Nymphaea, Victoria*) whereas *Nuphar* appears in a clade either with Cabombaceae (MP) or core Nymphaeaceae (BI, unsupported). Austrobaileyales are third with BI but resolved sister to magnoliids with MP. However, the latter topology is statistically unfounded. Monocots appear with either *Acorus* (MP) or *Nypa* (BI) branching first. BI indicates affinities of *Ceratophyllum* and Chloranthaceae to monocots, but their positions are not resolved with MP. Magnoliids are generally resolved, and even possess a synapomorphic SSR (Figure 2.6). Support for Magnoliales, Laurales, Canellales and Piperales is particularly high. Within eudicots, *Dicentra* (Ranunculales) comes first, followed by *Platanus* + *Nelumbo*, *Buxus*, *Trochodendron* and terminal core eudicots (*Aextoxicon*, *Impatiens*). Branches leading to *Piper*, Nymphaeales, and among Magnoliales to Annonaceae (Figure 2.7) are particularly long. The same nodes are supported by high numbers of microstructural changes (Figure 2.6).

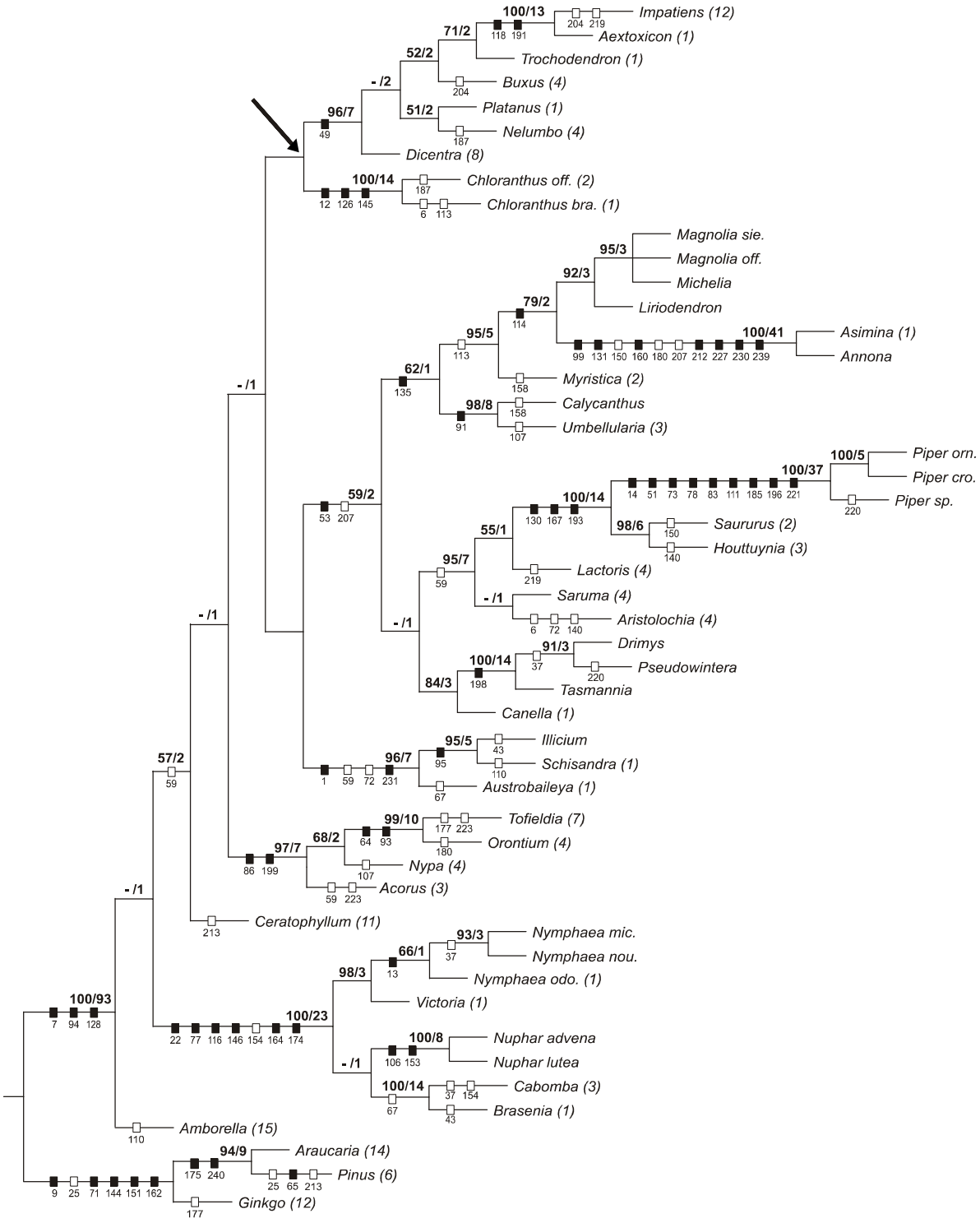


Figure 2.6 — One of two most-parsimonious trees obtained from MP analysis of the combined matrix. Bootstrap values (if higher than 50) and decay values are given above the branches. The arrow indicates the node that is collapsing in the strict consensus. Black boxes indicate synapomorphic indel characters. White boxes indicate homoplastic indels. Uninformative indels are not shown, but number of autapomorphies (if present) is given behind taxon names.

2.4 Discussion

2.4.1 Sequence variability within the *petD* region

Sequences of the *petD* intron and the *petB-petD* spacer are not as variable in their length as the intron and spacers of the *trnT-trnF* region (Borsch *et al.* 2003). Standard deviation (SD) of sequence length is not higher than 5% of the average sequence length in *petD*, but 10% of sequence length in the *trnL* intron, 32% in the *trnT-trnL* and 14% in the *trnT-trnF* spacer. Most of the highly length variable sequence stretches in *trnT-trnF* are located in hotspots. Thus, length deviation was significantly reduced by the exclusion of hotspots. In contrast, exclusion of hotspots in the *petD* dataset did not decrease SD indicating a more even distribution of microstructural changes.

Nevertheless, the matrix of the aligned *petD* intron and spacer sequences was with 1821 characters almost twice as long as the mean sequence length, indicating that length mutations play a major role in the evolution of both intron and spacer sequences. For a meaningful comparison of sequence variability between *petD* and other datasets the increased character numbers (caused by length mutations!) as compared to the nucleotide numbers has to be considered. As a better approximation percentages of variable or potentially parsimony informative characters in relation to mean sequence length instead of character number was suggested (Borsch *et al.* 2003), which were calculated here. The *petD* intron contains a similar number of variable (65%) and potentially informative (24%) characters as the *trnL* intron (63% variable, 19% informative). The *petB-petD* spacer is less variable (76%) than the *trnT-trnL* (83%) and the *trnL-trnF* spacer (98%) but the percentage of informative characters is not decreased (21% in *petB-petD* spacer versus 21 and 18%, respectively). Obviously, the *petB-petD* spacer is relatively conserved when it is compared to the spacers of the *trnT-trnF* region. This might be explained by its functional role in the expression of the gene cluster.

2.4.2 Kind and distribution of length mutations

Analysis of length mutations in the *petD* dataset allowed identification and coding of 219 indels. The resulting indel matrix is one of largest ever compiled (Simmons *et al.* 2001), allowing a thorough analysis of frequency, size distribution and kind of length mutations. Most indels (57%) in the *petD* region are simple sequence repeats. This is congruent with the findings of (Graham *et al.* 2000) in the chloroplast inverted repeat and (Borsch *et al.* 2003) in the *trnT-trnF* region. Among the SSRs most indels result from single-base events or are 4–5 bp long. Indels of 2–3 bp or more than 6 bp are considerably less frequent, what is in line with other non–coding cp DNA regions (Graham *et al.* 2000, Borsch *et al.* 2003)

2.4.3 Structural effects on sequence evolution in the group II intron

Characteristic for group II introns is their subdivision into six domains with domain I being the largest and most complex, whereas domains II to VI are simple stem-loop structures (Michel *et al.* 1989, Michel & Ferat 1995). Therefore, group II introns are ideal systems to compare sequence evolution of different structural partitions and to analyze their effect on

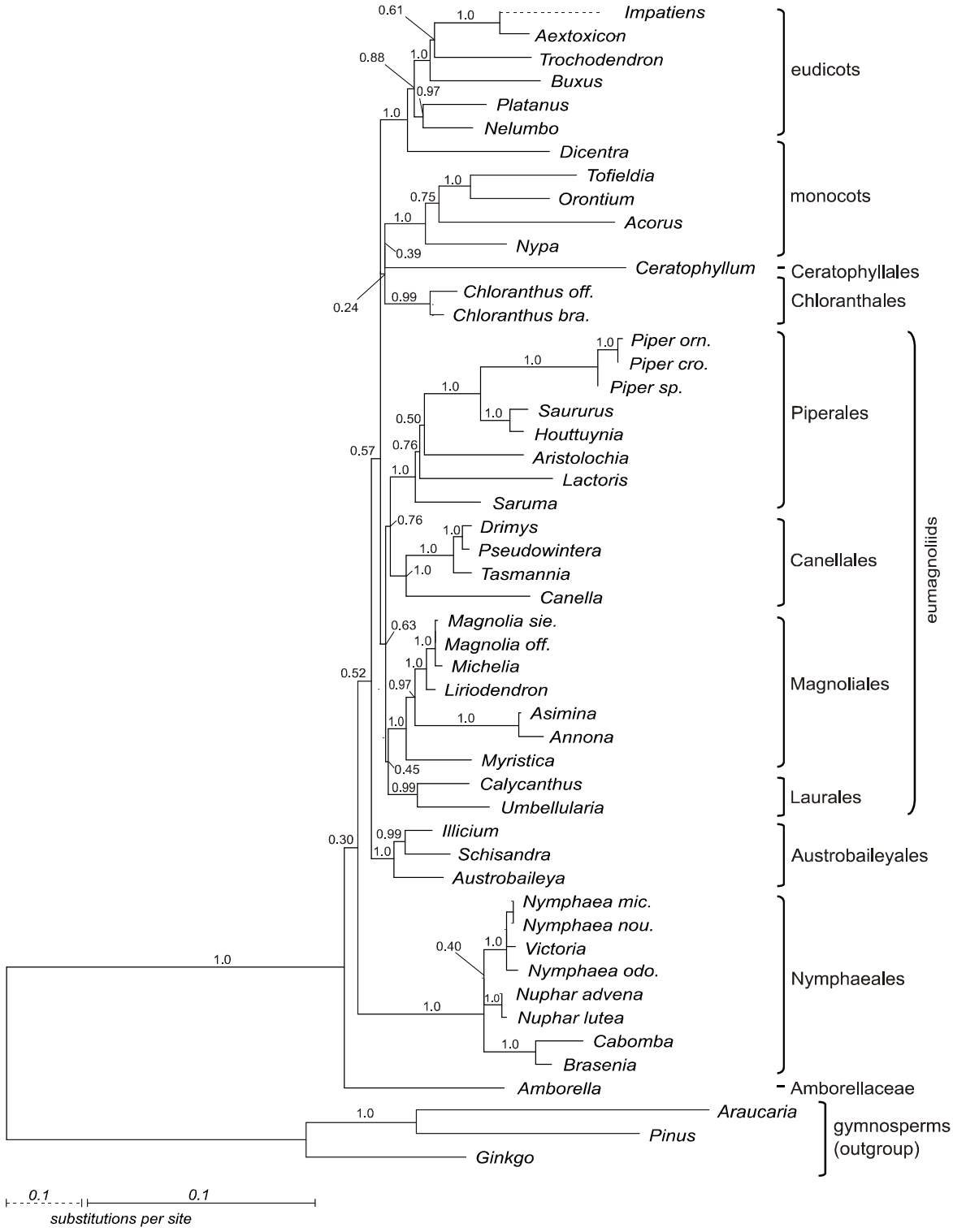


Figure 2.7 — 50% majority-rule consensus of 37,603 trees obtained from four runs of Bayesian analysis implementing the GTR+G+I model. Posterior probabilities are given above the branches. Note that the consensus is fully resolved, but some branches are too short to be visualized; for example, there seems to be a trichotomy in Nymphaeales, although *Nuphar* is placed sister to *Nymphaea*+*Victoria* in the consensus.

phylogeny reconstructions. Similar to the *rpl16* intron in Myoporaceae (Kelchner 2002) almost 90% of sequences could be unequivocally assigned to elements such as stems, loops, bulges and interhelical stretches in the *petD* intron. The comparison of *petD* intron sequences across basal angiosperms shows stem regions to contain fewer variable sites and also fewer potentially informative characters than loops and other non-pairing stretches. The same pattern is found in the *rpl16* intron in Myoporaceae (Kelchner 2002) although distances are generally smaller in this family level dataset. Thus, the group II intron core structure seems to be a governing factor for mutational dynamics regardless of the gene in which the intron is inserted.

Within the *petD* intron, GC content and ti:tv ratios are higher in stems and lower in loops and, corresponding to that, GC content and ti:tv ratios are higher in domains that contain large proportions of stem regions (domains I, V and VI) and lower in domains consisting mainly of loop stretches (Table 2.2). GC rich stems have a ti:tv of 4.12 (SD = 1.0), whereas AT rich loops have a ti:tv of 2.52 (SD = 0.9; Table 2.3). It appears that ti:tv ratios are a function of the respective GC content in line with observations of (Bakker *et al.* 2000) on *trnL-trnF* sequences. This is contrary to prevailing thinking that high ti:tv ratios depend on saturation through multiple transitions (Hillis *et al.* 1993), thereby reflecting their evolutionary distance. Maintenance of secondary structure in GC rich stem regions through compensatory mutations might further favour transitions (Rousset *et al.* 1991, Bakker *et al.* 2000). Obviously, stems evolve under high functional constraints because they are essential for the secondary structure and splicing function of the group II intron. Therefore, stems are less variable regarding nucleotide substitutions as well as length mutations, in line with findings in the mitochondrial *rps3* intron (Laroche & Bousquet 1999). In fact, only three indels have been observed in stems, all of which are 1 bp. Indels >1 bp only occur in non-pairing sequence stretches. In contrast, non-pairing DNA is more AT rich and bears more variable and potentially informative characters as well as length mutations.

The six intron domains are characterized not only by a typical secondary structure but also by their specific function in the splicing reaction (Dib-Hajj *et al.* 1993, Costa *et al.* 2000, Kelchner 2002). Domain VI, for instance, is one of the most important structures in the group II intron. Functional constraints on the evolution of this particular domain are supposed to be high. In fact, the frequency of length mutations in domain VI is considerably lower than in domain I, although both have almost the same amount of loop stretches relative to their total sequence length. Differences in domain conservation imply differences in phylogenetic utility. Less conserved regions (e.g. domains II – IV and loops in general) should provide information for the terminal nodes, whereas more conserved regions (domains V and VI, stems in general) possibly provide information for basal nodes. Since findings on different molecular evolution in domains of the *petD* intron corroborate findings made in *rpl16* (Kelchner 2002) these considerations might be valid for group II introns in general. Within the less conserved domain I c and d2 loops as well as in the domain VI loop of sequences are in part so variable that they could not be aligned unambiguously across seed plants. Similar to the *trnL* intron (Borsch *et al.* 2003) hotspots are strictly confined to loop stretches and do not comprise more than 18% of the whole intron sequence length (20% in *trnL*). Thus, they do not impair the general utility of that kind of non-coding cpDNA for phylogenetic reconstruction on higher taxonomic levels.

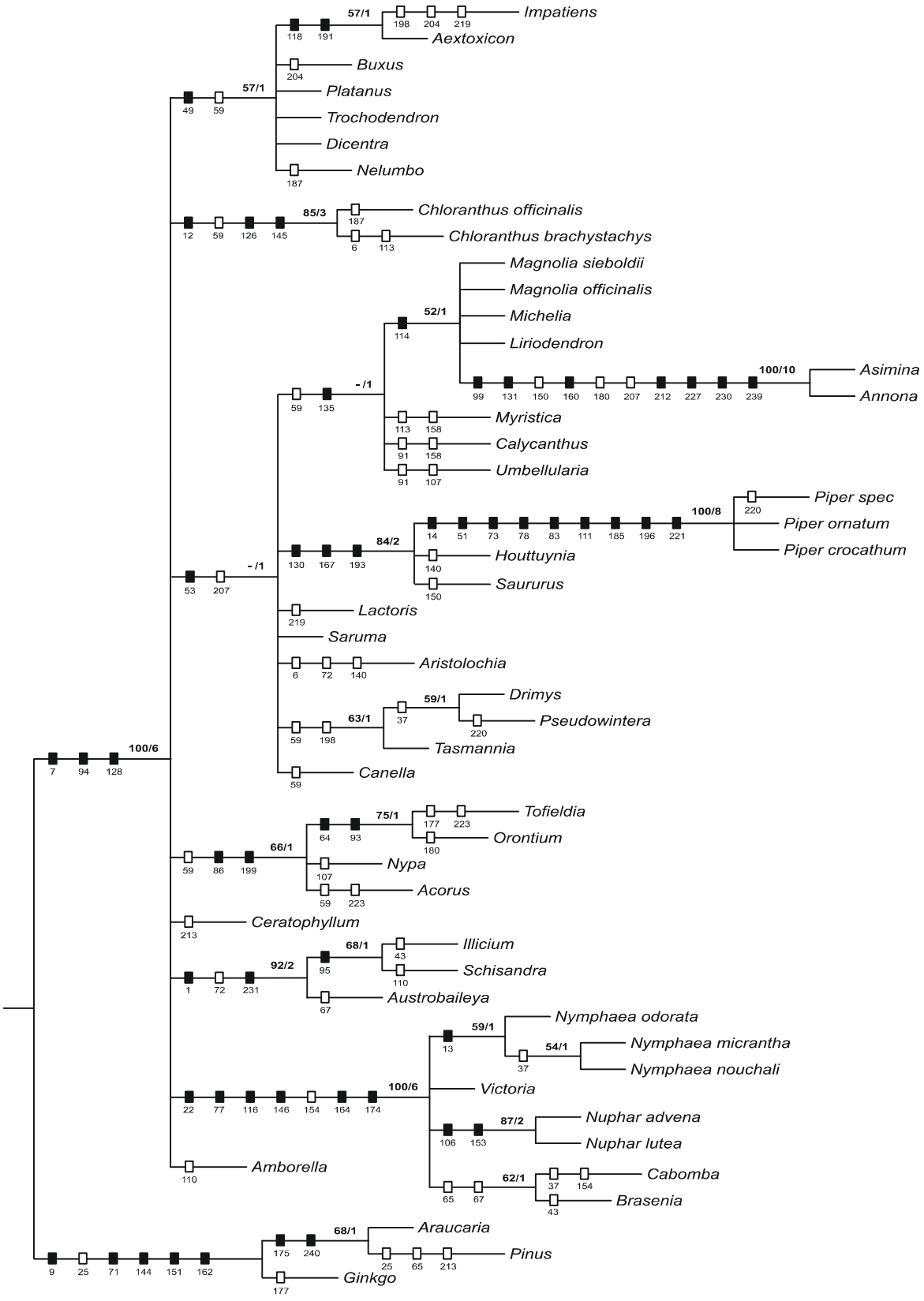


Figure 2.8 — Strict consensus of 10,000 most parsimonious trees obtained from the indel matrix (241 characters). Parsimony informative characters are mapped onto the tree. Black boxes show synapomorphies, white boxes indicate homoplastic indels. Numbers above branches indicate bootstrap and Bremer support values.

2.4.4 Molecular evolution of the spacer

In contrast to the *trnT-trnL* and *trnL-trnF* spacers the *petB-petD* intergenic spacer is less variable. Since the whole *petB-petD* region persists as a dicistronic mRNA after transcription it can be assumed that RNA secondary structure is important for the evolution of the *petB-petD* spacer. Moreover, this spacer is considered to play an important role in translation of the *petD* gene, by containing a sequence motif that allows ribosomes to detect the *petD* initiation codon (Monde *et al*. 2000). However, the secondary structure of the *petB-petD* intergenic spacer has not been analyzed. Further information on sites relevant for translation of the *petD* gene is, however, not available. The two inversions observed in *Cabomba* and *Impatiens* are associated with short inverted repeats indicating the presence of stem loop structures. In other taxa the regions enclosed by these inverted repeats show numerous length mutations and substitutions as well (see Figure 2.5, b).

2.4.5 Phylogenetic signal of *petD* sequence data

Differential analysis of data partitions and the respective tree statistics provide evidence for the high potential of microstructural changes as phylogenetic characters. Overall, indel characters are considerably less homoplastic than substitutions in this *petD* dataset (Table 2.4, columns 2 and 3). This becomes evident when comparing CI and RC values of the indel tree (Figure 2.8) with the trees inferred from substitutions only (indels: CI = 0.895, RC = 0.789; substitutions: CI = 0.566, RC = 0.368). The resolution of the indel tree is even more striking given that only 59 out of 168 indels in the intron are parsimony informative (35%). The spacer yields 21 parsimony informative indels out of 73. Most of the parsimony informative indels (71%) are synapomorphies (CI, RI = 1), whereas 23 out of 79 informative indels were reconstructed to have originated two or more times independently (empty boxes in Figure 2.6). Homoplastic indels in the intron are generally located in loop regions. In the spacer only indel 59 is homoplastic, which is part of a possible loop between inverted repeat stretches (Figure 2.5, b). This confirms the theoretical expectation that structural constraints have effects on the frequency of length mutations.

Indels are relatively rare at the deepest nodes within angiosperms. Nevertheless, indel no. 53, a simple sequence repeat of 4 nucleotides, is synapomorphic for the magnoliid clade (comprising Magnoliales, Laurales, Canellales, and Piperales; Figure 2.5,b). The magnoliids are one of the major, recently recovered angiosperm clades. This clade has also been revealed by substitution-based trees of this and of other non-coding (Borsch *et al*. 2003) and coding datasets (e.g., Qiu *et al*. 1999, Hilu *et al*. 2003) and now is clearly substantiated by indel information. Compared to substitutions, indels may be regarded as independent evidence because microstructural changes result from different mutational processes.

Although *petD* sequences reveal *Amborella* and Nymphaeales as first branching angiosperms, they provide no statistical support for the basal grade (Figures 2.6, 2.7). Short branch lengths (phylogram in Figure 2.7) indicate a possible lack of mutations that were fixed during early divergence of angiosperm lineages. Similarly, there is no support from the indels for these nodes (Figure 2.6, see also Figure 2.8). Plotting substitutions on the tree (Appendix 2 and 3) allowed assessing homoplasy at particular sites, and an assignment of individual characters to structural elements of the group II intron. In total, the backbone of

angiosperms (nodes from *Amborella* to eudicots) is supported by only 28 substitutions among the intron characters, averaging at 4-5 per node. All of these substitutions occur in non-pairing elements (21 in loops, 5 in bulges, 2 in interhelical elements). Synapomorphic states supporting angiosperms above *Amborella* are in character 1509, and for angiosperms above Nymphaeales in character 1014 (exon binding site 1). A substitution in character 1564 is synapomorphic for magnoliids. All other characters variable at the basal nodes are homoplastic, often exhibiting repeated substitutions or reversals within eudicots and monocots. Thus, short branches at the base of the angiosperm tree are predominantly caused by a lack of mutations, which may either be explained by a rapid radiation of these lineages or substitutional rates increasing through time. In addition, the respective variable characters are likely to be homoplastic because they are mostly located in rather freely evolving loops of the intron. Microstructural changes also have accumulated in unpaired elements, whereas stems are structurally conserved and show almost no length variability across angiosperms.

The rates of length mutations present in unpaired elements of *petD* obviously have not led to noticeable homoplasy. The above–mentioned structural conservation of stems extends to their conservation in primary sequence. Across the angiosperm tree, substitutions are very few in the stem partition of *petD* (Appendix 4, b). Contrary to the expectation that stem elements would provide no information for terminal nodes, some terminals are resolved in the stem partition tree, such as Nymphaeales, the genus *Piper* or the Annonaceae. These lineages accumulate both indels and substitutions (see Appendix 5, Fig. 6), which points to a lineage specific acceleration of mutational rates. Higher variability even in stem elements may be caused by possibly relaxed constraints on helical parts of the *petD* intron in these lineages. Accumulation of length mutations may be a trend in the chloroplast genome of Piperaceae and the Nymphaeales because this was also observed in the *trnT-trnF* region (Borsch *et al*. 2003).

Substitutional patterns in stem regions are often biased by compensatory mutations because maintenance of secondary structure is essential for the self-splicing mechanism. Since single events thus lead to double substitutions, signal from respective substitutions might be over–weighted in phylogenetic analysis. Therefore several authors argued for not including such characters into phylogenetic analysis. In the *petD* intron 192 out of 1055 characters belong to stem regions, but out of these characters only 64 are variable (in contrast to 406 variable characters in the whole intron). Thus, a maximum of 16% of all substitutions might be compensatory mutations. This percentage is higher than in the *trnL* intron (7%, Borsch *et al*. 2003) but still considerably lower than in 18S rDNA (73%, Soltis *et al*. 1999b). Compensatory mutations therefore probably play a minor role in sequence evolution of the *petD* intron.

This study has proven the partitioned analysis of structural elements (Appendix 4, a and b) as a valuable tool for unravelling effects of mutational dynamics on phylogenetic signal. However, the present *petD* dataset only contains 192 stem characters, 45 of which are potentially parsimony informative across seed plants. Thus, stem partitions will not have the potential to better resolve parts of the angiosperm tree, simply because of a too small amount of information. More group II intron datasets are needed to refine insights into signal obtained from different partitions.

2.4.6 Phylogeny of basal angiosperms

The trees inferred from the *petD* region are consistent with those based on multi-gene, multi-genome datasets (Qiu *et al.* 1999, Soltis *et al.* 2000, Zanis *et al.* 2002), and with trees obtained from other rapidly evolving genomic regions like *trnT-trnF* (Borsch *et al.* 2003) and *matK* (Hilu *et al.* 2003). *PetD* further substantiates *Amborella* as sister to all other angiosperms, although a contrary hypothesis on the root of angiosperms has recently been proposed (Goremykin *et al.* 2003).

Within Nymphaeales three clades are clearly supported: (1) Cabombaceae, (2) the genus *Nuphar*, and (3) a clade consisting of *Nymphaea* + *Victoria*. However, the monophyly of Nymphaeaceae as comprising *Nuphar*, *Nymphaea* and *Victoria* is not substantiated by the *petD* dataset. Actually, *Nuphar* is found either sister to Cabombaceae (MP, Fig. 6) or first branching in Nymphaeaceae (BI, Fig. 7). Further studies are needed to clarify relationships, since resolution and support among the three major Nymphaeales-lineages are also low in other studies (e.g., Borsch et al. 2003, Soltis et al. 2000, Qiu et al. 1999, Zanis et al. 2002), or monophyly of Nymphaeaceae s.str. was assumed a priori (Les et al. 1999).

The magnoliids consisting of Laurales + Magnoliales and Piperales + Canellales are consistently resolved in *petD* analyses. This underscores the utility of rapidly evolving regions including *petD*, which so far all unravelled this clade (Borsch *et al.* 2003, Hilu *et al.* 2003). Contrary to trees based on single, slowly evolving genes, magnoliids were otherwise only inferred by multi-gene analyses combining all three genomic compartments (Qiu *et al.* 1999). Within magnoliids results of this *petD* dataset are particularly relevant as BI and MP infer Piperales and Canellales as sisters. In all previous analyses there has been medium or only low statistical support for this sister group, and BI of *matK* sequences even found Piperales first branching in magnoliids (Hilu *et al.* 2003). Thus, molecular evidence seems to converge upon close relationships between Piperales and Canellales, contrary to phylogenetic analyses of morphological characters (Doyle & Endress 2000). Within Magnoliales the *petD* dataset provides high support and a topology that mirrors exactly the conclusions drawn in a recent analysis by Sauquet *et al.* (2003) using molecules and morphology. The respective positions of *Ceratophyllum* and Chloranthaceae as well as monocots and eudicots could not be clarified here but given the small amount of characters in *petD* (half of the trnT-trnF dataset, Borsch *et al.* 2003) this will have to await future combined analyses. Within eudicots the topology is fully congruent with datasets containing large numbers of taxa (Soltis *et al.* 2000, Hilu *et al.* 2003) suggesting that *petD* will provide valuable information for further resolving eudicot relationships.

2.5 Conclusion

Group II introns such as in *petD* can be aligned across seed plants. Similar to observations made with the *trnT-trnF* region high length variability is confined to mutational hotspots located in certain loops in their secondary structure. Overall, group II intron sequences as analyzed here for the *petD* intron are less variable compared to group I introns, which corresponds to the conserved group II intron secondary structure involving large proportions of helical elements. This *petD* intron dataset shows that mutational dynamics in helical

elements clearly differs from unpaired stretches. The combination of sequences from several group II introns will possibly serve as a model for contrasting molecular evolution and phylogenetic signal in different data partitions. Group II introns seem to reflect a mosaic pattern of very conserved stems and rather freely evolving unpaired elements. Mutational patterns and rates in *petD* loops, bulges an interhelical stretches obviously produce variability that resolves major clades of angiosperms such as magnoliids and eudicots as well as relationships within them. Spacers and group I introns may offer a wider spectrum of differently evolving sites as compared to rather uniform partitions in group II introns. At the level of angiosperms this seems to make the accumulation of information for nodes of the basal grade more likely. The latter fact might explain the even better performance of the *trnT-trnF* region for resolving an identical set of 42 basal angiosperm taxa (Borsch *et al*. 2003). Moreover, this study shows the potential of microstructural changes as phylogenetic markers. Indel information corroborates signal inferred from substitutions and leads to increased resolution and support of the recovered trees. Considering that levels of homoplasy are even lower compared to substitutions, a high indel frequency in introns and spacers underscores their effectiveness in phylogenetic reconstruction.

Chapter 3

Phylogenetic analysis of Nymphaeales using fast–evolving and non–coding chloroplast markers

Contents

3.1 Introduction

The basal angiosperm order of Nymphaeales (water lilies and their relatives) comprises approximately 70 aquatic species occurring in freshwater habitats all over the world. Generally, two families are recognized: Cabombaceae and Nymphaeaceae – although these have been combined into a broadly defined Nymphaeaceae, emphasizing their common descent but ignoring their substantial differences, by APG (2003). The Cabombaceae comprise the mostly neotropical genus *Cabomba* (5 species, with *C. caroliniana* reaching temperate North America) and the widespread monotypic genus *Brasenia* (e.g., Orgaard 1991, Williamson & Schneider 1993). The Nymphaeaceae (Schneider & Williamson 1993) consist of six genera: the monotypic *Euryale* (East Asia) and *Ondinea* (NW Australia), the neotropical *Victoria* (2 species), the Southeast Asian *Barclaya* (4 species), the north–temperate *Nuphar* (8 species in two sections), and *Nymphaea*, the largest and most cosmopolitan genus (47 species). Traditionally, *Nymphaea* is subdivided into five subgenera (Caspary 1891, Conard 1905): the Papuan-Australian subg. *Anecphya*, the neotropical subg. *Hydrocallis*, the palaeotropical subg. *Lotos*, the pantropical subg. *Brachyceras*, and the north–temperate subg. *Nymphaea*.

In earlier taxonomic treatments based on morphology or anatomy (e.g., Thorne 1976, Tamura 1982, Ito 1987, Cronquist 1988) circumscription of the order Nymphaeales also included the genera *Nelumbo* and *Ceratophyllum*. However, broad scale analyses of molecular markers (e.g, Chase *et al.* 1993, Savolainen *et al.* 2000b) justified the exclusion of *Nelumbo* and *Ceratophyllum* and substantiated the monophyly of Nymphaeales in the sense described above. The re–evaluation of morphological characters showed the presence of certain states such as tricolpate pollen (Nandi *et al.* 1998) or epicuticular wax tubules mainly composed of nonacosan-10-ol (Barthlott *et al.* 1996b) in *Nelumbo* and further substantiated its exclusion from Nymphaeales. As already noted contemporary treatments of Nymphaeales favour recognition of two families (Cabombaceae and Nymphaeaceae), although further families, such as Barclayaceae (Li 1955, Cronquist 1988), Euryalaceae (Li 1955), or Nupharaceae (Kerner von Marilaun 1891, Nakai 1943, Takhtajan 1997) have been suggested (see Les *et al.* 1991 for review of taxonomic history).

Molecular and morphological data generally indicate a close relationship of *Cabomba* and *Brasenia*, thus substantiating the monophyly of the family Cabombaceae (see Williamson & Schneider 1993). By contrast, the monophyly of the family Nymphaeaceae does not gain much support in phylogenetic analyses. The *rbcL* study of Les *et al.* (1991) yielded 50% bootstrap support and the more recent parsimony and Bayesian analyses of *trnT-trnF* sequence data by Borsch *et al.* (subm.) resulted in 77% JK support and 0.63 posterior probability, respectively. A recent analysis of basal angiosperm relationships (Löhne & Borsch 2005) using *petD* group II intron sequences found three major clades within Nymphaeales, i.e. Cabombaceae, *Nuphar*, and the remaining Nymphaeaceae. These changed their relative positions depending on combination of data partitions and tree inference method. Other studies of basal angiosperms show Nymphaeaceae as a monophylum but with low support (75 % bootstrap support in Borsch *et al.* 2003), or do not even resolve generic relationships (rbcL + atpB + 18S, Soltis *et al.* 2000). Analyses combining data from all three genomic compartments (e.g., Qiu *et al.* 1999, Zanis *et al.* 2002, Qiu *et al.* 2005) are usually short in taxon sampling, and comprise only one species each of

Nymphaea, *Nuphar*, *Cabomba* and *Brasenia*. The hitherto most extensive phylogenetic analysis of Nymphaeales was carried out by Les *et al*. (1999) and was based on *rbcL*, *matK*, 18S rDNA, as well as a morphological matrix. It included one representative of each of the eight genera of Nymphaeales. However, Nymphaeaceae were rooted with Cabombaceae, thus assuming the monophyly of Nymphaeaceae a priori. Most recently, nuclear ITS sequences were employed to reconstruct the phylogeny of Nymphaeales (Liu *et al*. 2005, Podoplelova & Ryzhakov 2005) but could not clarify relationships, mainly due to limitations in taxon and character sampling. Uncertainties also remain concerning the relationships within Nymphaeaceae. Based on the analysis by Les *et al*. (1999) relationships of Nymphaeaceae genera can be hypothesized as: *Barclaya* (*Ondinea* (*Nymphaea* (*Euryale*, *Victoria*))). However, the sister–group relationship of *Ondinea* to a clade consisting of *Nymphaea* and *Victoria* plus *Euryale* was not convincingly supported by the molecular data (51% bootstrap support), with the node being based on three morphological characters only (Bremer support = 3). A very recent molecular phylogenetic study involving non–coding *trnT–trnF* sequences from a much broader sampling of *Nymphaea* species (Borsch *et al*. subm.) indeed resolved *Ondinea* as part of a *Nymphaea* subg. *Anecphya* clade. Furthermore, Borsch *et al*. (subm.) could not even find convincing support for the monophyly of *Nymphaea* with respect to *Victoria* and *Euryale* (67% bootstrap, 0.85 posterior probability). The species of *Nymphaea* were consistently found in three well-supported clades: a clade consisting of *N*. subg. *Anecphya* and *N*. subg. *Brachyceras*, another clade consisting of *N*. subg. *Hydrocallis* and *N*. subg. *Lotos*, and a third clade of *N*. subg. *Nymphaea* that appeared sister to the other two. The *trnT–trnF* data set of Borsch *et al*. (subm.) in addition indicated that the pantropical *Nymphaea* subg. *Brachyceras* might be paraphyletic to the Australian subg. *Anecphya*.

In view of these recent results, the apparent effects of taxon sampling, and considering the still remaining questions on phylogenetic relationships in Nymphaeales, a comprehensive study was designed using a multi–gene data set and a more representative taxon sampling covering all genera of Nymphaeales. In this study, each subgenus of *Nymphaea* is represented by at least two species that are most distant in the *trnT–trnF* tree of Borsch *et al*. (subm.). Several fast–evolving regions of the chloroplast genome are employed to address the phylogeny of Nymphaeales. The present data set comprises a wide spectrum of structurally different markers: a group I intron (*trnL*), group II introns (*petD*, *rpl16*, *trnK*), spacers (*petB-petD*, *trnK-psbA*, *trnT-trnL*, *trnL-trnF*), and the *matK* gene. Phylogenetic utility of these fast–evolving regions has been substantiated in numerous studies for the *rpl16* intron (e.g., Kelchner & Clark 1997, Renner & Chanderbali 2000, Zhang 2000), for the *trnK-matK* region (e.g., Steele & Vilgalys 1994, Johnson & Soltis 1995, Hilu & Liang 1997, Hilu *et al*. 2003), and for the *trnT–trnF* region (e.g., Asmussen & Chase 2001, Sauquet *et al*. 2003, Neinhuis *et al*. 2005). Compared to rather slowly evolving genomic regions of conserved genes, rapidly evolving regions provide higher percentages of variable and informative characters because of higher rates of substitutions. A recent study of Borsch *et al*. (2003) on *trnT–trnF* indicated that non–coding regions like spacers and group I introns are not particularly prone to saturation and homoplasy. In fact, problems of homology assessment and alignment in data sets comprising sequences of a broad taxonomic range are confined to certain mutational hotspots. Moreover, it has been shown that microstructural changes, which are common in non–coding regions, provide valuable additional information for

phylogenetic inference (Simmons *et al.* 2001, Hamilton *et al.* 2003, Löhne & Borsch 2005, Müller 2006). Based on these results, the group II intron in *petD* was developed as an additional non–coding chloroplast marker for analyses at deeper taxonomic levels, and its high efficiency for phylogeny inference could be proved in a case study in basal angiosperms (Chapter 2). The *petD* intron, in combination with a set of structurally different, fast–evolving regions seemed to be a promising tool for elucidating the phylogeny and evolution of Nymphaeales.

For the present study there are three principal objectives. First, the aim is to generate a well-resolved and sufficiently supported chloroplast phylogeny of Nymphaeales based on an extensive taxon and character sampling. Second, the monophyly of Nymphaeaceae as well as the monophyly of the genus *Nymphaea* shall be evaluated. A third intention is to assess the phylogenetic utility of the structurally different chloroplast genomic regions used in this study. Ultimately, it will be necessary to compare trees obtained from different genomes, for example to address putative effects arising from ancient hybridization, thus well–supported trees for all three compartments will be needed.

3.2 Materials and Methods

3.2.1 Taxon sampling and plant material

The data set used in this study comprises 24 species of Nymphaeales, representing both genera of the Cabombaceae (*Brasenia*, *Cabomba*), each genus of the Nymphaeaceae (*Barclaya*, *Euryale*, *Nuphar*, *Nymphaea*, *Ondinea*, *Victoria*), and within the genus *Nymphaea* each of the five subgenera (*Anecphya*, *Brachyceras*, *Hydrocallis*, *Lotos*, *Nymphaea*) as well as both sections of *Nuphar* (*Astylus* and *Nuphar*). Most of the sequence data was generated during this thesis, with only a few sequences, especially for *trnT–trnF*, taken from previous studies (Borsch *et al.* 2003, but see also Chapter 2 and Löhne & Borsch 2005). All sequences of *Nymphaea alba* were taken from GenBank (AJ627251, Goremykin *et al.* 2004).

In recent molecular studies on angiosperm relationships Nymphaeales were inferred to be part of the basal grade of angiosperms, with either the New Caledonian shrub *Amborella* being the most basal angiosperm followed by Nymphaeales and Austrobaileyales as successive sister lineages (e.g., Parkinson *et al.* 1999, Mathews & Donoghue 2000, Soltis *et al.* 2000, Borsch *et al.* 2003, Hilu *et al.* 2003, Löhne & Borsch 2005) or with *Amborella* + Nymphaeales forming a clade sister to Austrobaileyales and all remaining angiosperms (e.g., Barkman *et al.* 2000, Zanis *et al.* 2002, Leebens-Mack *et al.* 2005). Although it is not possible at the moment to confirm one of the two hypotheses (Stefanovic *et al.* 2004, Leebens-Mack *et al.* 2005, Qiu *et al.* 2005) *Amborella* and Austrobaileyales can be regarded as the nearest branches to Nymphaeales in the tree of angiosperms. Thus, sequences of *Amborella trichopoda* (Amborellaceae) and four representatives of Austrobaileyales (*Austrobaileya*, *Illicium*, *Kadsura*, *Schisandra*) are included as outgroup taxa in the present study. Information on all investigated species, on the origin of material, as well as on deposited vouchers and GenBank accessions are summarized in Table 3.1.

Table 3.1 — Taxa used in the present study on Nymphaeales, their respective families, the source of material and location of voucher specimens. GenBank accession numbers of deposited sequences are given for those sequences that have already been submitted. All other sequence data (marked as "to be subm." in the table) will be submitted to GenBank for a publication based on this chapter of the thesis; accession numbers can be obtained from the author upon request.

Species	Family	Origin	Voucher	GenBank accession numbers			
				trnT-trnF	*trnK / matK*	*petD*	*rpl16*
Amborella trichopoda Baill.	AMB	University of California, Sta. Catarina BG	Borsch 3480 (VPI)	AY145324	DQ185522	AY590876	to be subm.
Austrobaileya scandens C. White	AUS	Bonn BG	Borsch 3464 (BONN)	AY145326	DQ185523	AY590867	to be subm.
Cabomba caroliniana Gray	CAB	USA, Virginia	J.C. Ludwig, s.n. (VPI)	AY145328	DQ185527	AY590868	to be subm.
Cabomba sp.	CAB	bought from a german water gardening company	Löhne 59 (BONN)	to be subm.	DQ185528	to be subm.	to be subm.
Brasenia schreberi J.F.Gmel.	CAB	USA, Virginia	Borsch & Wieboldt 3298 (VPI, FR)	AY145329	DQ185529	AY590869	to be subm.
Brasenia schreberi J.F.Gmel	CAB	Canada	Borsch, Wiersema & Hellquist 3390 (BONN)	to be subm.	DQ185530	to be subm.	to be subm.
Illicium floridanum J. Ellis (N241)	ILL	Bonn BG	Borsch 3552 (BONN)	—	DQ185524	AY590865	to be subm.
---------- (N117)	ILL	USA, Florida	Borsch & Wilde 3104 (BONN)	AY145325	—	—	—
Barclaya longifolia Wall.	NYM	bought from a german water gardening company	Löhne 60 (BONN)	to be subm.	DQ185534	to be subm.	to be subm.
Euryale ferox Salisb.	NYM	Bonn BG 14010	Borsch 3599 (BONN)	to be subm.	DQ185537	to be subm.	to be subm.
Nuphar advena (Aiton) W.T. Aiton	NYM	USA, Florida	T. Borsch & T. Wieboldt 3298 (VPI, BONN)	AY145351	DQ185531	AY590871	to be subm.
Nuphar japonica DC. (N400)	NYM	bought from a german water gardening company	Löhne 61 (BONN)	to be subm.	DQ185532	to be subm.	to be subm.
Nuphar lutea (L.) Sibth & Sm.	NYM	Germany, Hesse	Borsch 3337 (FR)	AY145330	DQ185533	AY590872	to be subm.
Nymphaea alba L.	NYM	see Goremykin *et al.* (2004)		AJ627251	AJ627251	AJ627251	AJ627251
Nymphaea amazonum Mart. & Zucc.	NYM	Mexico, Oaxaca	Novelo, Wiersema, Hellquist, Horn 1281 (MEXU)	to be subm.	DQ185543	to be subm.	to be subm.

Table 3.1 (continued)

Species	Family	Origin	Voucher	GenBank accession numbers			
				trnT-trnF	*trnK / matK*	*petD*	*rpl16*
Nymphaea elleniae S.W.L. Jacobs	NYM	Australia, Queensland	Hellquist 16757 (MASS)	to be subm.	DQ185539	to be subm.	to be subm.
Nymphaea gracilis Zucc.	NYM	Mexico, Jalisco	Novelo, Wiersema, Hellquist, Horn 1314 (MEXU)	to be subm.	DQ185542	to be subm.	to be subm.
Nymphaea jamesoniana Planch.	NYM	USA, Florida	Borsch & Summers 3220 (FR, MO)	to be subm.	DQ185544	to be subm.	to be subm.
Nymphaea lotus var. *thermalis* (DC.) Tuzson	NYM	Bonn Bot. Gard 11547-11 (Romania)	Borsch s.n. (BONN)	to be subm.	DQ185547	to be subm.	to be subm.
Nymphaea macrosperma Merr. & L.M. Perry	NYM	Australia, Northern Territory	Jacobs & Hellquist 8796 (NSW)	to be subm.	DQ185540	to be subm.	to be subm.
Nymphaea micrantha Guill. & Perr.	NYM	Bonn BG 5830 (Zimbabwe)	Koehnen s.n. (BONN)	to be subm.	DQ185541	AY590874	to be subm.
Nymphaea novogranatensis Wiersema	NYM	Mexico, Oaxaca	Novelo & Wiersema 1187 (MEXU)	to be subm.	DQ185545	to be subm.	to be subm.
Nymphaea odorata Ait. subsp. *tuberosa* (Paine) Wiersema & Hellq.	NYM	Canada, Manitoba	Borsch, Hellquist & Wiersema 3389 (BONN, NASC)	to be subm.	DQ185549	AY590873	to be subm.
Nymphaea oxypetala Planch.	NYM	Bolivia, Santa Cruz	Ritter, G.E. Crow, Garvizu, & C. Crow 4491 (NHA)	to be subm.	DQ185546	to be subm.	to be subm.
Nymphaea petersiana Klotzsch	NYM	Malawi	Chawanje, Ch. s.n. (FR, BONN)	to be subm.	DQ185548	to be subm.	to be subm.
Ondinea purpurea Hartog	NYM	Western Australia	Jacobs & Hellquist 8853 (NSW)	to be subm.	DQ185538	to be subm.	to be subm.
Victoria cruziana A.D.Orb.	NYM	Bonn BG	Löhne 55 (BONN)	to be subm.	DQ185535	AY590870	to be subm.
Victoria 'Longwood Hybrid'	NYM	Bonn BG	Borsch s.n. (BONN)	to be subm.	DQ185536	to be subm.	to be subm.
Kadsura japonica (L.) Dun.	SCH	Bonn Bot. Gard	Borsch 3411 (BONN)	to be subm.	DQ185525	to be subm.	to be subm.
Schisandra chinensis (Turcz.) Baill.	SCH	Bonn BG	Borsch & Löhne 3492 (BONN)	—	DQ185526	AY590866	to be subm.
Schisandra rubriflora Rehder & E. H. Wilson	SCH	Bonn BG 0727 ex BG Munich	Borsch 3477 (BONN)	AY145327	—	—	—

3.2.2 DNA isolation, amplification and sequencing

Total genomic DNA was isolated from fresh or silica-gel-dried leaf tissue, or from plant material preserved in CTAB. To gain an optimal quantity of high-quality DNA a CTAB method with triple extractions was used (Borsch *et al.* 2003), as modified from Liang & Hilu (1996). After chloroform extraction, DNA was precipitated with isopropanol, resuspended in TE and further purified by ammonium acetate and sodium acetate washing steps followed by ethanol precipitation.

Four regions of the chloroplast genome were amplified using different sets of primers (Table 3.2): 1) the *petD* region, comprising the *petB–petD* spacer, the *petD* 5′ exon (only 8bp) and the *petD* intron; 2) the *rpl16* intron; 3) the *trnK–matK* region, comprising the complete *trnK* intron, the *matK* gene and the *trnK–psbA* spacer; and 4) the *trnT–trnF* region, comprising the *trnT–trnL* spacer, the *trnL* gene with its intron, and the *trnL–trnF* spacer. PCR was conducted on a T3 Thermocycler (Biometra, Göttingen, Germany), using *Taq*-DNA-polymerase, buffer

Table 3.2 — Primers used in the phylogenetic analysis of Nymphaeales for amplification (A), sequencing (S) or both (A,S). Additionally, the reading direction (Dir.), i.e. forward (F) or reverse (R), as well as nucleotide sequences and authors of the primers are given.

Region	Primer	Use	Dir.	Sequence	Author
petD	PlpetB1411F	A,S	F	GCC GTM TTT ATG TTA ATG C	this study, Chapter 2
	PlpetD738R	A,S	R	AAT TTA GCY CTT AAT ACA GG	this study, Chapter 2
rpl16	NYrps3F	A,S	F	ATC TAT GGR GTA TTA GGG	this study
	rpl16F	A,S	F	CTA TGC TTA GTG TGT GAC TC	Campagna & Downie 1998
	rpl16R	A,S	R	TCT TCC TCT ATG TTG TTT ACG	Campagna & Downie 1998
trnK / *matK*	ARmatK660R	A,S	R	AYG GAT TCG CAT TCA TA	Scheplitz (unpubl.)
	NYmatK540R	A,S	R	CAA TTA TGA TGC TCG TGT AG	Borsch 2000
	NYmatK480F	A,S	F	CAT CTG GAA ATC TTG STT C	Borsch 2000
	psbA–R	A,S	R	CGC GTC TCT CTA AAA TTG CAG TCA T	Steele & Vilgalis 1994
	trnK2R	A,S	R	AAC TAG TCG GAT GGA GTA G	Johnson & Soltis 1995
	trnKf–bry	A,S	F	GGG TTG CTA ACT CAA TGG TAG AG	Quandt, in press
	MG15	S	F	ATC TGG GTT GCT AAC TCA ATG	Liang & Hilu 1996
	NYmatK100F	S	F	AGT ATA TCT ATG CAA CTT GC	Borsch 2000
	NYmatK180R	S	R	TAA GCG TTT CAC AAT TAG TG	Borsch 2000
	NYmatK1390F	S	F	GTA AAC ATA AAA GTA CGG	this study
	NYmatK1500R	S	R	ATA TCC AAA TAC CAA ATC CG	this study
trnT – trnF	rps4–5R	A,S	F	AGG CCC TCG GTA ACG SG	Sauquet *et al.* 2003
	trnL–110R	A,S	R	GAT TTG GCT CAG GAT TGC CC	Borsch *et al.* 2003
	C	A,S	F	CGA AAT CGG TAG ACG CTA CG	Taberlet *et al.* 1991
	F	A,S	R	ATT TGA ACT GGT GAC ACG AG	Taberlet *et al.* 1991
	D	S	R	GGG GAT AGA GGG ACT TGA AC	Taberlet *et al.* 1991
	E	S	F	GGT TCA AGT CCC TCT ATC CC	Taberlet *et al.* 1991

and dNTPs from Peqlab (Erlangen, Germany). See Appendix 4 for a detailed description of PCR conditions and reaction mixes optimized for each genomic region. PCR products were purified using a QiaQuick gel extraction kit (QIAGEN Inc., Valencia/CA, USA) and sequenced either with an ABI Prism™ BigDye Terminator Cycle Sequencing Ready Reaction Kit version 1.1 (Applied Biosystems, Foster City/CA, USA) on ABI 310 or 377 automated sequencers, or with a CEQ™ DTCS Quick Start Kit (Beckman Coulter, Fullerton/CA, USA) on a CEQ™ 8000 sequencer.

3.2.3 Sequence alignment and indel coding

Sequences were aligned manually following the rules described in Löhne & Borsch (2005). using BioEdit version 5.0.9 (Hall 1999). For each of the genomic regions a separate alignment was produced initially. Mutational hotspots, i.e. regions of uncertain homology, were excluded from analysis. Also, all exon parts flanking the *petD*, *rpl16*, *trnK* and *trnL* introns were excluded from analysis as they are short, and because they were often incomplete due to the placement of sequencing primers. After alignment, gaps were coded automatically in a binary matrix using SeqState version 1.21 (Müller 2005). The "simple indel coding" strategy after Simmons & Ochoterena (2000) was applied. Alignments of all genomic regions plus the respective indel matrices were then combined to a single nexus file comprising several data partitions. Alignments and nexus data files are available from the author upon request.

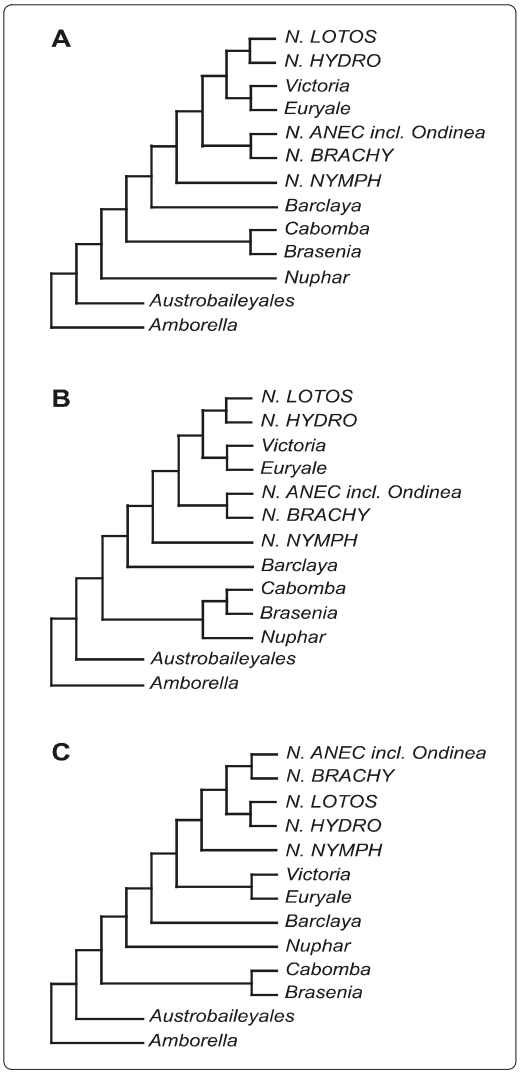
3.2.4 Phylogenetic analyses

Prior to phylogenetic analyses sequence divergence (uncorrected p distance), G/C content and other sequence statistics were calculated using SeqState (Müller 2005). To compare similarity of phylogenetic signal between genomic regions, partition homogeneity tests were performed in 1,000 replicates using PAUP* version 4.0b10 (Swofford 2002). For phylogeny reconstruction the following data partitions were analysed and compared: 1) all genomic regions separately, 2) all spacers together (*trnT-trnL*, *trnL-trnF*, *petB-petD*, *trnK-psbA*), 3) all introns (*trnL*, *trnK*, *petD*, *rpl16*), 4) all group II introns (*trnK*, *petD*, *rpl16*), 5) coding regions (only *matK*), 6) non-coding regions (all introns and spacers), 7) all indel characters, and 8) a combined data set including all nucleotide and indel characters.

Maximum parsimony. — Parsimony analyses (MP) were conducted with PAUP* version 4.0b10 employing heuristic searches with 1,000 random addition replicates and TBR branch swapping. The limit of trees saved was set to 10,000 for small matrices. Branch support was estimated through 10,000 jackknife (JK) replicates (simple addition, keeping 1 tree per replicate, deleting 36.788% of characters in each replicate). In addition, Bremer support (BS) was calculated using PAUP* with the help of PRAP version 1.21 (10 random addition replicates per constraint tree, parsimony ratchet not employed, Müller 2004). The consistency index CI (Kluge & Farris 1969) and the rescaled consistency index RC (Farris 1989) were calculated to assess levels of homoplasy. *Amborella* and the four representatives of Austrobaileyales were defined as outgroup for rooting the trees.

For a quantitative assignment of support or disagreement to each of the data partitions the partitioned Bremer support (PBS) was estimated using the program TreeRot version 2 (Sorenson 1999). Additionally, the phylogenetic structure *R* (Quandt *et al.* 2003) of each data partition was calculated in order to evaluate the phylogenetic utility of the respective markers. As a measure of average support per node *R* can range from 0 to 1, i.e. $R = 1$ if all nodes gain maximum support and $R = 0$ if not a single node is supported by more than 50% of jackknife or bootstrap replicates.

Bayesian inference. — Besides parsimony analyses, Bayesian inference (BI) of phylogeny for the combined dataset was conducted using MrBayes version 3.1 (Ronquist & Huelsenbeck 2003). Following the Akaike information criterion, Modeltest 3.06 (Posada & Crandall 1998) assigned the GTR + G model of molecular evolution to all nucleotide partitions, except the *petB–petD* spacer and the *trnT–trnL* spacer, to which the GTR model was assigned. The binary (restriction site) model implemented in MrBayes was applied to the indel partition. Analysis was performed for 1,000,000 generations applying the default settings (MCMCMC, 4 runs with 4 chains each, heating temperature 0.2, saving one tree every 100 generations). In all runs the probabilities had converged to a stable value after 15,000 generations. Thus, a consensus was calculated from 39,404 trees sampled after the burn-in.



Evaluating alternative topologies. — In order to compare the likelihood of the topology inferred from our data set with alternative phylogenetic hypotheses Kishino-Hasegawa (KH) tests (Kishino & Hasegawa 1989) were performed with the combined matrix (all substitutions) using the *Lscores* option in PAUP* and GTR + G model settings obtained with Modeltest 3.06 (Posada & Crandall 1998). The evaluated alternative hypotheses, as illustrated in Figure 3.1, refer to the position of *Nuphar* as either basal in Nymphaeales (A) or sister to Cabombaceae (B, see Löhne & Borsch 2005 and Chapter 2 of this thesis), and to the monophyly of the genus *Nymphaea* with *Victoria–Euryale* as sister clade (C, Borsch *et al.* subm.).

Figure 3.1 — Alternative hypotheses for relationships within Nymphaeales, tested against the optimal tree obtained from the present data set in Kishino-Hasegawa tests. Topologies A and B test different positions of *Nuphar*. Tree C refers to the monophyly of the genus *Nymphaea* with respect to *Victoria* + *Euryale*.

Analyses with reduced taxon set. — To assess the effects of taxon sampling on the inferred phylogeny we conducted several MP analyses with the data set reduced to 8 taxa, comprising only one representative of each genus of Nymphaeales. Five data subsets were created by choosing either *N. elleniae*, *N. micrantha*, *N. amazonum*, *N. lotus* or *N. odorata* (representing the five subgenera of *Nymphaea*) as the representative of the genus *Nymphaea*. In line with the analysis of Les *et al.* (1999), *Cabomba* and *Brasenia* were defined as outgroup in these 8–taxon data sets. MP trees, jackknife and Bremer support were calculated as described above.

3.3 Results

3.3.1 Sequence variability observed in the data sets

To obtain the multi–locus chloroplast dataset for Nymphaeales an average of 6633 nucleotides were sequenced per taxon. 658 nucleotides were not analysed because they were situated at the borders of the matrices or in exons. An average of 527 nucleotides per taxon fell within mutational hotspot regions and therefore was also not included. The final matrix containing all data partitions (spacers, introns and *matK*) comprised 6597 characters plus a presence-absence matrix of 369 indels, providing total numbers of 2224 variable and 1356 parsimony informative characters. A total of 19 hotspots was determined, occurring in all data partitions except the *petD* intron and the *matK* gene. Especially from the *trnK–psbA* spacer and from the *rpl16* intron a considerable amount of characters had to be excluded because of hotspots. The *rpl16* intron is also the region with highest length variability (standard deviation of mean sequence length = 84.6; Table 3.4), mainly caused by large

Table 3.3 — Results of partition homogeneity tests. Listed are the p values of all pairwise comparisons. The p value of the overall analysis (= partitions homogeneity test using all partitions in the same analysis) is 0.3040.

	***petB-petD* spacer**	***petD* intron**	***rpl16* intron**	***trnK* intron**	***matK***	***trnK-psbA* spacer**	***trnT-trnL* spacer**	***trnL* intron**	***trnL-trnF* spacer**
***petB-petD* spacer**	–	0.65	1.0	1.0	1.0	0.59	1.0	0.77	0.99
***petD* intron**	–	–	0.50	0.36	0.67	0.80	0.44	0.88	0.67
***rpl16* intron**	–	–	–	0.83	0.09	0.88	0.33	0.13	0.73
***trnK* intron**	–	–	–	–	0.80	0.71	0.13	0.29	0.28
matK	–	–	–	–	–	0.88	0.18	0.43	0.45
***trnK-psbA* spacer**	–	–	–	–	–	–	0.79	0.76	0.84
***trnT-trnL* spacer**	–	–	–	–	–	–	–	0.08	0.51
***trnL* intron**	–	–	–	–	–	–	–	–	0.28
***trnL-trnF* spacer**	–	–	–	–	–	–	–	–	–

sections being only present in the outgroup taxa. The *matK* gene, on the other hand, is most conserved in sequence length as is evident by a low SD of 10.6 although *matK* has twice the length of the *rpl16* intron. And the number of indels coded in the *matK* gene (N = 19) is low compared to its length. More indels per character were coded in the introns and, especially, in the spacers (see last column in Table 3.4). Sequence divergence is lowest in the *petB–petD* spacer and *petD* intron, highest in the *rpl16* intron and the *trnT-trnL* and *trnL-trnF* spacers. GC-content ranges between 30 and 40% in all partitions. Percentages of

Table 3.4 — Comparison of sequence statistics between all chloroplast genomic regions analysed for the 29 taxon data set of Nymphaeales. "Char." – number of characters in the alignment matrix (excluding hotspots); "Length range" – actual sequence length in nucleotides (minimal and maximal observed value, including hotspots); "Mean length" – mean of all observed sequence lengths (standard deviation in brackets); "% Div." – pairwise sequence distance in percent (uncorrected p distance, overall mean, lowest and highest scores in brackets); "% GC" – GC content in percent; "% var." – percentage of variable characters; "% inf." – percentage of parsimony informative characters; "indels" – number of length mutations that were coded by SeqState (percentage of informative indels in brackets); "Indels/Char." – number of indels per number of characters (as a measure of indel frequency).

Data Partition	Mean length	Length range	No. Char.	% Div.	% GC	% var.	% inf.	Indels % inf.	Indels / Char.
Individual data sets									
petB–petD spacer	196.8 (13.2)	164–224	273	4.5 (0.0–15.4)	30.0	20.5	8.4	27 (63%)	0.10
petD intron	652.7 (33.3)	621–733	817	4.6 (0.0–14.0)	39.5	22.4	12.2	42 (55%)	0.05
rpl16 intron	825.7 (84.6)	763–1109	858	8.0 (0.0–21.3)	37.8	30.3	20.0	49 (59%)	0.06
trnK intron	1007.9 (21.8)	957–1068	1130	6.6 (0.0–19.1)	38.1	28.2	18.1	66 (58%)	0.06
matK	1517.8 (10.6)	1499–1542	1590	6.5 (0.0–19.0)	36.1	31.3	20.6	19 (58%)	0.01
trnK–psbA spacer	316.6 (17.0)	287–396	170	6.6 (0.0–24.0)	39.4	30.6	15.9	10 (50%)	0.06
trnT–trnL spacer	483.8 (54.4)	349–684	600	8.8 (0.0–25.2)	38.8	31.5	20.3	59 (49%)	0.10
trnL intron	537.9 (32.8)	474–606	581	5.6 (0.0–14.3)	39.5	23.1	13.8	41 (54%)	0.07
trnL–trnF spacer	385.2 (33.0)	244–441	579	9.5 (0.0–28.2)	35.1	28.0	17.3	56 (46%)	0.10
Combined partitions									
introns	3024.1 (98.4)	2916–3384	3386	6.3 (0.0–17.6)	38.6	26.5	16.5	198 (57%)	0.06
spacers	1382.4 (52.5)	1225–1569	1621	7.9 (0.0–23.3)	36.3	28.3	16.8	152 (51%)	0.09
non-coding	4406.5 (121.1)	4268–4852	5007	6.7 (0.0–19.1)	37.9	27.1	16.6	350 (54%)	0.07
all substitutions	5924.4 (119.7)	5767–6356	6597	6.7 (0.0–19.1)	37.4	28.1	17.5	369 (54%)	0.06
total (incl. indels)	–	–	6966	–	–	31.9	19.5	–	–

variable characters range from 20.6 to 31.5%, and percentages of parsimony informative characters range from 8.4 to 20.6%. Lowest amounts of informative characters relative to the total number of characters are present in the *petD* region (spacer and intron), highest in the *matK* gene.

3.3.2 Trees obtained from different approaches

Partition homogeneity tests (Table 3.3) indicated that data partitions are not significantly incongruent (P value of overall analysis 0.30, P values in all pairwise comparisons ≥0.08). Therefore, all data partitions could be analysed in a combined matrix. Maximum parsimony analysis of the combined matrix yielded one shortest tree with 2562 steps (CI = 0.85, RC = 0.76), which is shown in Figure 3.2 with jackknife support above and Bremer support below branches. Inclusion of the indel matrix led to the same tree (2979 steps, CI = 0.85, RC = 0.77) with similar jackknife, but slightly increased Bremer support (Table 3.6). This fully resolved tree could be obtained from combined analysis of all non–coding partitions using only substitutions, and when adding *matK* and indels, respectively. Consensus trees of individual data sets differed in degree of resolution, ranging from 8 nodes resolved by the *petB–petD* and *trnK–psbA* spacers to 24 nodes resolved by the *trnK* intron (Table 3.6, column 7). In parallel, *R* is lowest in the *petB–petD* and *trnK–psbA* spacers (*R* = 0.18) and highest in the *trnK* intron (*R* = 0.59; Table 3, column 6). Partitioned Bremer support (PBS) for selected nodes and summed PBS values for all data partitions are presented in Table 3.5. Six nodes (node 5, 8, 14, 22, plus 20 and 23 which are not shown in the table) receive PBS values below zero from one of the data partitions (or two in case of node 14), indicating phylogenetic signal that contradicts the total evidence tree. However, since in these few cases values do not drop below -1, the overall conflict is rather small. The tree found in the Bayesian analysis is fully congruent to the total evidence MP tree. It is shown in Figure 3.3 with branch lengths estimated from 39,404 trees. Posterior probabilities (PP) of clades are given above branches.

Table 3.5 — Partitioned Bremer Support for selected nodes of the total evidence tree.
Node numbers refer to nodes in the MP tree (Figure 3.2).

Data Partition	Partitioned Bremer Support (PBS) for selected nodes										Total PBS
	5	8	11	12	14	15	17	19	22	24	
petB–petD spacer	5	0	1	2	0	1	0	0	0	0	19.0
petD intron	12	0	1	1	0	1	0	0	0	0	81.0
rpl16 intron	5	2	1	2	2.5	3.4	0	0	2	2	134.9
trnK intron	11	1	1.5	2	0.5	3.1	2	1	1	3	158.1
matK	24	2	5	8	-1	1.4	1	0	3	2	270.4
trnK–psbA spacer	1	0	1	0	0	0	0	0	0	0	23.0
trnT–trnL spacer	9	3	3	0	0.5	7.4	1	0	-1	0	96.9
trnL intron	9	-1	3	5	-0.5	0.6	0	0	0	0	70.1
trnL–trnF spacer	-1	0	3.5	3	0	3	1	0	2	0	64.5
all substitutions	75	7	20	23	2	21	5	1	7	7	918.0

Table 3.6 — Results of maximum parsimony analyses and phylogenetic structure (R) of different data partitions. Node numbers refer to nodes in Figure 1. "Trees" — number of trees saved; "TL" – length of shortest trees; "CI" – consistency index; "RC" – rescaled consistency index; "R" – phylogenetic structure; "No. of Nodes" – number of nodes present in the strict consensus; "n.p." – node not present in the respective tree.

Data partition	Trees	TL	CI	RC	*R*	No. of Nodes	Jackknife (Bremer support) for selected nodes									
							5	8	11	12	14	15	17	19	22	24
Individual data sets																
petB–petD spacer	42	69	0.94	0.89	0.18	8	99	n.p.	63	82	n.p.	63	n.p.	n.p.	n.p.	n.p.
petD intron	7	231	0.88	0.81	0.40	13	100	n.p.	62	n.p.	n.p.	63	n.p.	n.p.	n.p.	n.p.
rpl16 intron	2	387	0.81	0.71	0.58	17	98	59	n.p.	89	82	87	n.p.	n.p.	84	87
trnK intron	5	464	0.84	0.73	0.59	24	100	n.p.	60	83	n.p.	85	91	64	63	95
matK	15	687	0.84	0.75	0.58	18	100	n.p.	n.p.	100	n.p.	87	63	n.p.	94	86
trnK–psbA spacer	10,000	77	0.87	0.76	0.18	8	66	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.
trnT–trnL spacer	40	250	0.87	0.80	0.53	18	100	99	96	n.p.	n.p.	100	63	n.p.	n.p.	n.p.
trnL intron	26	167	0.89	0.83	0.39	13	100	n.p.	94	99	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.
trnL–trnF spacer	540	218	0.85	0.76	0.42	16	n.p.	65	81	87	n.p.	94	73	n.p.	60	n.p.
Combined partitions																
introns	2	1255	0.84	0.75	0.82	25	100(37)	81(2)	96(6)	100(9)	98(2)	100(7)	91(2)	84(1)	97(3)	99(5)
spacers	99	618	0.87	0.78	0.65	20	100(14)	97(3)	100(7)	99(4)	n.p.	100(10)	90(2)	n.p.	n.p.	n.p.
non–coding	1	1874	0.85	0.76	0.88	26	100(51)	97(5)	100(15)	100(15)	97(3)	100(19)	99(4)	88(1)	98(4)	100(5)
all substitutions	1	2562	0.85	0.76	0.90	26	100(75)	99(7)	100(20)	100(23)	90(2)	100(21)	100(5)	89(1)	100(7)	100(7)
indels	12	414	0.89	0.83	0.54	22	100(7)	n.p.	94(5)	99(5)	n.p.	100(8)	n.p.	n.p.	63(1)	64(1)
total (incl. indels)	1	2979	0.85	0.77	0.90	26	100(82)	99(8)	100(26)	100(30)	90(2)	100(31)	100(6)	88(1)	100(8)	100(8)

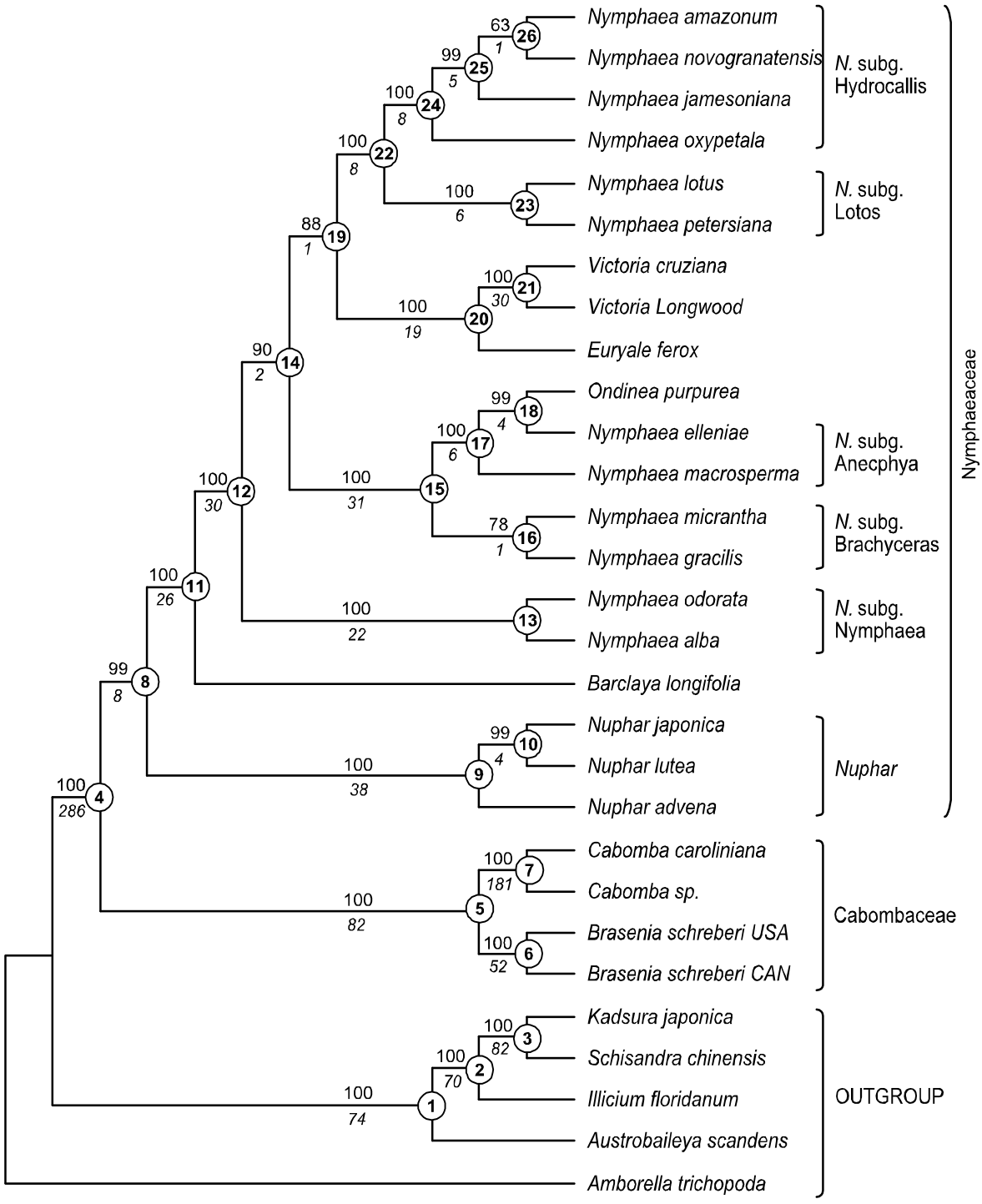


Figure 3.2 — Single most parsimonious tree of Nymphaeales obtained from a combined analysis of all markers (substitutions and indel matrix). Jackknife values are given above branches, Bremer support below. Nodes are numbered consecutively (in circles).

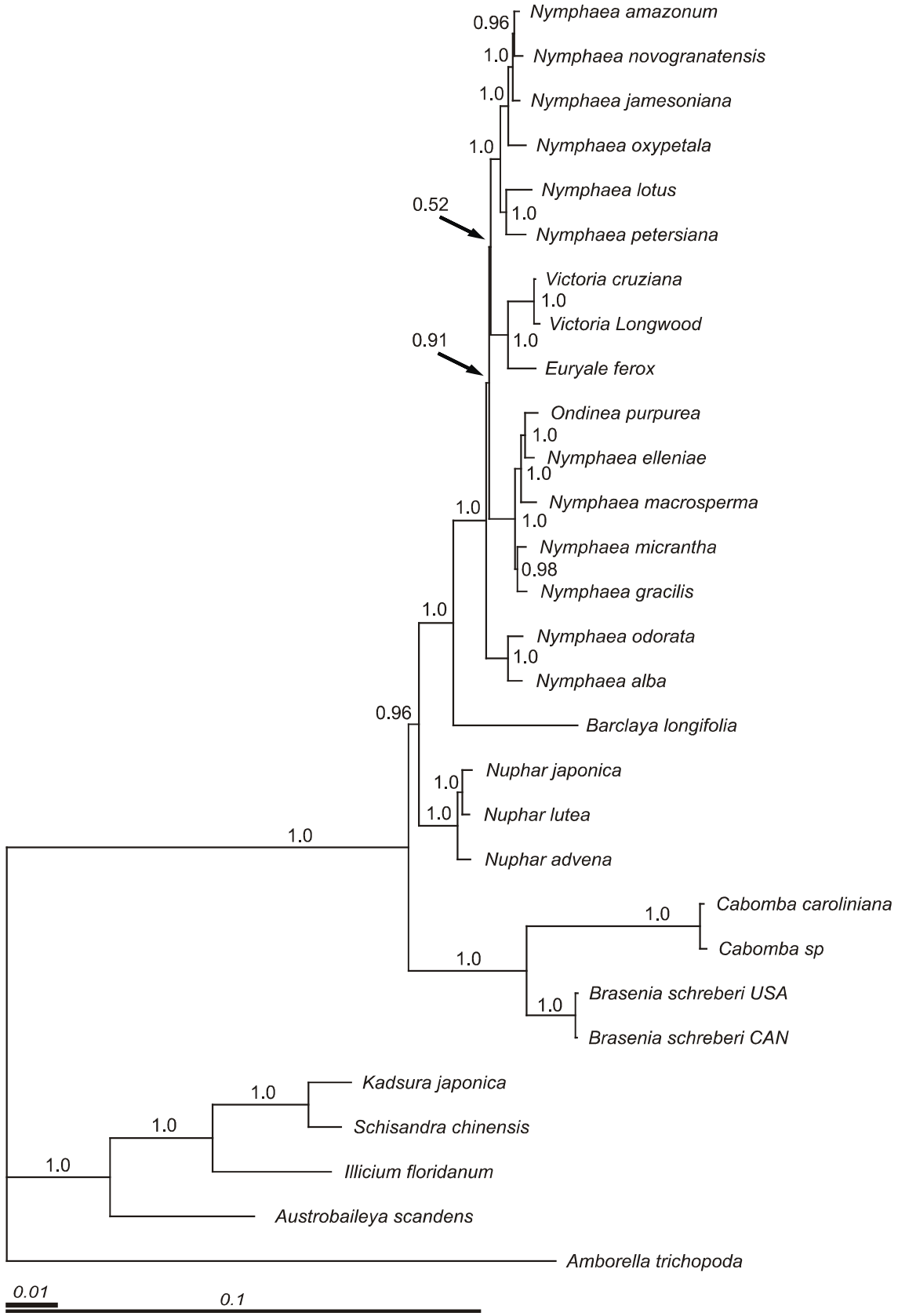


Figure 3.3 — Phylogram of the combined data set of Nymphaeales including all markers (substitutions and indels). The tree is a 50% majority rule consensus of 39,404 trees obtained from 4 runs of Bayesian analysis implementing the GTR+G model. Branch lengths reflect changes per site. Posterior probabilities are given above branches.

3.3.3 Phylogeny of Nymphaeales

In both total evidence trees (Figure 3.2 and Figure 3.3) the Cabombaceae, consisting of *Cabomba* and *Brasenia*, are revealed as a well-supported clade (JK = 100, BrS = 82, PP = 1.0). Nymphaeaceae, consisting of *Nuphar*, *Barclaya*, *Nymphaea*, *Ondinea*, *Euryale* and *Victoria*, also emerge as monophyletic with high jackknife (JK = 99) but rather low Bremer support (BrS = 8) and medium posterior probability (PP = 0.96). *Nuphar* is depicted basal in Nymphaeaceae, followed by *Barclaya*. *Victoria* and *Euryale* form a well–supported sister clade. However, the genus *Nymphaea* appears to be paraphyletic with respect to both *Ondinea* and the *Victoria-Euryale* clade. *Ondinea* is placed within *Nymphaea* subg. *Anecphya* with maximum support in both MP and Bayesian analysis. *Nymphaea* subg. *Anecphya* and subg. *Brachyceras* together form a well-supported clade (JK = 100, BrS = 31, PP = 1.0), whereas *N*. subg. *Hydrocallis* is grouped with *N*. subg. *Lotos* (JK = 100, BrS = 8, PP = 1.0). *Victoria* and *Euryale* emerge as a sister group to the *Hydrocallis-Lotos* clade with average jackknife (JK = 88), but low Bremer support (BrS = 1) and low posterior probability (PP = 0.52). *Nymphaea* subg. *Nymphaea* appears basal (JK = 90, BrS = 2, PP = 0.91) to all other subgenera as well as *Victoria–Euryale* and *Ondinea*.

To evaluate alternative hypotheses (see Figure 3.1) on phylogenetic relationships in Nymphaeales Kishino-Hasegawa tests were performed. The total evidence tree obtained from the present data set (see Figure 3.2 or Figure 3.3) was significantly favoured over all alternatives (P = 0.000, Table 3.7).

Table 3.7 — Results of Kishino-Hasegawa tests. The "optimal tree" is the total evidence tree of Fig. 2, see Fig. 1 for topologies of alternatives A to C. Difference can be regarded as significant if $P < 0.05$.

	Optimal tree	Alternative A	Alternative B	Alternative C
-ln	22832.36905	22833.98303	22836.29277	22836.56115
Diff. to -ln of MP tree	-	1.61398	3.92372	4.19210
P	-	0.000	0.000	0.000

3.3.4 Effects of taxon sampling

In each of the MP analyses of the taxon subsets a single most parsimonious tree was obtained. The trees differ in topology, i.e. in the position of *Ondinea*, dependent on which species of *Nymphaea* was included (Figure 3.4, A–E). If *N. elleniae* (subg. *Anecphya*) or *N. micrantha* (subg. *Brachyceras*) were chosen, *Ondinea* grouped with them with maximum support (JK = 100, Figure 3.4, A, B). *Ondinea* appears basal to *Victoria–Euryale* and *Nymphaea* if *N. amazonum* (subg. *Hydrocallis*) or *N. lotus* (subg. *Lotos*) were chosen as representatives (Figure 3.4, C, D), with high support (JK = 96) in the first case and medium (JK = 87) in the latter. *Nymphaea odorata* (subg. *Nymphaea*), on the other hand, emerges basal to a clade consisting of *Ondinea* and *Victoria-Euryale* (JK = 74, Figure 3.4, E).

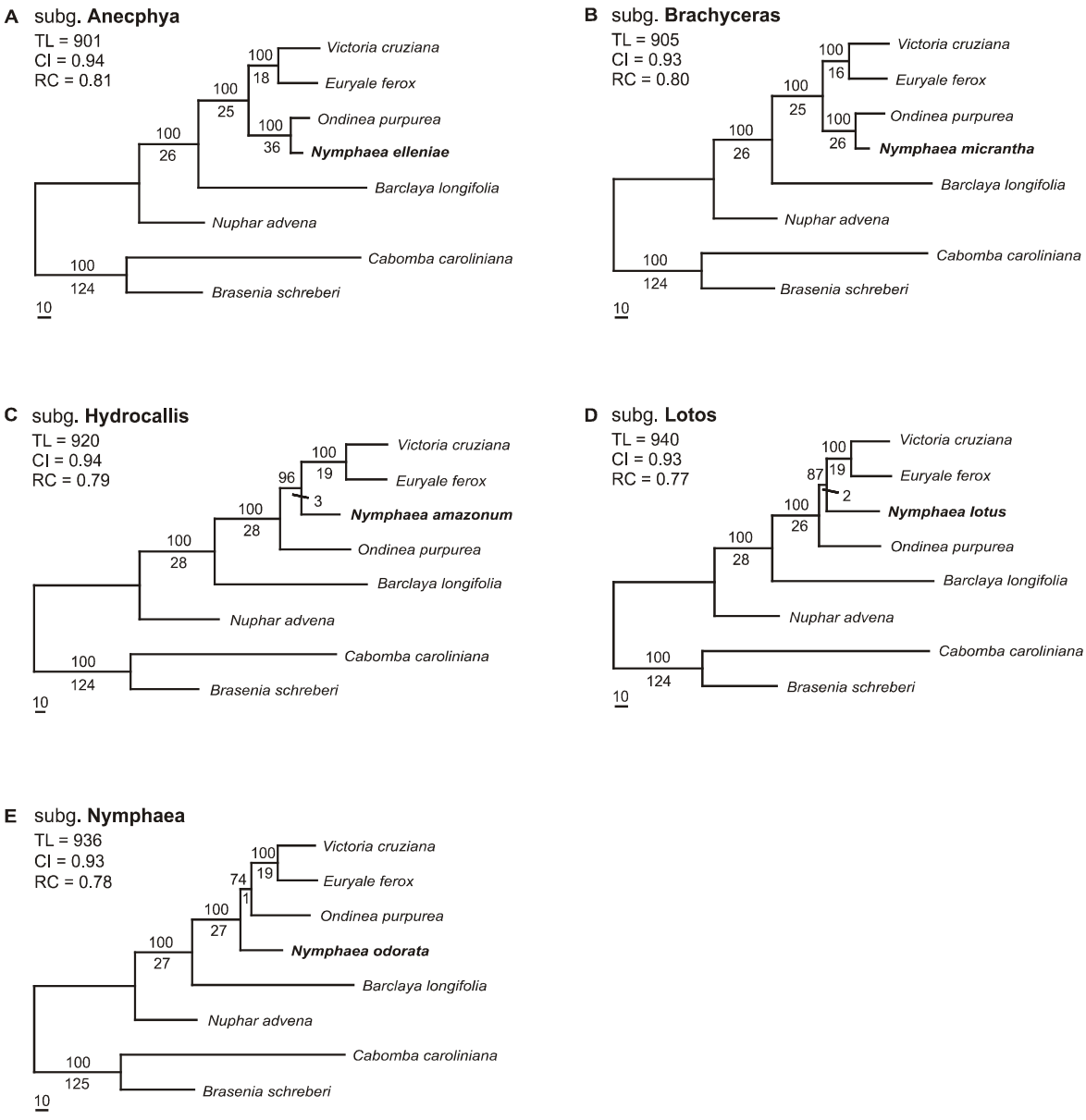


Figure 3.4 — Results of phylogenetic analyses with taxon sampling reduced from 29 to 8 taxa, thus reflecting the number of taxa sampled by Les *et al.* (1999). Five independent parsimony analyses were run to test whether members of different subgenera of Nymphaea yield different trees (A, subg. *Anecphya*; B, subg. *Brachyceras*; C, subg. *Hydrocallis*; D, subg. *Lotos*; E, subg. *Nymphaea*). Cabomba and Brasenia were used as outgroup taxa. Jackknife values are given above branches, Bremer support below. Branch lengths reflect number of changes.

3.4 Discussion

3.4.1 Structure of data and reliability of the trees

In the present study we compiled the most extensive molecular data set to date for Nymphaeales. Due to a relatively high percentage of variable and informative characters, that is comparable to other data sets based on non–coding chloroplast DNA (e.g., Renner & Chanderbali 2000, Borsch *et al*. 2003, Löhne & Borsch 2005, Neinhuis *et al*. 2005), a total of 1156 parsimony informative alignment positions plus 200 informative characters from coded length mutations could be obtained. Also similar to previous analyses based on non–coding genomic regions, the occurrence of extremely variable nucleotide stretches, which impair a clear alignment, was confined to several, clearly demarcated mutational hotspots comprising less than 10% of all nucleotides (Borsch *et al*. 2005).

The present data set not only comprises a high number of informative characters (compared to, e.g., 108 in Les *et al*. 1999 for Nymphaeales with different taxon sampling,, or 652 in Renner & Chanderbali 2000 for Laurales of comparable taxon sampling), but it is characterized by generally low degrees of homoplasy (apparent from high CI and RC values) and a strong phylogenetic signal. Both methods used for phylogenetic analysis (maximum parsimony and Bayesian inference) yielded the same results, i.e. an exactly matching topology and comparable support for nodes. The compelling statistical support obtained in both analyses (MP and BI) emphasizes the homogeneous structure and strong phylogenetic signal of the present data set. Thus, a majority of nodes in the total evidence tree (17 out of 26) gains maximum support (JK = 100, PP = 1.0). In total, 21 of 26 nodes are well supported by the underlying data, even if we follow rather conservative approaches in interpreting support values (see Simmons *et al*. 2004, Zander 2004, Schönenberger *et al*. 2005 for discussion), and regard only nodes with at least 95% Jackknife support and posterior probalities of 0.95 as well supported. Furthermore, estimation of partitioned Bremer support revealed no considerable conflict between the nine data partitions. Negative PBS values of -1 or smaller, which occurred in a few cases, can be regarded as non–significant (Creer *et al*. 2003, Lambkin 2004).

The fact that none of the data partitions alone provided enough information to fully resolve the relationships among the sampled taxa, but that combining markers resulted in a single most parsimonious tree and strongly increased the number of well–supported clades, demonstrates the necessity of a certain amount of informative characters to clarify phylogenetic relationships at this taxonomic level. The inclusion of further markers might therefore be helpful to verify those relationships that do not gain sufficient support by the present dataset.

Microstructural changes

The indel matrix compiled from length mutations in the different data partitions provides phylogenetic information that is congruent with the topology obtained from substitutions. Clades with high statistical support from substitutions and with long branches are characterized by high numbers of autapomorphic indels, such as the Cabombaceae (7 synapomorphic indel characters) and the genus *Cabomba* (25 indels), or the clade consisting

of *Nymphaea* subg. *Brachyceras*, subg. *Anecphya* and *Ondinea* (7 indels). Some examples of parsimony informative length mutations are illustrated in Figure 3.5. The indel matrix of the present data set is one of the largest that has been compiled so far (see Simmons *et al.* 2001). The fact that most clades in Nymphaeales are resolved, even from indels alone (nodes not resolved: 8, 14, 17, 18, 19, 23, 26), demonstrates the high value of length mutations in phylogenetic analyses. Since the increasing understanding of molecular evolutionary processes leading to such microstructural changes allows a reasonable, unambiguous alignment of length-variable sequences, it is becoming a matter of course to include information from length mutations into sequence analysis (e.g., Kjer 1995, Graham *et al.* 2000, Kelchner 2000, 2002, Quandt *et al.* 2003). I agree with Simmons & Ochoterena (2000) and other authors (e.g., Giribert & Wheeler 1999, Kelchner 2000, Müller 2006), that not coding length mutations is equivalent to discarding data.

Figure 3.5 — Details of the alignment matrix for Nymphaeales and outgroup showing indels synapomorphic for major clades.

Indel no. 23, from the *petB–petD* spacer, is a synapomorphy for the *Hydrocallis–Lotos* clade in the genus *Nymphaea*. Indel no. 151, from the *trnK* intron, unites all members of Nymphaeaceae except *Nuphar*. Indels no. 191 and 192 occur in the *matK* gene. No. 191 is synapomorphic for a *Nymphaea–Ondinea–Victoria–Euryale* clade while no. 192 is one of 7 indels uniting *Ondinea* with the *Nymphaea* subgenera *Anecphya* and *Brachyceras*. Note that all four indels shown are simple sequence repeats. Nucleotide substitutions that are parsimony informative in the 29 taxon data set are marked with a dot.

Branch lengths

A closer look at Figure 3.3 reveals that all nodes which do not gain maximum support in Parsimony or Bayesian analyses are characterized by rather short branches. This is true for node 8 (uniting all members of Nymphaeaceae), node 14 (uniting the *Anecphya–Brachyceras* and the *Hydrocallis–Lotos–Victoria* clades), node 16 (*N.* subg. *Brachyceras*), node 19 (uniting *Hydrocallis–Lotos* with *Victoria–Euryale*), and node 26 (*N. amazonum* sister to *N. novogranatensis*). Short branch lengths in both the Bayesian and the MP tree (branch length not shown), in combination with low Bremer support values (Figure 3.2),

indicate the presence of only a few phylogenetically informative characters. This can be due to rather rapid diversification at these levels (Richardson *et al.* 2001, Shaw *et al.* 2003). Long branches, on the other hand, are leading to the clade of Cabombaceae (and within Cabombaceae to *Cabomba*) and to *Barclaya*. It is remarkable that *Cabomba* and *Barclaya* represent two extremes of life forms in the Nymphaeales. Whereas *Cabomba* is a completely submersed plant with strongly dissected leaves, one species of *Barclaya* (*B. rotundifolia* Hotta) is semi–terrestrial. This indicates that strong shifts in morphological features or life form may be related to accelerated molecular evolution (Müller *et al.* 2004).

3.4.2 Phylogeny and evolution of Nymphaeales

The results of the present study confirm some previous hypotheses on phylogenetic relationships in the order Nymphaeales, but due to the comprehensive taxon and character sampling we also get new insights into the evolutionary history of this group. Thus, these results corroborate the monophyly of Cabombaceae, which has been hypothesized before based on morphological, anatomical and molecular characters (see Williamson & Schneider 1993). Within Nymphaeaceae the position of *Barclaya* as sister to a well–supported clade consisting of *Nymphaea*, *Ondinea*, *Victoria* and *Euryale* has been confirmed (similar to Les *et al.* 1999). Furthermore, the sister relationship of *Victoria* and *Euryale*, which had already been proposed by Caspary (1891, based on seed morphology and the presence of spines), is clearly supported here. *Nuphar* is inferred to be basal in a monophyletic clade comprising all members of Nymphaeaceae, similar to previous studies based on anatomy, morphology and molecules (Ito 1987, Les *et al.* 1999). However, despite the high amount of characters sampled the monophyly of the Nymphaeaceae is not convincingly supported (see below). More strikingly, the present data set does not give support for the monophyly of the genus *Nymphaea*. In contrast to all previous phylogenetic studies and classifications, it is inferred to be paraphyletic with respect to the *Victoria–Euryale* clade and to *Ondinea*. Because of their relevance for understanding the evolution of the water lily family, these findings will be address in closer detail in the following paragraphs.

Nuphar and the monophyly of Nymphaeaceae

Nymphaeaceae are revealed as a monophyletic group in the present analysis with *Nuphar* branching first in this clade (node 8). Alternative topologies, i.e. *Nuphar* basal in Nymphaeales (Figure 3.1, A) or *Nuphar* sister to Cabombaceae (Figure 3.1, B), were rejected by Kishino-Hasegawa tests. However, this inference is based on only a few characters, as evident from the low Bremer support (Figure 3.2) as well as the short branch leading to this node (Figure 3.3). A reason for the scarcity of informative characters at the base of Nymphaeales could be a rapid, early diversification into the three major lineages. Based on the observation that support values in the present data set increased remarkably after combining different partitions it can be expected that the monophyly of Nymphaeaceae can be corroborated by future studies including more characters. However, there are several morphological and anatomical features, such as its stout creeping rhizomes, a superior syncarpous gynoecium with discontinuous stigmatic rays, echinate anasulcate pollen, emergent fruits having smooth exarillate seeds with distinctive dehiscence, which demonstrate the distinctness of this genus from other Nymphaeaceae and from

Cabombaceae. Until the monophyly of Nymphaeaceae has been proved by further data, the idea of Nupharaceae as a separate family (suggested by Kerner von Marilaun 1891, Nakai 1943, Takhtajan 1997) should not be ignored completely. Relationships among the three sampled species of *Nuphar* agree with the hypothesis advanced by Padgett *et al.* (1999) in their comprehensive analysis of this genus. *Nuphar lutea* and *N. japonica*, as representatives of the Old World *Nuphar* sect. *Nuphar* form a well–supported clade that is sister to *N. advena*, the only representative of New World *Nuphar* sect. *Astylus* in this study.

Polyphyly of the genus Nymphaea

Perhaps the most striking result of this analysis is the inferred polyphyly of the genus *Nymphaea* with respect to *Victoria* and *Euryale*. Since previous studies did not consider subgenera of *Nymphaea* in their analyses (e.g., Ito 1987, Moseley *et al.* 1993, Les *et al.* 1999), did not resolve relationships among them (Borsch 2000), or found only weak support for the monophyly of *Nymphaea* this is the first time that this phenomenon could be observed. Here, *Victoria* and *Euryale* are inferred to be sister to a clade comprising the water lily subg. *Hydrocallis* and subg. *Lotos*. This node gains considerable jackknife support (JK = 88), but Bremer support and posterior probabilities are rather low (BS = 1, PP = 0.51). The same applies for node 14, which separates the temperate subg. *Nymphaea* from the rest of the genus including *Ondinea*, *Euryale* and *Victoria* (JK = 90, BS = 2, PP = 0.91). Although an alternative topology (Figure 3.1, C), constraining the genus *Nymphaea* to be monophyletic with respect to *Victoria–Euryale* (but not with respect to *Ondinea*!), would be only 3 to 5 steps longer, the total evidence tree was significantly favoured over this alternative as shown by Kishino–Hasegawa tests.

Although it may seem premature to propose a close affinity of *Victoria–Euryale* to subg. *Hydrocallis* and subg. *Lotos*, some morphological and ecological similarities of these plants are remarkable: *Hydrocallis* and *Lotos* are night–flowering water lilies, just as *Victoria* (Valla & Cirino 1972, Prance & Arias 1975, Wiersema 1988). Both subgenera and *Victoria* are characterized by relatively large, whitish flowers and prominent carpellary appendages (e.g., Wiersema 1987, 1988, Schneider & Williamson 1993). Furthermore, they share the same pollinators, scarab beetles of the tribe Cyclocephalini, even though *Victoria* and subg. *Hydrocallis* only occur in South America while subg. *Lotos* is palaeotropic (Ervik & Knudsen 2003, Hirthe & Porembski 2003). Only *Euryale* does not fit this pattern, as it has purple, completely cleistogamous flowers without conspicuous carpellary appendages (Okada & Otaya 1930, Kadono & Schneider 1987). However, these differences could be concomitant with the shift from chasmogamy to cleistogamy in *Euryale*.

Ondinea — the apetalous water lily

Although doubt may exist on the inferred position of *Victoria* and *Euryale*, this comprehensive molecular data set adds further compelling evidence to the hypothesis of a close affinity between *Ondinea* and the Australian water–lilies (*Nymphaea* subg. *Anecphya*), which was first uncovered by (Borsch *et al.* subm.) using data from the *trnT–trnF* region. The clade consisting of *N. elleniae*, *N. macrosperma* and *Ondinea purpurea* receives high support from all statistical tests (JK = 100, BrS = 6, PP = 1.0). In fact, the results strongly indicate an origin of *Ondinea* from within the *Anecphya* subgenus, since *Ondinea* is

depicted as sister to the small–seeded *Nymphaea elleniae*. The significance of these results for understanding the evolution of Nymphaeales becomes evident if one considers the striking morphological differences between *Ondinea* and *Nymphaea*. Obviously, dramatic morphological changes involving many parts of the plant have occurred during the evolution of *Ondinea*. A possible mechanism that could explain such drastic shifts is neoteny. Schmucker (1932) observed comparable phenomena, i.e. dwarfing and reduction of floral organ number, in *Nymphaea micrantha* clones grown from leaf offshoots. Schneider *et al.* (2003) also discuss neoteny in Nymphaeales, but with regard to the reduced flowers in Cabombaceae.

However, despite its distinct morphological features *Ondinea* has been considered to be closely related to *Nymphaea* by other scientists before (Den Hartog 1970, Williamson & Moseley 1989, Williamson *et al.* 1989). In fact, den Hartog (1970) stated in his description of *Ondinea purpurea*, that "it is closer to *Nymphaea* than to any of the other genera within the Nymphaeaceae. In general the *Ondinea* flower can be regarded as an apetalous *Nymphaea* flower." Les *et al.* (1999) observed in their analysis "a weak tendency for *Nymphaea* and *Ondinea* to resolve as a separate clade", which was induced mainly by molecular data (18S and *matK*). Since Les *et al.* (1999) sampled only one species of *Nymphaea* subg. *Nymphaea* (*N. odorata*) they could not detect the close affinity of *Ondinea* to subg. *Anecphya*, which underscores the importance of judicious taxon sampling (see below).

Considering the fact that the present tree is solely based on the chloroplast genome, the possibility of a hybrid origin of *Ondinea* remains plausible, with the maternal parent being a water lily and the paternal being an unknown, presumably extinct member of Nymphaeales. Thus, a cross check with nuclear markers will be necessary to track the real phylogenetic history of *Ondinea*. This issue will be addressed in the following chapter of this thesis, which presents a detailed study of relationships among Australian water lilies using nuclear and chloroplast markers.

Other relationships among and within the subgenera of Nymphaea

The results of the present study strongly corroborate a close relationship of the *Nymphaea* subgenera *Hydrocallis* and *Lotos*. This affinity has been suggested earlier (e.g., Wiersema 1987), since it is indicated by several morphological synapomorphies, such as pollen morphology, presence of highly developed carpellary appendages, a medial position of anthers on the stamens and the nocturnal flowering pattern. The *Hydrocallis–Lotos* clade is well supported also from data subsets, e.g. the *matK* gene (see Table 3.6), as well as the *trnT–trnF* data set of (Borsch *et al.* subm.). This analysis confirm the previous findings of Borsch *et al.* (subm.) that the SE African species *Nymphaea petersiana* does not belong to subg. *Brachyceras*, as previously assumed, but appears to be closely related to *Nymphaea lotus* (JK = 100, BrS = 6, PP = 1.0).

Within subg. *Hydrocallis* *Nymphaea oxypetala* is sister to a clade consisting of *N. amazonum*, *N. jamesoniana* and *N. novogranatensis*. Among the subg. *Hydrocallis* species sampled *N. oxypetala* is exceptional due to its submersed habit, an unusual floral morphology, presence of spherical staminal sclereids, and its polyploid nature (6n = 84, see

Wiersema 1987). A possible affinity of *N. novogranatensis* (2n = 28) to a 2n = 18 group of species including *N. amazonum* has been proposed by Wiersema (1987) based on similarities in morphology, phytochemistry and floral biology. However, since the *N. amazonum–novogranatensis* clade does not gain high statistical support in the present analysis (JK = 63, BrS = 1, PP = 0.96), any conclusions will have to await substantiation by further data.

Another highly supported relationship within *Nymphaea* is the clade uniting the subgenera *Anecphya* and *Brachyceras*. Several morphological and anatomical characters coincide with this grouping. The most prominent is the incomplete carpellary fusion in both subgenera, which was observed already by Caspary (1865, 1891) and Conard (1905) and led them group these two subgenera together in the so-called "Leptopleura" (Caspary 1865, 1891) or "Apocarpiae" (Conard 1905). However, the previously established hypothesis of the paraphyly of subg. *Brachyceras* with respect to subg. *Anecphya* (Borsch *et al.* subm.) could not be confirmed here. The two sampled representatives of subg. *Brachyceras*, *N. micrantha* and *N. gracilis*, emerge as a monophylum although it is one of the least supported clades in the total evidence trees (JK = 78, BrS = 1, PP = 0.98). Similar to the results of Borsch *et al.* (subm.) the subgenus *Nymphaea* is revealed basal to all other subgenera of *Nymphaea*, but statistical support is not high enough to dispel any doubt about this position (JK = 90, BrS = 2, PP = 0.91). However, a basal position of the hardy, north–hemispheric water lilies would shed new light on the evolution of the genus *Nymphaea*, in which all other members inhabit tropical to subtropical regions.

3.4.3 Taxon sampling

Because previous studies on molecular phylogenetics in the Nymphaeaceae (Borsch 2000, Borsch *et al.* subm.) showed that the hypothesis of a monophyletic genus *Nymphaea* did not gain much statistical support, it appeared necessary to include a more representative taxon sampling of all *Nymphaea* subgenera for the present analysis. By combining multiple chloroplast regions a well–supported hypotheses on intergeneric relationships in Nymphaeaceae could be obtained. The polyphyletic nature of the genus *Nymphaea* was not revealed before simply because other studies failed to sample across the subgenera of *Nymphaea* (e.g., Les *et al.* 1999) or did not include a sufficient amount of characters (Borsch 2000, Borsch *et al.* subm.). Figure 3.4 points out the relevance of judicious taxon sampling: Including only one representative of each genus of Nymphaeales leads to different well–supported topologies, depending on which subgenus of *Nymphaea* was sampled. The correct position of *Ondinea* (according to the total evidence tree) could be inferred only by including *N. elleniae* (subg. *Anecphya*) or *N. micrantha* (subg. *Brachyceras*). This effect is also illustrated by indel no. 192 in Figure 3.5 (occurring in the *matK* gene): if subgenera *Anecphya* and *Brachyceras* had not been sampled, this indel would appear autapomorphic for *Ondinea*, as was the case in the data set of Les *et al.* (1999, but indels have not been analysed there). Thus, the present study well exemplifies that increasing knowledge on the diversity and phylogenetic relationships within a group of taxa can improve study design and thereby also the scientific output of follow–up analyses.

3.4.4 Morphological characters in Nymphaeales

The morphological and anatomical characteristics of Nymphaeaceae and Cabombaceae have been studied intensively since the 1950's (Li 1955, Moseley 1958, 1961, Williamson & Moseley 1989, Schneider *et al*. 1995, Schneider *et al*. 2003). A detailed treatment and evaluation of those characters in this study would be out of scope. However, the new insights into the evolution of Nymphaeales reveal the need for a critical reconsideration of morphological and anatomical characters with special regard to variability among the subgenera of *Nymphaea*. Up to now, variability in *Nymphaea* has not been considered in phylogenetic analyses of Nymphaeacae (Les *et al*. 1999) or Nymphaeales (Li 1955, Ito 1987, Moseley *et al*. 1993). The inferred close relationship of *Nymphaea* to *Victoria–Euryale* but not to *Ondinea* in the study of Les *et al*. (1999), was supported only by floral vasculature characters (vascular supply from the receptacular plexus, source and structure of the petal trace, characters 26–28 in their matrix). Although (Moseley 1961) reports some variability of floral vasculature within *Nymphaea*, complete information on states of anatomical characters in all five subgenera are mostly missing. A critical re–examination of anatomy and morphology might help to substantiate the new hypotheses on the evolution of Nymphaeales. Further possibly informative characters, that are variable in Nymphaeaceae and within *Nymphaea*, are, e.g., structure and number of petiolar and peduncular air canals (Wiersema 1987); seed and pollen surface morphology (Wiersema 1987, Borsch 2000); arrangement and form of stamens and petals (Conard 1905, Wiersema 1987); leaf margins, sclereids, or overall morphology (Conard 1905, Wiersema 1987); and developmental morphology of juvenile plants and seedlings (J. Wiersema, pers. communication).

Furthermore, it is necessary to scrutinize the expression of morphological characters in other basal angiosperms in order to define the plesiomorphic states. Some character states that are currently interpreted as being autapomorphic for, e.g., the Nymphaeaceae including *Nuphar*, such as the presence of lactifers, a compound ovary with laminar placentation, numerous seeds with a distinct apical cap, and numerous petals and stamens, could in fact be plesiomorphic in Nymphaeales. If this was the case, such characters would not contradict a position of *Nuphar* as basal in Nymphaeales or sister to Cabombaceae. However, it might be a challenging task to compare morphological and anatomical traits of Nymphaeales with those of outgroup taxa, since this plant group is characterized by very unique features, mainly as a result of their early separation from the rest of angiosperms and drastic phenotypic changes in the course of their adaptation to the aquatic habit.

3.4.5 Phylogenetic signal of data partitions

The increased application of molecular data in phylogenetics has led to an enormous amount of sequence data sets. Since more and more molecular data is becoming available, a debate on whether information from different genomic regions should be analysed in combination or individually to test phylogenetic hypotheses is emerging (e.g., Bull *et al*. 1993, Cunningham 1997, Castoe *et al*. 2004). Whereas the focus of this debate is on the treatment of heterogeneous data partitions, the combined analysis of homogeneous data is generally accepted (Bull *et al*. 1993, Chippindale & Wiens 1994, De Queiros *et al*. 1996).

There is no apparent conflict among the data partitions in the present analysis, as revealed by partition homogeneity tests as well as partitioned Bremer support. Furthermore, homoplasy is generally low across all partitions. Therefore, the assessment of phylogenetic utility can be confined to the evaluation of sequence divergence, number of informative characters, frequency of length mutations and hotspots, as well as the phylogenetic structure *R* inherent in each data partition.

Highest phylogenetic structure *R*, i.e. the highest average support per node, was observed in the *rpl16* intron, the *trnK* intron and the *matK* gene. These three markers are also characterized by the highest percentages and absolute numbers of informative characters in the present data set. In general, the introns and the *matK* gene provide more information than spacers at this taxonomic level. The spacers are more variable (both in nucleotide substitution as well as in sequence length) but the percentage of informative nucleotide and indel characters is similar or smaller than in introns. This coincides with a lower overall phylogenetic structure (R = 0.65 in spacers vs. R = 0.82 in introns, Table 3.6).

Obviously, the spacers included in the present analysis tend to be too variable for an ordinal taxonomic level. However, a general conclusion on the phylogenetic utility of spacers can not be drawn here, since there can be considerable difference among differently evolving spacers. Transcribed spacers — like the *petB–petD* and the *trnL–trnF* spacer in the present analysis — are, for example, much more conserved than non–transcribed spacers. Differences in information content and phylogenetic utility exist also among the introns investigated here. The introns in *trnL* and *petD*, which have been shown to provide valuable information in a broader taxonomic context (i.e. the basal angiosperms, Borsch *et al.* 2003, Löhne & Borsch 2005, see also Chapter 2 of this thesis), perform below average in Nymphaeales. The *petD* intron so far appears to be the most conserved group II intron in the large single copy region of the cp genome.

The introns in *trnK* and *rpl16*, which are among the best markers in the present analysis, seem to be generally more suitable at this taxonomic level since they have been applied successfully in previous analyses comparable to this study (e.g., Renner & Chanderbali 2000 for rpl16, Müller & Borsch 2005 for trnK). The study of Shaw *et al.* (2005) also revealed the *rpl16* intron as one of the more variable introns, which is true for the present data set as well. In contrast to my results, Shaw *et al.* (2005) designated the *trnK–matK* region as well as *trnT–trnF* as less phylogenetic informative than most of the 21 regions included in their analysis. However, they were searching for suitable markers for analyses at infrageneric levels, which is a completely different goal. The special patterns of molecular evolution in the *matK* gene, that cause the high quantity of information and high quality of characters compared to other coding regions of the chloroplast region (Müller *et al.* in prep.), are the reason for the broad utility of this marker. The *matK* gene has been shown to be more informative than any other single marker at higher taxonomic levels (Hilu *et al.* 2003, Müller *et al.* in prep.), and again in this phylogenetic analysis of Nymphaeales it was one of the most effective markers.

This study adds another piece of evidence to the general phylogenetic utility of non-coding and fast-evolving regions at higher taxonomic levels. Since intense sequence variability and homoplasy is generally confined to mutational hot spots, those markers can easily be

employed in investigations including a broad taxonomic spectrum, and often they provide better resolution and support than rather slowly evolving genes. The increasing number of data sets implementing non–coding and fast–evolving sequences at higher taxonomic levels (e.g., Hilu *et al*. 2003, Quandt *et al*. 2004, 2005, Löhne & Borsch 2005, Qiu *et al*. 2005, see Borsch *et al*. 2005 for review) confirm those markers as promising tools in molecular phylogenetics.

Chapter 4

Phylogenetic relationships among Australian water lilies (*Nymphaea* subg. *Anecphya*)

Contents

4.1 Introduction

Within the family Nymphaeaceae the water lilies (*Nymphaea*) represent the most diverse and most widespread genus, occurring with about 70 species on every continent except Antarctica. The genus *Nymphaea* is traditionally subdivided into five subgenera: *Anecphya*, *Hydrocallis*, *Lotos*, *Brachyceras*, and *Nymphaea* (Caspary 1891, Conard 1905). Three of these subgenera are widespread such as the palaeotropical subg. *Lotos*, the pantropical subg. *Brachyceras* or the northern temperate subg. *Nymphaea*, whereas subg. *Hydrocallis* is restricted to South and Central America and subg. *Anecphya* to Australia (see also Chapter 5 for a detailed description of biogeography).

From these Australian water lilies belonging to subg. *Anecphya* currently seven species are generally recognized: *N. atrans* S.W.L. Jacobs, *N. elleniae* S.W.L Jacobs, *N. gigantea* Hook., *N. hastifolia* Domin, *N. immutabilis* S.W.L. Jacobs, *N. macrosperma* Merr & L.M. Perry, and *N. violacea* Lehm (Jacobs 1992, Jacobs & Porter in prep.). Most species are restricted to the monsoonal parts of the Australian tropics, only *N. gigantea* also occurs south of the Tropic of Capricorn in Queensland and New South Wales (Jacobs & Porter in prep.). Some species appear to be widespread in the monsoonal parts of Australia (*N. immutabilis, N. macrosperma, N. violacea*), whereas others are less widely distributed, e.g. *N. hastifolia* in the higher rainfall areas of tropical Western Australia and Northern Territory, and *N. atrans* and *N. elleniae* on the Cape York Peninsula (northern Queensland). *N. macrosperma*, *N. violacea* and *N. elleniae* also occur in New Guinea. Recently, Jacobs & Hellquist (in press.) described three new species from the Gulf Savannah region of northern Queensland (*N. alexii* S.W.L. Jacobs & Hellq., *N. carpentariae* S.W.L. Jacobs & Hellq., *N. georginae* S.W.L. Jacobs & Hellq.).

Besides *Nymphaea* subg. *Anecphya* a few species from other nymphaean subgenera and another genus of Nymphaeaceae occur naturally in Australia: *Ondinea purpurea* Hartog is endemic to some coastal areas of the Kimberley district in Western Australia. *Nymphaea nouchali* Burm. f. (subg. *Brachyceras*) and *N. pubescens* Willd. (subg. *Lotos*) – both rather widespread paleotropical species – naturally inhabit coastal areas in Northern Territory and northern Queensland (*N. pubescens* only occasionally in Queensland). *Nymphaea caerulea* Savigny ssp. *zanzibarensis* (Caspary) S.W.L. Jacobs (subg. *Brachyceras*), which is originally from Africa, has become naturalised along the east coast in Queensland and New South Wales, and *Nymphaea mexicana* Zucc. (subg. *Nymphaea*) has become naturalised around Brisbane and Sydney (Jacobs 1994, Jacobs & Porter in prep.).

Water lilies of the subgenus *Anecphya* are characterized by clearly emergent, often very large flowers. Some floral characters are shared with the closely related subgenus *Brachyceras*, such as the incomplete fusion of carpel walls ("Apocarpiae" sensu Conard 1905), and the common blue petals. In contrast to subg. *Brachyceras*, members of subg. *Anecphya* do not possess carpellary appendages and, additionally, often show extremely high numbers of stamens (up to 600 in some members, Jacobs & Porter in prep.). Some morphological distinctions can be made among the members of subg. *Anecphya*: one group of species is characterized by a distinctive gap between petals and stamens, rather large seeds and toothed leaf margins (*N. macrosperma, N. georginae, N. gigantea, N.*

carpentariae, N. immutabilis, N. atrans), whereas another group of species is characterized by petals grading into stamens, relatively small seeds and entire to sinuate leaf margins (*N. violacea, N. elleniae, N. hastifolia, N. alexii*). All Australian water lilies of subg. *Anecphya* are day–flowering and possess erect rhizomes.

A close relationship of subgenera *Anecphya* and *Brachyceras* has been demonstrated in recent phylogenetic analyses based on molecular data (Borsch *et al.* subm., Löhne *et al.* subm., and Chapter 3 of this thesis). However, whereas *Anecphya* clearly appears as a well–supported clade in both studies, there is evidence from the chloroplast *trnT–trnF* region and a dense taxon sampling that *Brachyceras* might be paraphyletic with respect to *Anecphya* (Borsch *et al.* subm.). Another, more striking result of these two recent studies is the lack of statistical support for the monophyly of the genus *Nymphaea*. In fact, a comprehensive analysis of Nymphaeales including 12 species of *Nymphaea* and multiple regions of the cp genome depicted the South American genus *Victoria* and the Asian genus *Euryale* as sister to a *Hydrocallis–Lotos* clade, but with rather low support (Löhne *et al.* subm.). Furthermore, combined evidence from the chloroplast genome strongly indicated that the Australian endemic *Ondinea purpurea* is derived from within the Australian water lilies *Nymphaea* subg. *Anecphya* (Borsch *et al.* subm., Löhne *et al.* subm., and Chapter 3 of this thesis).

In the present study, a first examination of phylogenetic relationships in *Nymphaea* subg. *Anecphya* is intended. In order to get a full picture and to unravel possible reticulation events at the species level, the analysis is based on both the nuclear internal transcribed spacer region (ITS) and the chloroplast *trnT-trnF* region. Both markers have been widely used for phylogenetic studies at low taxonomic levels (e.g., Álvarez & Wendel 2003, Shaw *et al.* 2005, and references therein), and have been effective in previous studies on the genus *Nymphaea* (Borsch 2000, Woods *et al.* 2005a, Borsch *et al.* subm.). The present study comprises the first extensive taxon sampling with 47 individuals of different populations, representing all but one (*N. alexii*) of the currently described species and subspecies and the entire range of distribution of subg. *Anecphya* in Australia. The large sample base became available through the extensive collections of S. Jacobs, B. Hellquist and J. Wiersema, which were established during field expeditions in 1997 and 2002. *Ondinea purpurea* is also included in the taxon sampling in order to verify the previous results from the chloroplast genome (see above). Representatives of subg. *Brachyceras* are chosen as outgroup taxa. Based on the described project design, the following questions will be addressed: Can the observed two major groups in *N.* subg. *Anecphya* be revealed by molecular data, and what are the relationships within these groups? Is there any evidence for recent or ancient hybridisation or introgression among the taxa? Can the previously inferred close affinity of *Ondinea purpurea* to the Australian water lilies be reproduced with nuclear and chloroplast sequences and an extensive taxon sampling? Is there any specific taxon in *N.* subg. *Anecphya* that *Ondinea* might be more closely related to?

4.2 Materials and Methods

4.2.1 Taxon sampling and plant material

The data set used in this study comprises 47 representatives of *Nymphaea* subg. *Anecphya* plus 4 members of *Nymphaea* subg. *Brachyceras* as outgroup taxa. The plant material of Australian water lilies was collected during field trips of Surrey Jacobs, Barre Hellquist and John Wiersema in 1997 and 2002. Plant material from the outgroup taxa was collected in North America by Thomas Borsch (1997) or was taken from plants cultivated in Bonn Botanic Gardens. Detailed information on all specimens, including field localities, collectors and vouchers are given in Table 4.1. Most of the sequence data were generated due the course of this thesis. However, about one quarter of the sequences was taken from previous, unpublished studies of Borsch (2000).

4.2.2 DNA isolation, amplification and sequencing

DNA was isolated from silica gel dried leaf tissue or from material conserved in CTAB using the triple extraction method described in Chapter 2.2.2 (after Borsch *et al*. 2003). After chloroform extraction, DNA was precipitated with isopropanol, resuspended in TE, and further purified by ammonium acetate and sodium acetate washing steps followed by ethanol precipitation.

ITS. — The nuclear marker was amplified in a single fragment with the standard primers ITS4 and ITS5 (White *et al*. 1990). This regions spans the internal transcribed spacer 1 (ITS1) between 18S and 5.8S rDNA, the 5.8S rDNA itself, and the internal transcribed spacer 2 (ITS2) between 5.8S and 26S rDNA (see also Figure 4.1). PCR was conducted on a T3 Thermocycler (Biometra, Göttingen, Germany), using *Taq*-DNA-polymerase, buffer and dNTPs from Peqlab (Erlangen, Germany). See Appendix 5 for a detailed description of PCR conditions and reaction mixes. PCR products were purified using a QiaQuick gel extraction kit (QIAGEN Inc., Valencia/CA, USA) and sequenced either with an ABI Prism™ BigDye Terminator Cycle Sequencing Ready Reaction Kit version 1.1 (Applied Biosystems, Foster City/CA, USA) on ABI 310 or 377 automated sequencers, or with a CEQ™ DTCS Quick Start Kit (Beckman Coulter, Fullerton/CA, USA) on a CEQ™ 8000 sequencer. Amplification primers were also used for sequencing both strands, but in some cases these primers were not sufficient to read through the complete fragment because of length polymorphisms in the original DNA (see results). Additional internal primers (NY–5.8F, NY–5.8R, and ITS1–R) were designed to overcome these problems (see Figure 4.1 and Table 4.2).

***trnT–trnF*.** — The chloroplast marker was amplified in two fragments, using primers rps4-5R and trnL110R for the 5′–fragment and primers C and F for the 3′–fragment (see Table 4.2 for primer sequences and authors). Amplification, purification of products, and sequencing followed the same procedures as described for ITS, but see also Appendix 5 for further information. In addition to the amplification primers, two internal sequencing primers (D and E, Table 4.2 and Figure 4.1) were necessary in order to read through a complex microsatellite region (poly–T and poly–TA in the P8 region of the *trnL* intron, see results).

Table 4.1 — Plant material of 51 samples used in present study on *Nymphaea* subg. *Anecphya*. Including information on the collectors (AL = A. Leu, BH = B. Hellquist, JW = J. Wiersema, KWi = K. Wilson, KW = K. Woods, RN = R. A. Novelo, SJ = S. W. L. Jacobs, TB = T. Borsch, VW = V. Wilde), location of voucher specimens, and geographic information on field origin. Coordinates were approximated based on the name of the locality, if not given by the collectors (marked with "~"). Minutes are given in decimal values. Coordinates have not been recorded for outgroup taxa (*).

Species	DNA-Sample Number	Voucher Information			Field Origin					
		Collector	Voucher Number	Herbarium	Region	Locality	Latitude		Longitude	
							degr	min	degr	min
N. atrans S.W.L.Jacobs	NY102	SJ, BH & JW	8212	NASC, NSW, BRI	Queensland	Low Lake, Lakefield National Park	14	38,352	143	54,381
N. atrans S.W.L.Jacobs	NY432	BH & AL	16766	NASC, NSW, BRI	Queensland	Low Lake, Lakefield National Park	14	38,352	143	54,381
N. carpentariae S.W.L Jacobs & C. B. Hellquist	NY396	SJ & BH	8757	NASC, NSW, BRI	Queensland	NE of Normanton	17	31,712	141	82,696
N. carpentariae S.W.L Jacobs & C. B. Hellquist	NY398	SJ & BH	8768	NASC, NSW, BRI	Queensland	73 km W of Normanton on Burketown Road	18	7,362	140	32,249
N. carpentariae S.W.L Jacobs & C. B. Hellquist	NY399	SJ & BH	8770	NASC, NSW, BRI	Queensland	Burketown	17	44,774	139	32,888
N. cf. *carpentariae* S.W.L Jacobs & C. B. Hellquist	NY434	BH & AL	16774	NASC, NSW, BRI	Queensland	Cattle Dam S of Greenvale	19	0,258	145	1,302
N. elleniae S.W.L. Jacobs	NY381	BH & AL	16757	NASC, NSW, BRI	Queensland	Jardine River	11	9,046	142	21,338
N. elleniae S.W.L. Jacobs	NY137	SJ, BH & JW	8227	NASC, NSW, BRI	Queensland	Jardine River	11	9,046	142	21,338
N. elleniae S.W.L. Jacobs	NY103	SJ, BH & JW	8224	NASC, NSW, BRI	Queensland	Bamaga	10	53,604	142	23,237
N. georginae S.W.L Jacobs & C. B. Hellquist	NY425	SJ & BH	8868	NASC, NSW, PERTH	Northern Territory	James River	20	1,517	137	29,55
N. gigantea Hook.	NY126	SJ & BH	8357	NASC, NSW, BRI	Queensland	Condamine River	26	59,08	150	6,69
N. gigantea Hook.	NY426	SJ & BH	8870	NASC, NSW, BRI	Queensland	Chinamans lagoon, 1 km S of Miles	26	40,114	150	11,167
N. gigantea Hook.	NY435	BH & AL	16772	NASC, NSW, BRI	Queensland	Cattle Creek, S of Ingham	18	44,765	146	8,492
N. cf. *gigantea* Hook.	NY395	SJ & BH	8752	NASC, NSW, BRI	Queensland	Reid River	19	45,739	146	50,067
N. hastifolia Domin	NY134	JW & BH	(no voucher)		Northern Territory	Darwin	12	47	130	92
N. cf. *immutabilis* S.W.L. Jacobs	NY450	BH & AL	16767	NASC, NSW, BRI	Queensland	Low Lake, Lakefield National Park	14	38,352	143	54,381

Table 4.1 (continued)

Species	DNA-Sample Number	Voucher Information			Field Origin					
		Collector	Voucher Number	Herbarium	Region	Locality	Latitude		Longitude	
							degr	min	degr	min
N. immutabilis S.W.L. Jacobs	NY121	BH, JW & KB	(no voucher)		Queensland	Cabbage Tree Creek	15	18,19	144	36,97
N. immutabilis S.W.L. Jacobs	NY136	BH, JW & KB	(no voucher)		Queensland	Mt. Molloy-Mareeba Road	~ 16	50	145	20
N. immutabilis S.W.L. Jacobs	NY462	BH & AL	16770	NASC, NSW, BRI	Queensland	Ninds Ck., Innisfail	17	34,263	146	2,298
N. immutabilis S.W.L. Jacobs	NY503	BH & AL	16775	NASC, NSW, BRI	Queensland	Waluma Swamp, Mt. Garnet	17	42,229	145	7,749
N. immutabilis S.W.L. Jacobs	NY383	BH & AL	16760	NASC, NSW, BRI	Queensland	Langi Lagoon, Rokeby	13	27,052	142	42,004
N. immutabilis S.W.L. Jacobs	NY427	BH & AL	16773	NASC, NSW, BRI	Queensland	Ross River, Townsville	19	21,21	146	43,94
N. immutabilis ssp. *kimberleyensis* S.W.L. Jacobs	NY380	SJ & BH	8813	NASC, NSW, PERTH, K, B	Western Australia	Brooking Springs	18	7,248	125	36,121
N. macrosperma Merr. & Perry	NY373	SJ & BH	8802	NASC, NSW, DNA	Northern Territory	Yellow water, Kakadu Nat. Park	~ 12	54	132	31
N. macrosperma Merr. & Perry	NY391	SJ & BH	8796	NASC, NSW, DNA, B, G	Northern Territory	Wildman river, Kakadu Nat. Park	12	34,25	132	13,118
N. macrosperma Merr. & Perry	NY127	BH, JW & KB	16181	NASC, NSW, DNA	Northern Territory	Island Billabong W of road to Ubirr along Magela Creek, N of Jabiru, Kakadu National Park	~ 12	45	132	30
N. cf. *macrosperma* Merr. & Perry	NY418	SJ & BH	8864	NASC, NSW, DNA	Northern Territory	W of Roper Bar	14	42,273	132	27,653
N. cf. *macrosperma* Merr. & Perry	NY433	SJ, BH & KWi	10143	NASC, NSW, PERTH	Western Australia	Perrys Lagoon	15	32,969	128	15,592
N. violacea Lehm.	NY372	SJ & BH	8779	NASC, NSW, DNA	Northern Territory	Kangaroo Creek	16	49,841	137	9,546
N. violacea Lehm.	NY374	SJ & BH	8862	NASC, NSW, DNA	Northern Territory	E of Timber Creek	15	44,014	130	32,573
N. violacea Lehm.	NY405	SJ & BH	8792	NASC, NSW, DNA	Northern Territory	Jabiru Lake	12	40,429	132	50,534
N. violacea Lehm.	NY407	SJ & BH	8799	NASC, NSW, DNA	Northern Territory	Jim Jim Billabong, Kakadu	12	56,538	132	32,076
N. violacea Lehm.	NY409	SJ & BH	8803	NASC, NSW, DNA	Northern Territory	Fogg Dam	12	36,631	131	17,989

Table 4.1 (continued)

Species	DNA-Sample Number	Voucher Information			Field Origin					
		Collector	Voucher Number	Herbarium	Region	Locality	Latitude		Longitude	
							degr	min	degr	min
N. violacea Lehm.	NY410	SJ & BH	8863	NASC, NSW, DNA	Northern Territory	Little Roper River	14	45,583	132	37,166
N. violacea Lehm.	NY110	SJ, BH & JW	8230	NASC, NSW, BRI	Queensland	ca. 22km W of Batavia Downs-Weipa road	12	42,62	142	30,82
N. violacea Lehm.	NY131	BH & AL	16589	NASC, NSW, BRI	Queensland	Mt. Molloy	16	41,188	145	19,572
N. violacea Lehm.	NY382	BH & AL	16759	NASC, NSW, BRI	Queensland	Langi Lagoon, Rokeby	13	27,052	142	42,004
N. violacea Lehm.	NY436	BH & AL	16754	NASC, NSW, BRI	Queensland	Palmer River Roadhouse	16	6,443	144	46,63
N. violacea Lehm.	NY501	BH & AL	16755	NASC, NSW, BRI	Queensland	S of Coen	14	0,288	143	11,436
N. violacea Lehm.	NY502	BH & AL	16765	NASC, NSW, BRI	Queensland	Sweetwater Lake, Lakefield	14	39,813	143	50,315
N. violacea Lehm.	NY504	BH & AL	16779	NASC, NSW, BRI	Queensland	Wonga Beach	16	19,774	145	24,545
N. violacea Lehm.	NY424	SJ & BH	8860	NASC, NSW, PERTH	Western Australia	Hidden Valley, Kununurra	15	46,334	128	45,051
N. cf. *violacea* Lehm.	NY413	SJ & BH	8834	NASC, NSW, PERTH	Western Australia	Apex Creek	17	6,437	125	10,923
N. cf. *violacea* Lehm.	NY419	SJ & BH	8835	NASC, NSW, PERTH	Western Australia	March Fly Glen	17	9,784	125	18,62
N. cf. *violacea* Lehm.	NY420	SJ & BH	8845	NASC, NSW, PERTH	Northern Territory	Drysdale Station	~ 15	34	126	13
N. cf. *violacea* Lehm.	NY448	BH & AL	16761	NASC, NSW, BRI	Queensland	Bobs Lagoon, Rokeby	13	27,305	142	43,603
Ondinea purpurea Hartog	NY377	SJ & BH	8853	NASC, NSW, PERTH	Western Australia	ca. 6 km N of Kalumbaru, Pego Road, Ephemeral Creek	14	15,363	126	37,216
* *N. elegans* Hook.	NY006	TB & VW	3084	FR	USA, Florida, Collier County					
* *N. elegans* Hook.	NY370	TB & KW		BONN, VPI	USA, Texas					
* *N. micrantha* Guill. & Perr.	NY007	Koehnen	s.n.	BONN	Bonn Botanic Gardens Acc.No. 5830 [Zimbabwe]					
* *N.* cf. *nouchali* Burm. f.	NY066	E. Fischer	s.n.	BONN	Bonn Botanic Gardens 14244 [Rwanda]					
* *N. pulchella* DC.	NY100	RN et al.	1295	MEXU	Mexico, Veracruz					

Table 4.2 — Primers used in this study for amplification (A), sequencing (S) or both (A,S). Additionally, the reading direction ("Dir."), i.e. forward (F) or reverse (R), as well as nucleotide sequences and authors of the primers are given. See also Figure 4.1 for position of primers.

Region	Primer	Use	Dir.	Sequence	Author
ITS	ITS4	A,S	R	TCC TCC GCT TAT TGA TAT GC	White *et al.* 1990
	ITS5	A,S	F	GGA AGT AAA AGT CGT AAC AAG G	White *et al.* 1990
	NYITS1–R	S	R	ATA GCA AAG AAG GGA ACC	this study
	NY5.8S–F	S	F	AAG AAC GTA GCG AAA TGC	this study
	NY5.8S–R	S	R	CRC ATA GCT TGA CGC CCA GG	this study
trnT–trnF	rps4–5R	A,S	F	AGG CCC TCG GTA ACG SG	Sauquet *et al.* 2003
	trnL–110R	A,S	R	GAT TTG GCT CAG GAT TGC CC	Borsch *et al.* 2003
	C	A,S	F	CGA AAT CGG TAG ACG CTA CG	Taberlet *et al.* 1991
	F	A,S	R	ATT TGA ACT GGT GAC ACG AG	Taberlet *et al.* 1991
	D	S	R	GGG GAT AGA GGG ACT TGA AC	Taberlet *et al.* 1991
	E	S	F	GGT TCA AGT CCC TCT ATC CC	Taberlet *et al.* 1991

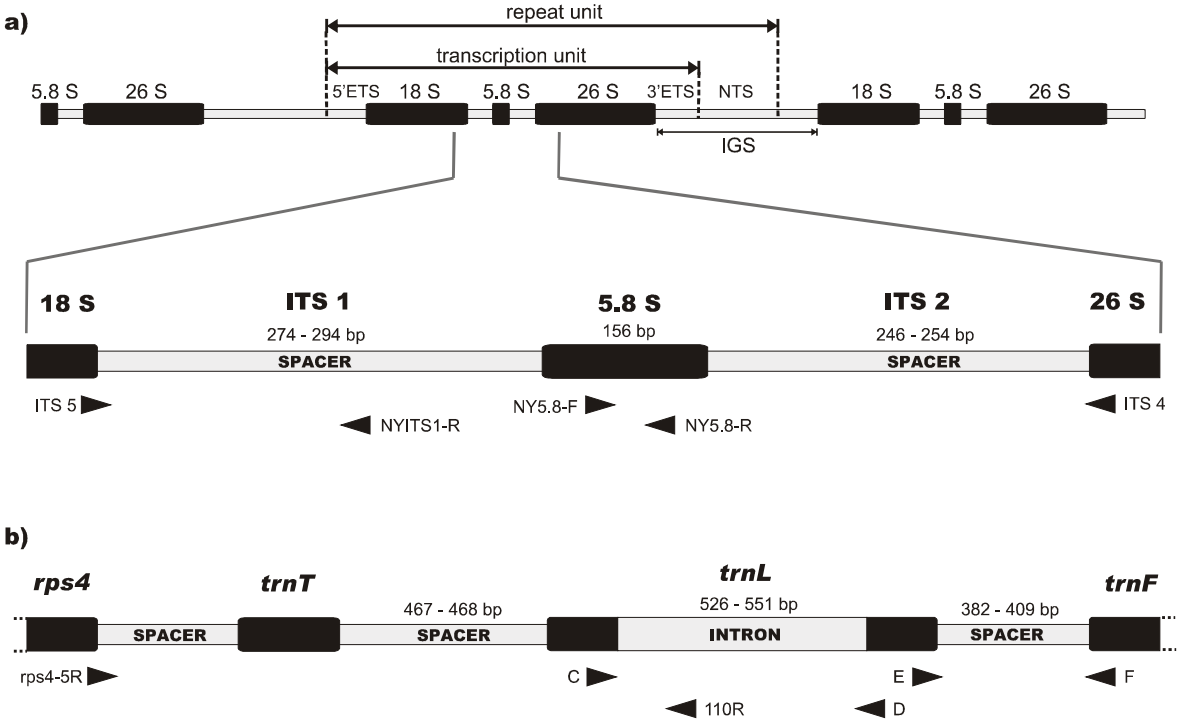


Figure 4.1 — Structure of the genomic markers used for phylogenetic analysis of *Nymphaea* subg. *Anecphya* and position of primers. a) The ITS (internal transcribed spacer) region is part of a larger repeat unit of the nuclear genome comprising the ribosomal 18S, 5.8S and 26S DNA, the intergenic spacer (IGS) between 26S and 18S, and the internal transcribed spacers between 18S and 5.8S (ITS1) and between 5.8S and 26S (ITS2). Parts of the IGS are also transcribed (ETS = external transcribed spacer, NTS = non-transcribed spacer). **b)** the *trnT–trnF* region of the chloroplast genome. Coding stretches are marked by black boxes; non-coding regions are grey. Arrows mark the position and reading direction of primers used in this study (for reference see Table 4.2). Length ranges observed in the present data set are given above the boxes (see also Table 4.3 and 4.4).

4.2.3 Sequence alignment and phylogenetic analysis

Both genomic regions were aligned manually using BioEdit version 7.0.5 (Hall 1999), following the rules outlined in Chapter 2. Mutational hotspots (see Chapter 2 and Borsch *et al*. 2003 for definition), the *trnL*–exons and the incomplete edges of the matrices were excluded from analysis (for explanation see Chapter 3.2.3, page 58). Indels were coded automatically in a "01"–matrix with SeqState version 1.25 (Müller 2005), applying the "simple indel coding" strategy after Simmons & Ochoterena (2000). Indel matrices of *trnT–trnF* and ITS were appended to the respective sequence alignment for a combined analysis. Both, the ITS and the *trnT–trnF* data set, were analysed separately including either only substitutions or substitutions + indels. Combined analyses of the nuclear and the chloroplast data sets were not conducted because of apparent incongruences in the single–marker trees. Sequence statistics, such as G/C content and sequence divergence (uncorrected p distance) were calculated using SeqState.

Maximum Parsimony. — MP analyses were conducted with PAUP* 4.0b10 (Swofford 2002) and PRAP (Müller 2004), which generates command files for PAUP* that allow parsimony ratchet searches (Nixon 1999). Twenty random addition cycles of 200 ratchet iterations were applied. Iterations comprised two rounds of TBR branch swapping each, one on a randomly reweighted data set (25% of the positions) and the other on the original matrix, saving one shortest tree. Shortest trees collected from the different tree islands were used to compute a strict consensus tree. The four representatives of *Nymphaea* subg. *Brachyceras* were chosen as outgroup to root the trees. Node support was estimated through jackknifing (JK) 10,000 replicates (simple addition, keeping 1 tree per replicate, deleting 36.788% of characters in each replicate). Additionally, Bremer support was calculated using PAUP* and PRAP version 1.21 (10 random addition replicates per constraint tree, parsimony ratchet employed, Müller 2004).

Bayesian Inference. — Modeltest version 3.06 (Posada & Crandall 1998) was used to determine the best models of molecular evolution for ITS and *trnT–trnF*. In both cases the GTR + G model was selected according to the Akaike information criterion. Bayesian analyses of the total evidence data sets (substitutions + indels) were performed using MrBayes version 3.1 (Ronquist & Huelsenbeck 2003), with the binary (restriction site) model applied to the indel partition. Analysis was performed for 1,000,000 generations (settings: MCMCMC, 4 runs with 4 chains each, saving one tree every 100 generations). In the case of the ITS 44–taxon data set tree probabilities had converged to a stable value after 10,000 generations. Thus, the burn–in was set to 100 and 39,603 trees were sampled for calculating the consensus tree. In the four runs of the *trnT–trnF* 51–taxon analysis probabilities had converged after 20,000 generation or earlier, therefore burn–in was set to 200 and 39,204 trees were sampled.

4.2.4 Biogeographic data

Coordinates of field localities from all sampled specimens were digitized using the geographic information software ArcView 3.2., and localities were plotted on a map of Australia. If coordinates were not given by the collectors they were approximated based on the name of the locality (marked by "~" in Table 1). For *Ondinea purpurea* and *Nymphaea*

hastifolia a distribution map was generated based on field records of these species. Distribution data were available from the "Australia's Virtual Herbarium" online resource http://www.chah.gov.au/avh/), which provides information for 37 specimens of *Nymphaea hastifolia* and 64 specimens of *Ondinea purpure* deposited in Australian Herbaria (New South Wales, Northern Territory, Western Australia, Tasmania, Canberra, Victoria, Queensland, and South Australia), including downloadable tables with latitude–longitude data. The date of data accession was December, 8th 2004.

4.3 Results

4.3.1 Sequence variability in ITS

The ITS data set comprised in total 695 characters of which 40% were obtained from ITS1, 38% from ITS2 and 22% from the 5.8S rDNA. The length variability among the sequences is relatively low: Several length mutations of 1–3 nucleotides occur in ITS1 and 2, which were coded to 35 indel characters in total. The 5.8S rDNA has a conserved length of 156 nucleotides in all sequences. Sequence divergence is highest in ITS1 (8.0%), lower in ITS2 (5.9%) and extremely low in 5.8S (0.6%, Table 4.4). In most of the samples a stretch up to 14 "A"s in ITS1 caused problems in sequencing due to slippage. Thus, internal primers annealing to the respective opposite strands had to be used to complete the sequences. This highly length–variable microsatellite was present in all ingroup samples examined except the individuals found in the SS–I clade (Fig. 4) plus the putative *N. violacea* hybrids NY382, NY409, NY424, NY501, NY502 (see below). This microsatellite region (alignment pos. 172–188) was excluded from phylogenetic analysis, because such microsatellites normally show a high rate of insertions and deletions leading to uncertain homology assessment and spurious signal in reconstructing phylogenetic relationships of more distant species.

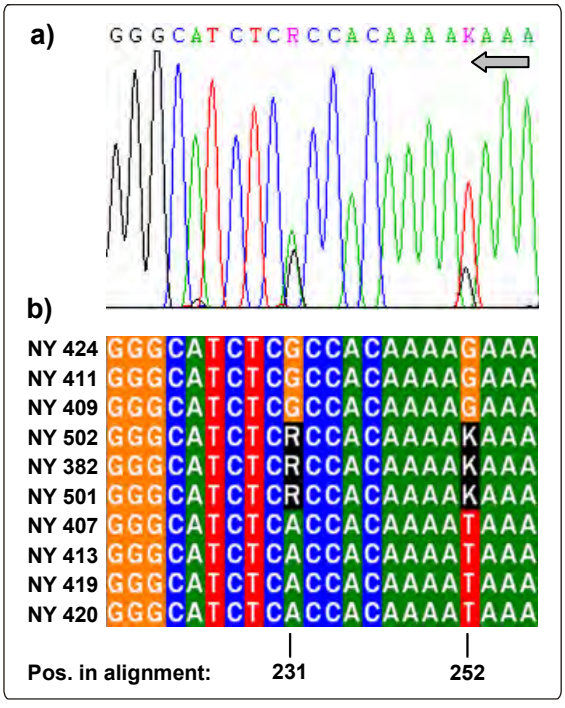


Figure 4.2 — Polymorphic nucleotide sites in ITS.

a) Section of the pherogram of the NY5.8–R primer in taxon NY382 (*Nymphaea violacea*). The arrow indicates the reading direction of the primer. Two sites with double peaks are observable.

b) Section of the ITS sequence alignment showing the same positions as in a). The respective nucleotide positions (231, 252) are marked with ambiguity codes (R= A or G, K= G or T) for those sequences in which double peaks occur.

Polymorphisms in ITS. — In several of the examined specimens of *Nymphaea* subg. *Anecphya* there were polymorphic sites, hinting at divergent ITS paralogues. In 29 sequences additive signal, i.e. double peaks in the pherograms, could be observed (see illustrated example in Figure 4.2). In 22 of these sequences polymorphisms occurred at parsimony informative sites and 14 sequences contained more than one polymorphic site (these taxa are listed in Table 3). Polymorphic nucleotide sites were coded with IUPAC ambiguity codes in the sequence alignment (Figure 4.2, b). However, polymorphisms among the paralogues were not restricted to single nucleotide positions. Additionally, length differences were present. One of these length polymorphisms was due to different repeat numbers of an otherwise conserved "AG"–motif in ITS2 (alignment pos. 503–509). In most individuals 4 AG-repeats are present whereas five samples of *Nymphaea violacea* (NY131, NY372, NY405, NY436, NY504) possess only 3, leading to a 2 bp gap in the alignment. There are some additional individuals of *Nymphaea violacea* (NY382, NY409, NY424, NY501, NY502; Figure 4.3) in which length polymorphisms at the respective position could be observed. Apparently, this is caused by different ITS paralogues possessing either 3 or 4 AG-repeats. These length polymorphisms as well as the numerous nucleotide polymorphisms among paralogues point to hybridisation events between different subclades of *Nymphaea* subg. *Anecphya*, which must have occurred so recently that they are not yet completely obscured by concerted evolution in the nrDNA repeat unit (see also section 4.3.3 and discussion).

Table 4.3 — Taxa with polymorphic nucleotide sites and length polymorphisms in ITS. Only those samples with 2 ore more polymorphic nucleotides at parsimony informative sites are listed (see text for further explanations). A question mark indicates that no polymorphisms have been observed, but the respective sequence was incomplete due to sequencing problems related to length polymorphisms. The first set of 7 samples was excluded from the data set for 44–taxa analyses, all 14 samples were excluded for the 36–taxa analyses.

Taxon (DNA sample number and species name)		Number of polymorphic nucleotide sites (ITS1 / 5.8S / ITS2)	Number of length polymorphisms (ITS1 / 5.8S / ITS2)
NY427	*N. immutabilis*	12 (7/0/5)	0
NY380	*N. imm.* ssp. *kimberleyensis*	4 (?/1/3)	2 (1/0/1)
NY382	*N. violacea*	8 (8/0/?)	1 (0/0/1)
NY409	*N. violacea*	6 (6/0/?)	1 (0/0/1)
NY424	*N. violacea*	8 (3/0/5)	2 (1/0/1)
NY501	*N. violacea*	8 (4/0/4)	2 (1/0/1)
NY502	*N. violacea*	7 (7/0/?)	3 (2/0/1)
NY425	*N. georginae*	2 (0/0/2)	0
NY126	*N. gigantea*	4 (2/0/2)	0
NY395	*N.* cf. *gigantea*	2 (0/0/2)	0
NY426	*N. gigantea*	4 (2/0/2)	0
NY435	*N. gigantea*	4 (2/0/2)	0
NY418	*N.* cf. *macrosperma*	4 (0/0/4)	0
NY110	*N. violacea*	3 (1/0/2)	0
total number		38 (20/1/17)	5 (3/0/2)

Further length polymorphisms were observed in ITS1 of the *N. violacea* samples NY424, NY501, and NY502 (with 2 length–polymorphic sites in the latter taxon). In the sequence of *N. immutabilis* ssp. *kimberleyensis* (NY380) length polymorphisms also occurred in both ITS1 and ITS2. In general, all samples in which length polymorphisms were observed showed also a high frequency of ambiguous nucleotide sites. In the case of sample NY427 (*N. immutabilis*) no length polymorphisms occurred, but the pherograms showed double peaks for 12 parsimony informative sites. To test the effect of polymorphic sites on phylogenetic inference, analyses were run with either the complete taxon sampling (51 sequences) or with data sets reduced by those sequences with frequent polymorphisms (44 sequences, after excluding NY380, NY382, NY409, NY424, NY427, NY501, and NY502). Whereas the seven above mentioned individuals exhibited a high number of polymorphisms suggesting hybridisation among rather distant genomes, there were other individuals of potential hybrid origin but with only very few polymorphic sites. These individuals with few (two to four) polymorphic sites belonged to *N. georginae* (NY425), *N. gigantea* (NY126, NY395, NY426, NY435), *N.* cf. *macrosperma* (NY418), and *N. violacea* (NY110). Their inclusion in the matrix also had no significant effect on the reconstructed trees (see below) as was elucidated by a matrix further reduced to 36 taxa that excluded all before mentioned samples.

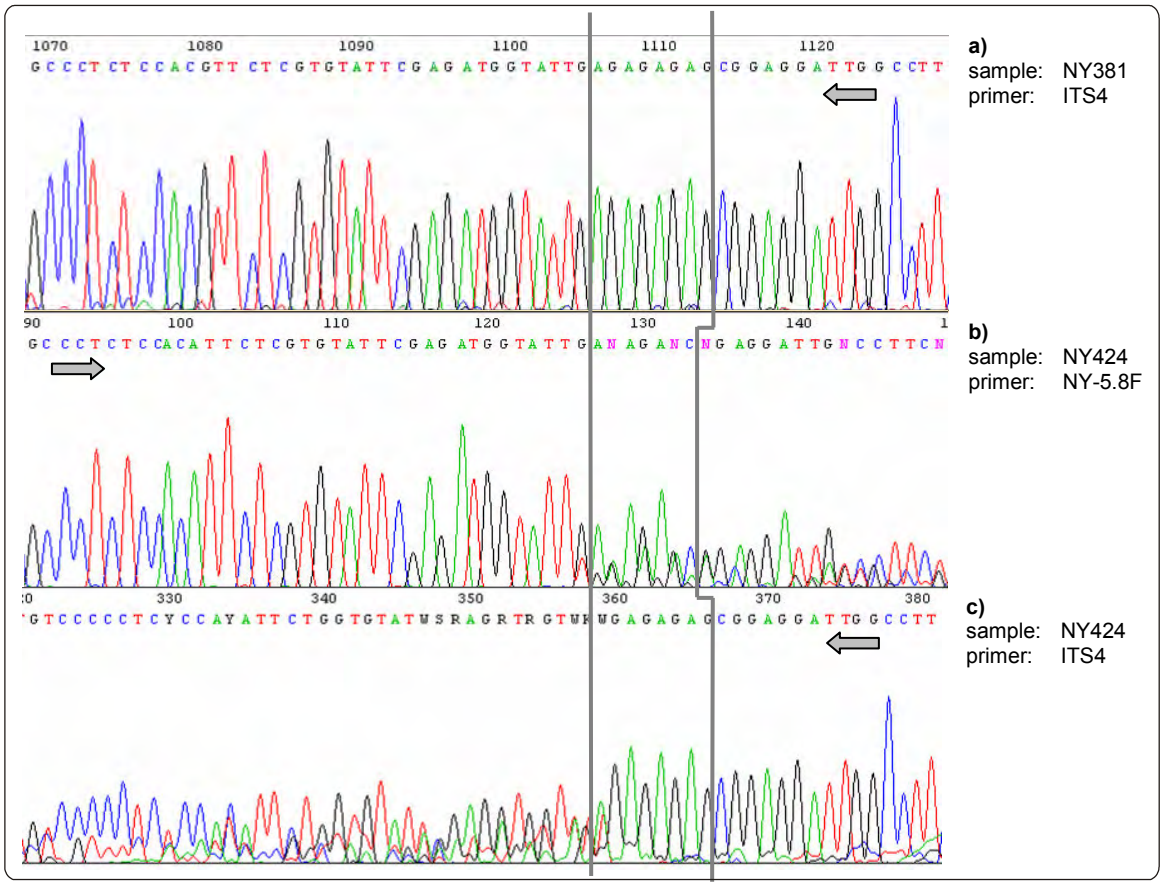


Figure 4.3 — Length polymorphism observed in ITS2 sequences of several samples. The respective section (an "AG"–repeat) is marked by vertical lines. Arrows indicate the reading directions of the primers. a) Section of the pherogram of the ITS4 primer in taxon NY381 (N. elleniae). This is a "normal" sequence with 4 "AG" units. b, c) Pherogram sections of the forward primer NY–5.8F and the reverse primer ITS4 in taxon NY424 (N. violacea). Breakdown of the signal and overlapping peaks after the respective "AG"–repeat indicate the presence of templates differing in the length of the "AG"–repeat.

4.3.2 Sequence variability of the *trnT–trnF* region

The *trnT–trnF* data set comprises 1465 characters. The largest partition is the *trnL* intron (568 characters, 39%), followed by the *trnT–trnL* spacer (32%) and the *trnL–trnF* spacer (29%, Table 4.4). Length variability was almost exclusively restricted to the *trnL* intron and the *trnL–trnF* spacer, whereas only one single–nucleotide indel is found in *trnT–trnL* spacer. Within the *trnL* intron length mutations occurred in an "AT"–rich section (Figure 4.8, on page 98) that corresponds to the P8 stem loop in the *trnL* intron secondary structure (Borsch *et al.* 2003, Quandt *et al.* 2004). The majority of all indels were simple sequence repeats involving 4 to 27 nucleotides. A mutational hot spot close to the 3' end of the *trnL–trnF* spacer (poly–A stretch, alignment pos. 1875–1884) was excluded from the analysis. Another hot spot (poly-T) in the P8 region of the *trnL* intron was also excluded from analysis. The average sequence divergence is relatively low in the overall *trnT–trnF* data set (0.6%), although it is somewhat higher in the *trnT-trnL* spacer (0.8%, Table 4.4). In total, the *trnT–trnF* data set provides 42 informative characters (substitutions and indels).

Table 4.4 — Characteristics of ITS and *trnT-trnF* sequences in the *Nymphaea* subg. *Anecphya* data set (including 4 outgroup taxa of *Nymphaea* subg. *Brachyceras*). "Mean length" – mean of all observed sequence lengths (SD in brackets); "Length range" – actual sequence length in nucleotides (minimal and maximal observed value, including hotspots), "No. Char." – number of characters in the alignment matrices (excluding hotspots), "% Div." – pairwise sequence distance in percent (uncorrected p distance, overall mean, SE in brackets); "% GC" – GC content in percent; "Var.", number of variable characters (percentage in brackets); "Inf." –number of parsimony informative characters (percentage in brackets); "indels" – number of length mutations that were coded by SeqState (number of informative indels in brackets).

Genomic Region	Mean Length	Length Range	Hot Spots	No. Char.	% Div. (SE)	% GC (SD)	Var. (%)	Inf. (%)	Indels (inf.)
ITS (total)	684 (3.4)	681-700	1	695	5.5% (0.5)	49.3 (0.8)	175 (25%)	135 (19%)	35 (23)
ITS 1	280 (3.2)	274-294	1	277	8.0% (1.0)	47.3 (0.9)	92 (33%)	76 (27%)	18 (12)
5.8 S	156 (0.1)	156	0	156	0.6% (0.3)	55.1 (0.2)	11 (7%)	3 (2%)	0 (0)
ITS 2	249 (2.0)	246-254	0	261	5.9% (0.7)	47.9 (1.7)	72 (28%)	56 (21%)	17 (11)
***trnT–trnF* (total)**	1400 (12.7)	1377 -1414	2	1464	0.6% (0.1)	35.4 (0.3)	65 (4%)	35 (2%)	14 (7)
trnT–trnL	468 (0.1)	467-468	0	468	0.8% (0.2)	37.0 (0.6)	22 (5%)	14 (3%)	1 (0)
trnL intron	539 (8.9)	526-551	1	568	0.5% (0.1)	35.2 (0.7)	27 (5%)	13 (2%)	6 (4)
trnL–trnF	393 (10.9)	382-409	1	428	0.5% (0.2)	34.0 (0.5)	16 (4%)	8 (2%)	7 (3)

4.3.3 Trees obtained from nuclear ITS sequences

Figure 4.4 shows the strict consensus of 40 trees yield by maximum parsimony analysis of the ITS 44–taxon data set. Bayesian analysis (Figure 4.5) revealed basically the same topology. There is only a single, hardly supported node in each of the trees that is not present in the respective other tree (marked with an asterisk in both figures). The topology obtained from the 51–taxon set (including hybrid taxa) is less resolved and much less supported (Figure 4.9, see also Table 4.5 for tree statistics and comparison of node support). The inclusion or exclusion of indel characters had, on the other hand no effect on the inferred topologies, but generally the support values are higher if indels are included (Table 4.5).

Both, MP and Bayesian analyses of the ITS data set reveal a major subdivision of *Nymphaea* subg. *Anecphya* into two clades: one clade (node 1 in Figure 4.4) comprises *N. macrosperma*, *N. georginae*, *N. gigantea*, *N. carpentariae*, *N. atrans,* and *N. immutabilis*, and a second clade (node 4 in Figure 4.4) comprises *N. violacea*, *N. elleniae*, *N. hastifolia* as well as *Ondinea purpurea*. These two clades, which are separated by relatively long branches (Figure 4.5), correspond to a grouping suggested by morphological characters such as seed size. Therefore, these two groups are in the following referred to as the "large–seeded" *Anecphya* clade (*N. macrosperma*, *N. gigantea* etc.) and the "small–seeded" *Anecphya* clade (*N. violacea*, etc.). Among the "large–seeded" (LS) species samples further group into two clades, one comprising *N. macrosperma*, *N. georginae*, *N. gigantea*, and *N. carpentariae* ("LS–I", node 2 in Figure 4.4), and another comprising *N. atrans* and *N. immutabilis* ("LS–II", node 3). The LS–II clade gains maximum support in all trees (irrespective of taxon sampling differences). In contrast, support for the LS–I clade is low in the 51–taxa analysis (JK = 52, PP = 0.88, BrS = 1), high in the 36–taxa analysis (JK = 99, PP = 0.95, BrS = 3, Figure 4.4), and intermediate in the 44–taxon analysis (Figure 4.4 and Figure 4.5). The LS–I and LS–II groups are not further resolved internally, and ITS sequences do not allow to recognize species within these clades. There is only weak indication that NY425 (*N. georginae*) and NY395 (*N.* cf. *gigantea*) might be more closely related to each other than to other plants (JK = 63, PP = 0.95, BrS = 1). Similar to the large–seeded clade, there are also two well–supported groups within the small–seeded (SS) lineage of *Anecphya*. Subclade SS–I (node 5, JK = 100, PP = 1.0, BrS = 11) comprises all samples of *N. elleniae* and six of the *N. violacea* samples. Subclade SS–II (node 6, JK = 100, PP = 1.0, BrS = 7) comprises all other *N. violacea* samples as well as *N. hastifolia* and *Ondinea purpurea*. The sister group relationship of the latter two taxa gains high statistical support (JK = 100, PP = 1.0, BrS = 11). Two samples of *N. violacea* (NY110, NY448) appear as a well supported clade, separated from the other *N. violacea* samples in the SS–II group. Generally, resolution is poor within the subclades largely due to the sampling of several individuals from different localities of the same putative species. The 50% majority rule consensus (tree not shown) is identical to the strict consensus tree (Figure 4.4).

Closer examination of sequences and pherograms of those individuals with a high frequency of polymorphic sites (not included in the 44–taxon data set) showed that paralogues present in these individuals are a mixture of sequences otherwise occurring in two different subclades of subg. *Anecphya*. In the case of the five potential hybrids identified as *N. violacea* (NY382, NY409, NY424, NY501, and NY502) the polymorphic sites are composed

of character states present in the SS–I and the SS–II group. The putative hybrid in *N. immutabilis* (NY427) shares character states among the LS–I and the LS–II groups, whereas ITS from *N. immutabilis* ssp. *kimberleyensis* shows sites characteristic of the LS–I and the SS–I clades. The latter might therefore result from hybridisation between the large–seeded and the small–seeded clades of subg. *Anecphya*.

Table 4.5 — Results of phylogenetic analyses of the ITS data set using Maximum Parsimony (MP) and Bayesian (BI) approaches. Both, MP and BI have been conducted for the 51–taxon data set, as well as for data sets reduced to 44 and to 36 taxa, respectively. For the 44–taxon data set the following samples were excluded: NY382, NY409, NY424, NY501, NY502 (all *N. violacea*), NY427 (*N. immutabilis*) as well as NY380 (*N. immutabilis* ssp. *kimberleyensis*). See text for explanations. For the 36–taxon data set the following taxa were excluded additionally: NY110 (*N. violacea*), NY418 (*N.* cf. *macrosperma*), *N. georginae* (NY425) and all samples of *N. gigantea* (NY126, NY395, NY426, and NY435). For MP, analyses of only substitutions (first line) and substitutions + indels (second line) were run separately. For BI, only the complete data sets (substitutions + indels) were analysed.

		MP 51 taxa	MP 44 taxa	MP 36 taxa	Bayesian 51 taxa	Bayesian 44 taxa	Bayesian 36 taxa
a)	***Tree statistics***						
	Number of characters	695 *730*	695 *730*	695 *730*	730	730	730
	Number of trees	30 *29*	40 *40*	30 *29*	39604	39603	37802
	Tree length	248 *288*	242 *281*	222 *255*	–	–	–
	CI	0.860 *0.863*	0.847 *0.851*	0.860 *0.863*	–	–	–
	RC	0.831 *0.834*	0.817 *0.821*	0.831 *0.834*	–	–	–
b)	***Bootstrap values, Bremer support (in brackets), and posterior probabilities***						
	Node 1 (large–seeded Anecphya clade)	82 (1) *96 (2)*	100 (8) *100 (12)*	*100 (7)* *100 (11)*	0.99	1.00	1.00
	Node 2 (LS-I group)	52 (1) *51 (1)*	80 (2) *84 (2)*	99 (3) *99 (3)*	0.88	0.72	0.95
	Node 3 (LS-II group)	100 (1) *100 (9)*	100 (9) *100 (10)*	100 (11) *100 (11)*	0.99	1.00	1.00
	Node 4 (small–seeded Anecphya clade)	83 (1) *96 (12)*	100 (13) *100 (16)*	100 (13) *100 (16)*	1.00	1.00	1.00
	Node 5 (SS-I group)	57 (1) *58 (1)*	100 (10) *100 (11)*	100 (8) *100 (9)*	0.87	1.00	1.00
	Node 6 (SS-II group)	54 (1) *61(1)*	99 (6) *100 (7)*	98 (5) *99 (6)*	1.00	1.00	1.00

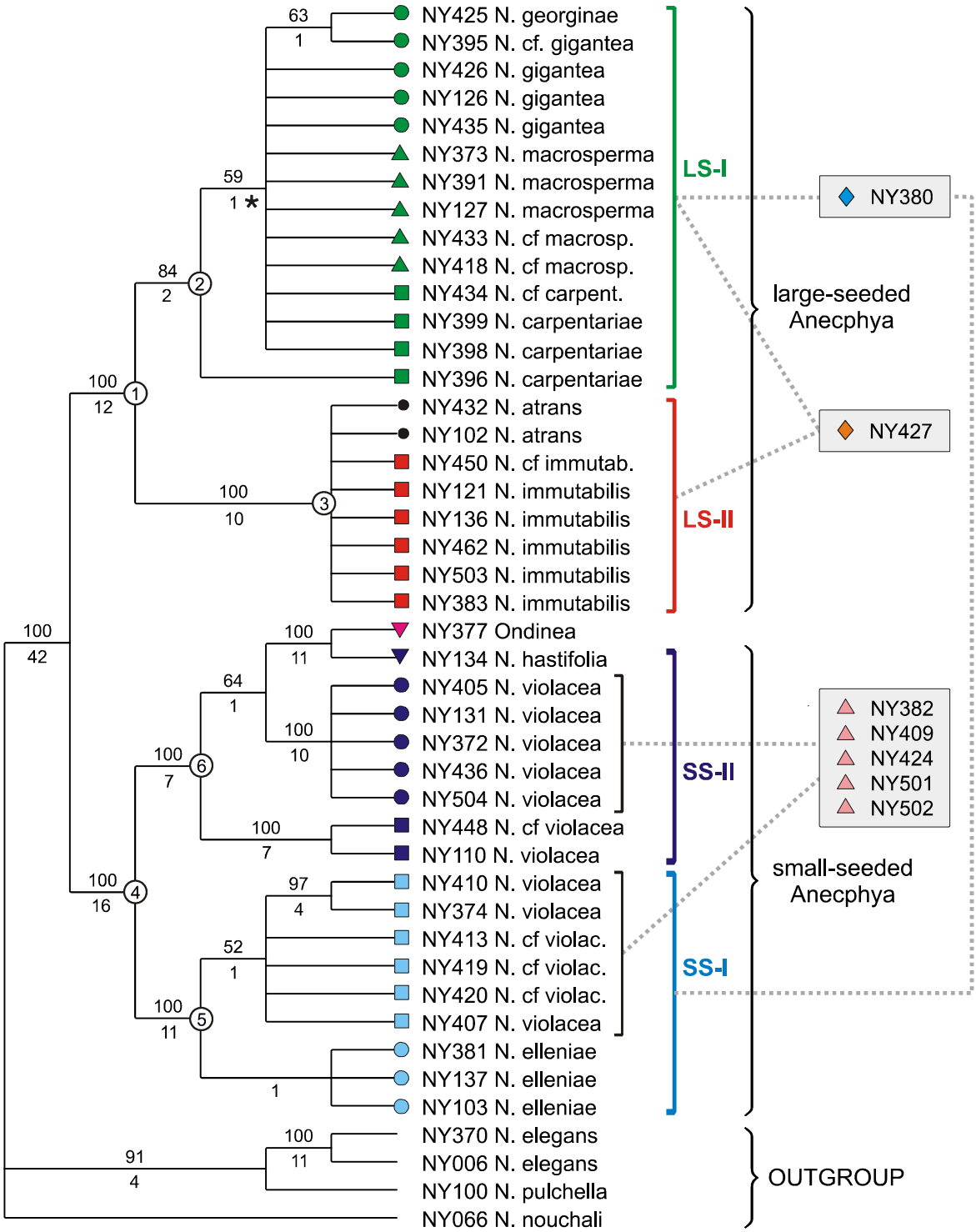


Figure 4.4 — Strict consensus of 40 shortest trees obtained from MP analysis of ITS sequence data from 44 taxa (substitutions + indels). Jackknife values (if higher than 50) are given above the branches, Bremer support below. Coloured symbols correspond to the ones used in the *trnT–trnF* tree (Figure 4.6) and in the distribution maps (Figure 4.10 and Figure 4.11). Nodes relevant for discussion are numbered (in circles, see also Table 4.5). The asterisk marks the node that is not present in the Bayesian tree (Figure 4.5). The specimens in grey boxes have not been included in the 44–taxon analysis because their sequences are characterized by several nucleotide and length polymorphisms. Dotted lines indicate their affinities to subgroups of *Nymphaea* subg. *Anecphya*, based on visual examination of the pherograms and character states in the alignment. See also Figures 4.2 and 4.3 and explanations in the text.

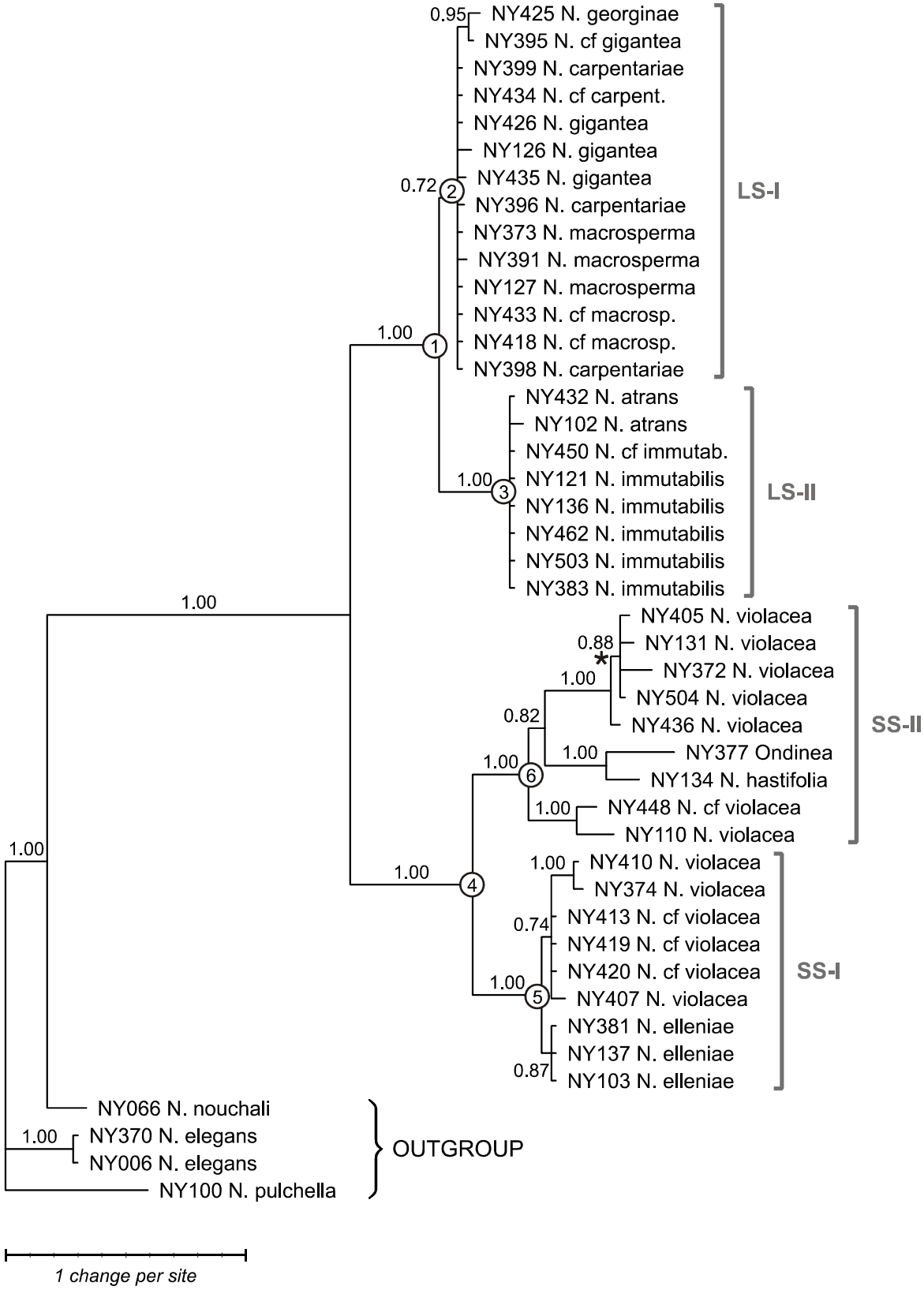


Figure 4.5 — Phylogram of phylogenetic relationships in *Nymphaea* subg. *Anecphya* inferred from ITS. The tree is a 50% majority rule consensus of 39,603 trees obtained from four runs of Bayesian analysis of the 44–taxon data set (substitutions + indels). Numbers above branches indicate posterior probabilities of the respective nodes. Nodes relevant for discussion are numbered (in circles, see also Table 4.5). The asterisk marks the node that is not present in the MP tree (see Figure 4.4). Besides this node both trees are identical.

4.3.4 Trees obtained from the plastid marker *trnT–trnF*

Bayesian and MP analyses of the *trnT-trnF* sequences were conducted in parallel to ITS using the complete taxon sampling (51 sequences) and data sets reduced to 44 and 36 taxa, respectively. In contrast to ITS, taxon sampling had no effects on the inferred topologies and only minor effects on branch support (Table 4.6). Thus, only the full evidence trees (51 taxa, substitutions + indels) are discussed in the following, but see Table 4.6 for statistics of the other trees. Parsimony (Figure 4.6) and Bayesian analyses (Figure 4.7) yielded congruent topologies. There is only one node resolved in addition in each of the respective trees (marked with an asterisk in the respective figures).

The plastid marker did not resolve the large–seeded and small–seeded clades. Instead, four major clades appear in a polytomy. Two of these clades correspond to the large–seeded groups LS–I and LS–II as inferred from ITS. The other two clades contain only small–seeded species, but the composition of the two clades is different from the SS–I and SS–II groups found with ITS. Resolution within the LS–I and LS–II clades is poor, although many of the individuals, even those assigned to the same species, possess autapomorphic mutations. This is also evident from the Bayesian phylogram (Figure 4.7). In LS–I most samples of *N. macrosperma* (NY121, NY 373, NY391, NY433) form a clade together with one individual identified as *N. carpentariae* (NY398) and one as *N. georginae* (NY425; JK = 84, PP = 1.0, BrS = 2).

The coloured symbols used in the MP trees of ITS (Figure 4.4) and *trnT–trnF* (Figure 4.6) clearly illustrate that the nuclear and the chloroplast genomic partitions provide incongruent signal regarding the relationships among the small–seeded species of *Anecphya*. Representatives of the SS–I and SS–II groups (as inferred from ITS) are completely intermingled in the *trnT–trnF* tree (Figure 4.6). Contrary to the ITS tree, chloroplast sequences of *Nymphaea elleniae* (NY103, NY137, NY381) group with two samples of *N. violacea* (NY448, NY110). Three samples designated as *N. cf. violacea* (NY413, NY419, NY420) appear as a well–supported clade (JK = 95, PP = 1.0, BrS = 3) in the chloroplast tree, which was not resolved with ITS. In the other small–seeded clade the two *N. violacea* samples NY405 and NY407 are inferred to be sisters, although these samples emerge in different subclades of the ITS tree (NY405: SS–II, NY407: SS–I). Another well-supported *trnT-trnF* subclade comprises all remaining SS–II representatives (NY131, NY372, NY436, NY504) as well as two members of the ITS SS–I clade (NY374, NY410) and the five samples of *N. violacea* that were not included in the 44–taxa analyses because they were considered as hybrids (NY382, NY409, NY424, NY501, NY502). The other two samples that were not included in the 44–taxa analyses, NY380 (*N. immutabilis* ssp. *kimberleyensis*) and NY427 (*N. immutabilis*), appear in the *trnT-trnF* trees in the well–supported LS–I clade.

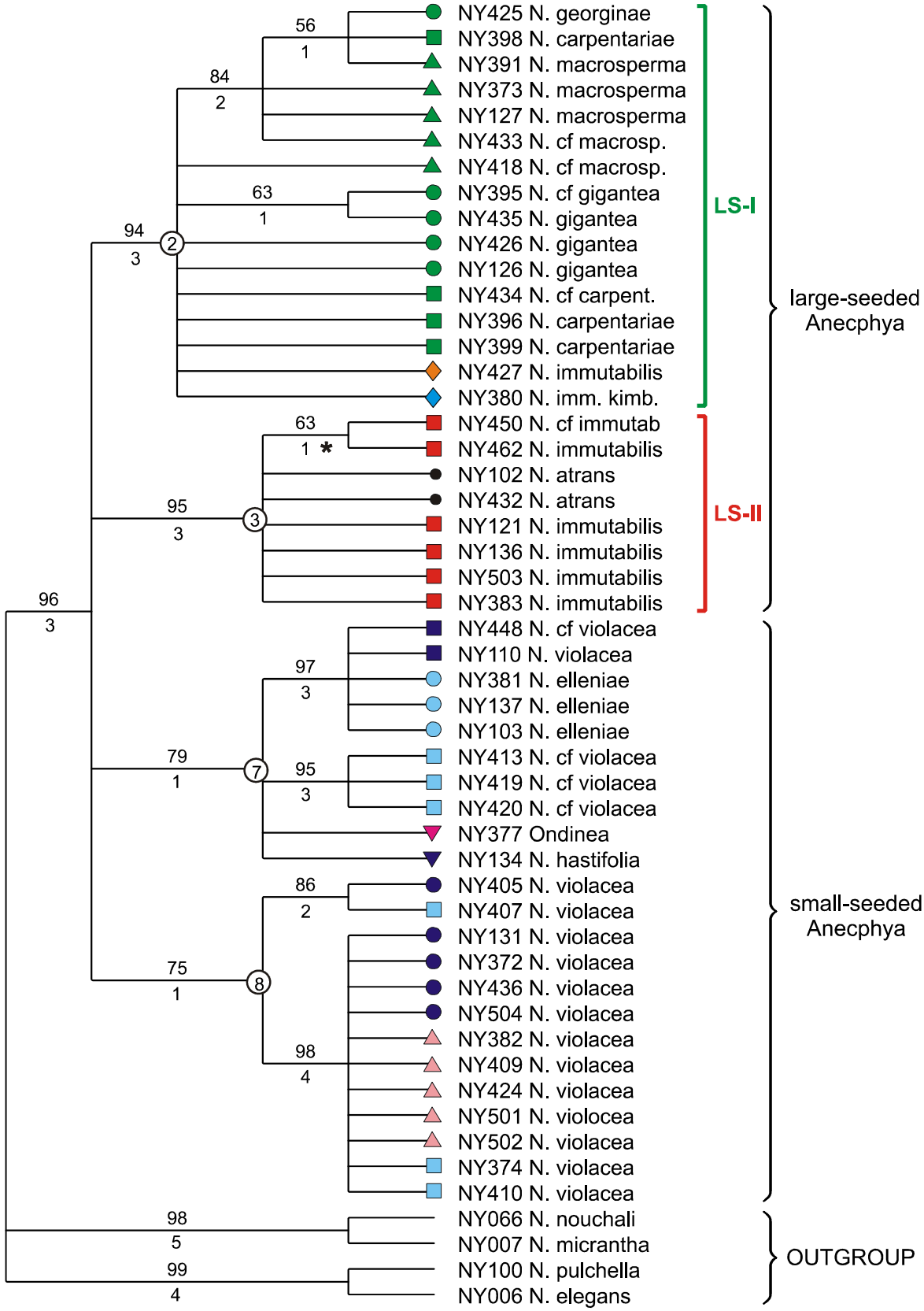


Figure 4.6 — Strict consensus of 44 shortest trees obtained from MP analysis of *trnT–trnF* sequence data from the complete data set (51 taxa, substitutions + indels). Jackknife values are given above the branches, Bremer support below. Coloured symbols correspond to the ones used in the ITS tree (Figure 4.4) and in the distribution maps (Figure 4.10 and Figure 4.11). Nodes relevant for discussion are numbered (in circles, see also Table 4.6). The asterisk marks the node that is not present in the Bayesian tree (Figure 4.7).

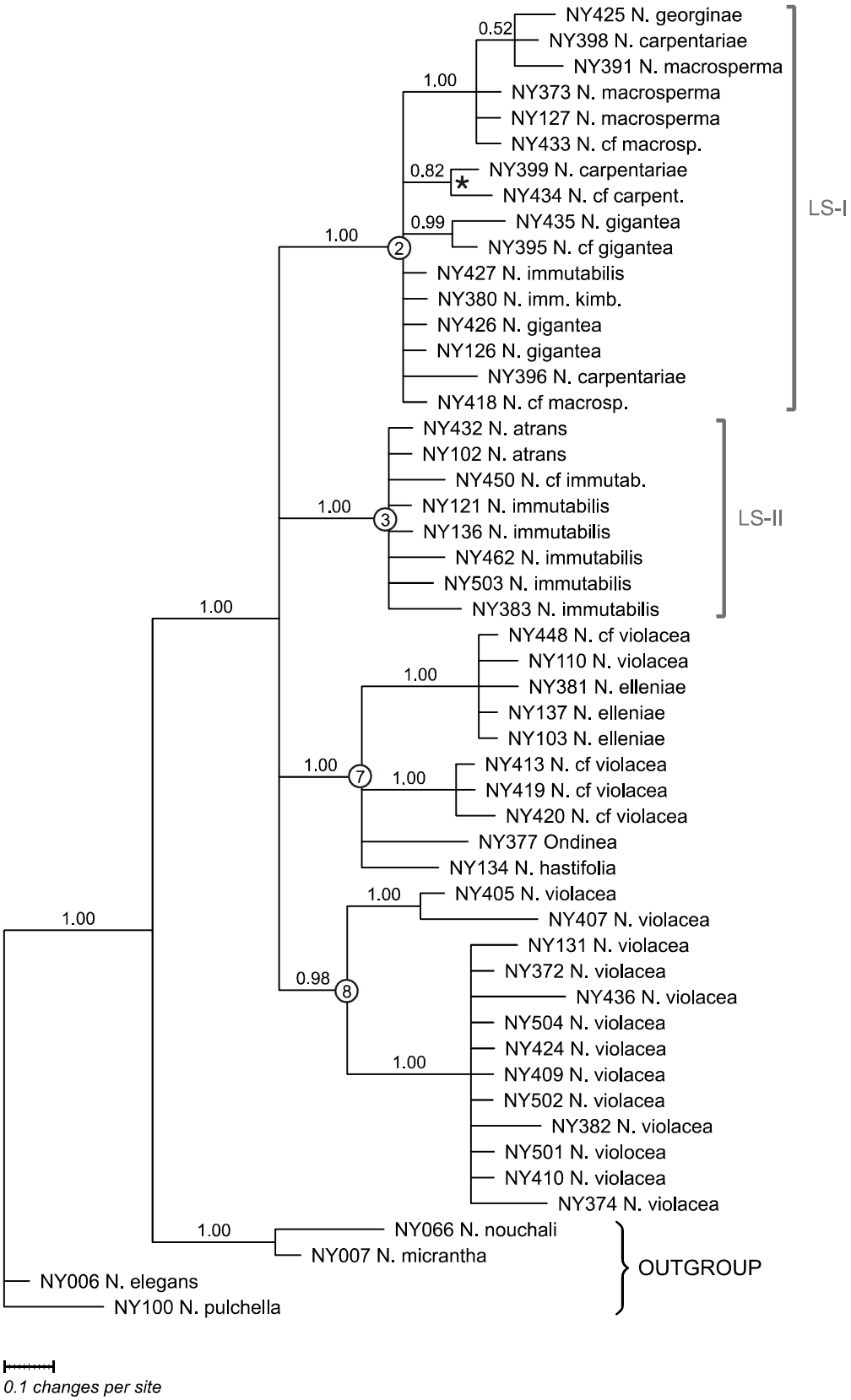


Figure 4.7 — Phylogram of phylogenetic relationships in *Nymphaea* subg. *Anecphya* inferred from the chloroplast marker *trnT–trnF*. The tree is a 50% majority rule consensus of 39,204 trees obtained from four runs of Bayesian analysis of the 51–taxon data set (substitutions + indels). Nodes relevant for discussion are numbered (in circles, see also Table 4.6).Numbers above branches indicate posterior probabilities of the respective nodes. The asterisk marks the node that is not present in the MP tree (Figure 4.6). Besides this node both trees are identical.

Table 4.6 — Results of phylogenetic analyses of the *trnT-trnF* data set using Maximum Parsimony (MP) and Bayesian (BI) approaches. Both, MP and BI have been conducted for the 51–taxon data set, as well as for data sets reduced to 44 and to 36 taxa, respectively. For the 44–taxon data set the following samples were excluded: NY382, NY409, NY424, NY501, NY502 (all *N. violacea*), NY427 (*N. immutabilis*) as well as NY380 (*N. immutabilis* ssp. *kimberleyensis*). For the 36–taxon data set the following taxa were excluded additionally: NY110 (*N. violacea*), NY418 (*N.* cf. *macrosperma*), *N. georginae* (NY425), and all samples of *N. gigantea* (NY126, NY395, NY426, and NY435). In contrast to ITS, the inclusion of putative hybrid taxa has no effect on topology and node support. See text for explanations. For MP, analyses of only substitutions (first line) and substitutions + indels (second line) were run separately. For BI, only the complete data sets (substitutions + indels) were analysed.

		MP 51 taxa	**MP 44 taxa**	**MP 36 taxa**	**Bayesian 51 taxa**	**Bayesian 44 taxa**	**Bayesian 36 taxa**
a)	***Tree statistics***						
	Number of characters	1464 *1478*	1464 *1478*	1464 *1478*	*1478*	*1478*	*1478*
	Number of trees	23 *45*	22 *44*	15 *34*	39204	39404	38998
	tree length	68 *84*	66 *82*	63 *78*	–	–	–
	CI	0.971 *0.952*	0.970 *0.951*	0.968 *0.949*	–	–	–
	RC	0.960 *0.937*	0.957 *0.933*	0.953 *0.926*	–	–	–
b)	***Bootstrap values, Bremer support (in brackets), and posterior probabilities (right column)*** ***Note:*** *nodes 1, 5 and 6 (as inferred from ITS) are not present in any of the trnT–trnTF trees*						
	Node 2 (LS-I group, but including NY380, NY427)	87 (2) *94 (3)*	87 (3) *94 (3)*	87 (3) *93 (3)*	1.0	1.0	1.0
	Node 3 (LS-II group)	95 (3) *95 (3)*	95 (2) *95 (3)*	95 (2) *96 (3)*	1.0	1.0	1.0
	Node 4 (small–seeded Anecphya clade)	63 (1) *n.p.*	63 (1) *n.p.*	63 (1) *n.p.*	n.p.	0.62	0.62
	Node 7 (only present in trnT–trnF)	63 (1) *79 (1)*	64 (1) *77 (1)*	63 (1) *74 (1)*	1.0	1.0	1.0
	Node 8 (only present in trnT–trnF)	n.p. *75 (1)*	n.p. 74 (1)	n.p. *72 (1)*	0.98	0.93	0.93

4.4 Discussion

4.4.1 Sequence variability and phylogenetic utility of *trnT–trnF* and ITS in *Anecphya*

A comparison of sequence statistics summarized in Table 4 clearly reveals that ITS is much more variable than *trnT–trnF*. Despite the fact that the ITS matrix comprises not even half the number of characters as the *trnT–trnF* matrix, ITS provides more than three times as much informative characters (158 in ITS vs. 42 in *trnT–trnF*, Table 4.4). This difference in variability and information content is also reflected in the low mean sequence divergence in the *trnT–trnF* data set (0.6% compared to 5.5% in ITS, Table 4.4). Sequence divergence is higher in the *trnT–trnL* spacer than in the *trnL* intron or the *trnL–trnF* spacer, in line with the general patterns in this region (Borsch *et al*. 2003, Shaw *et al*. 2005). Similar to other Nymphaeales there is an extended terminal part of the P8 stem–loop region of the *trnL* intron as was shown in secondary structure analyses (Borsch *et al*. 2003, Quandt *et al*. 2004). Long AT–rich sequence parts are thought to have evolved through a stepwise mutation process leading to an independent growth of this region in different groups of land

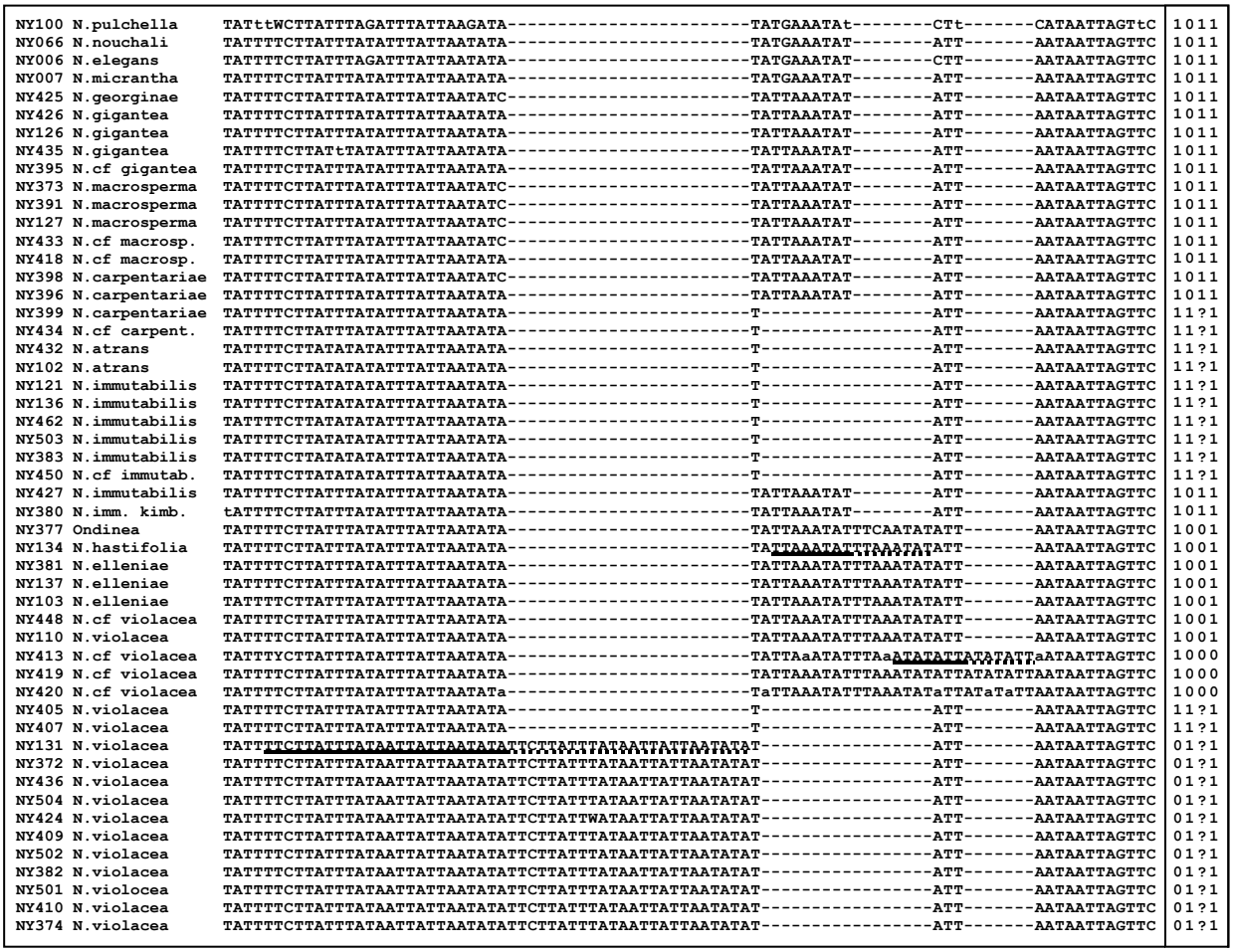


Figure 4.8 — Section of the *trnT–trnF* matrix, showing a highly variable part of the P8 stem–loop in the *trnL* intron. This section yields four of the seven parsimony informative indel characters of the whole *trnT–trnF* data set. Three of these indels are simple sequence repeats (template and repeat indicated by full and dotted lines, respectively), the other might be a deletion. In the right column of the figure, the binary codes for these length mutations are illustrated (1 = gap present, 0 = no gap).

plant (Quandt et al. 2004). In *Nymphaea*, Borsch *et al.* (subm.) recently have shown that subg. *Brachyceras* and *Anecphya* and the genus *Ondinea* have highly similar P8 stem–loop regions. Therefore, the *trnL* sequences could be aligned completely throughout this dataset, allowing to make use of the numerous mutations in this AT–rich part. This part entails length mutations of up to 24 nucleotides, even among different samples of *N. violacea* or *N. immutabilis* (Figure 4.8). Polytomies in the *trnT–trnF* tree largely seem to be due to not enough information rather than being the result of homoplasy, as evident from high CI and RC values (Table 4.5 and Table 4.6) in the MP trees, and from the fact that the 50% majority rule consensus is not better resolved than the strict consensus (tree not shown).

In the ITS region, 25% of all characters are variable and 19% are parsimony informative. When comparing the different partitions in ITS it becomes obvious that ITS1 (8% mean sequence divergence, 27% informative characters) is more variable than ITS2 (5.9% mean sequence divergence, 21% informative characters), which is in line with what Woods et al. (2005) found in hardy water lilies. Moreover, it is congruent to the general patterns of variability observed in the ITS region in angiosperms (Baldwin *et al.* 1995, Fuertes Aguilar & Nieto Feliner 2003). In parallel to the distribution of variable and informative characters, the frequency of polymorphic nucleotides sites in ITS1 is somewhat higher than in ITS2 (20 vs. 17 polymorphic nucleotides at parsimony informative sites). Only a single polymorphism was observed in the 5.8S rDNA, occurring in the *N. immutabilis* ssp. *kimberleyensis* sample (NY380). Polymorphic sites are extremely valuable for identifying putative hybrids in a given data set, especially if the two bases involved in a polymorphic site were also found separately in other accessions of the data set (Sang *et al.* 1995, Fuertes Aguilar *et al.* 1999, Andreasen & Baldwin 2003, Fuertes Aguilar & Nieto Feliner 2003).

4.4.2 Polymorphic sites and incongruent trees provide evidence for reticulate evolution

In contrast to the chloroplast genome the nuclear genome is inherited biparentally. Thus, a successful fertilization is followed by recombination of the parental alleles. In the case sequences of a given locus are divergent among parents, polymorphisms will be observed if the hybridisation event happened rather recently. In ITS, these differences among paralogues are levelled out over time through concerted evolution (Fuertes Aguilar *et al.* 1999). Concerted evolution means that all repeat copies of the nuclear ribosomal DNA, evolve in unison due to unequal crossing–over during meiosis and gene conversion (Arnheim 1983, Hillis *et al.* 1991, Elder & Turner 1995). However, there is recent evidence that the tempo and degree of concerted evolution strongly depends on factors specific to the organisms, such as generation time (annual vs. perennial) and mode of reproduction (sexual vs. vegetative, e.g., Suh *et al.* 1993, Baldwin *et al.* 1995, Buckler *et al.* 1997). In some cases divergent repeat types may persist, especially if hybridisation is coupled to an allopolyploidisation event or if the hybrids reproduce only vegetatively (e.g. Campbell *et al.* 1997, Zhang & Sang 1999). As an alternative explanation for the observed polymorphisms in ITS one might think of random mutations in some of the ITS copies of the same genome that have not been levelled out yet (Baldwin *et al.* 1995, Buckler *et al.* 1997). In the case of the *Anecphya* dataset such an explanation seems unlikely because most of the nucleotide polymorphisms and all of the length polymorphisms are parsimony informative sites, often occurring in several individuals, rather than being distributed at random. Both of the

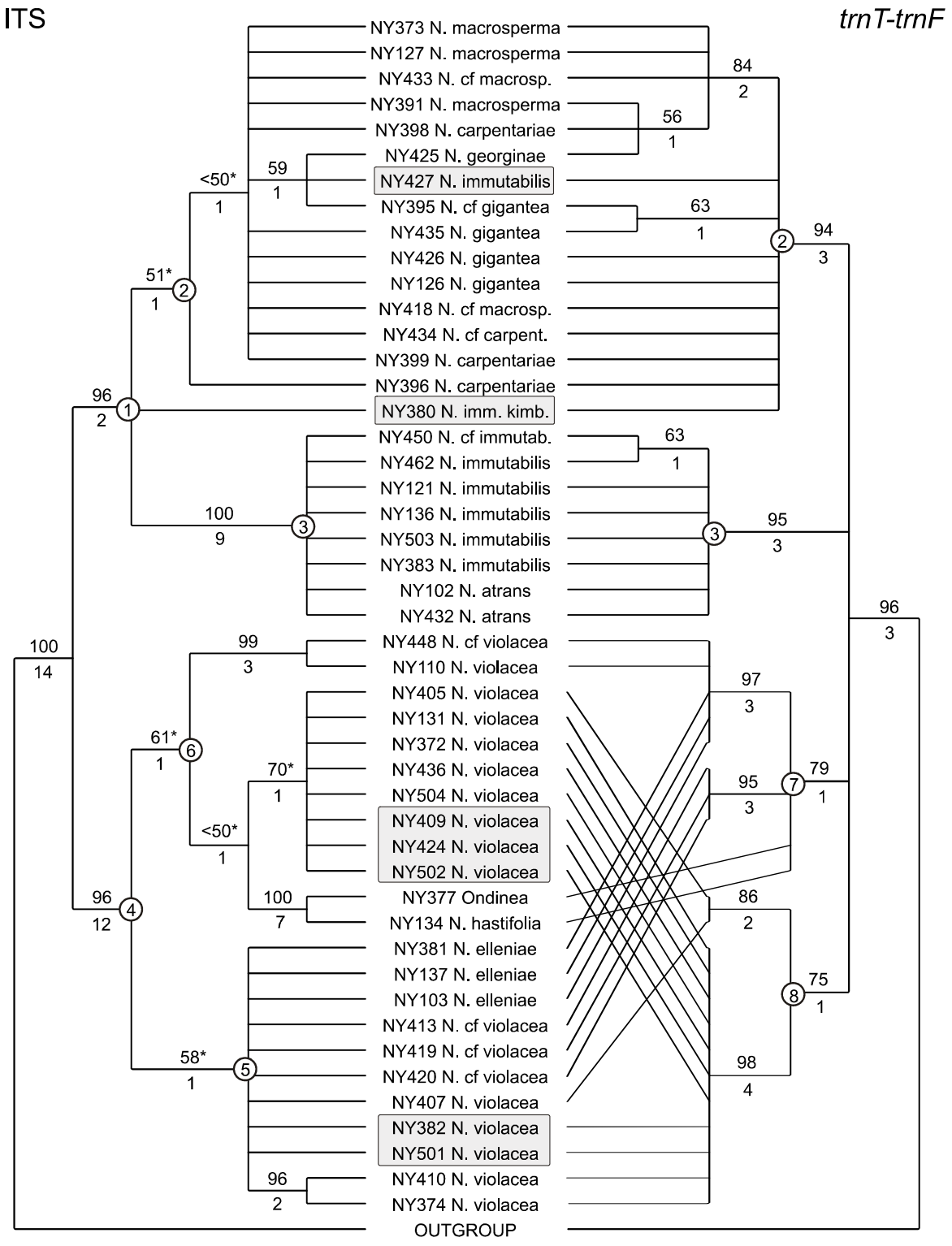


Figure 4.9 — Incongruence of the ITS tree (left) and the *trnT–trnF* tree (right) for *Nymphaea* subg. *Anecphya*. The figure shows the respective strict consensus trees of MP analyses of the 51–taxon data sets. Jackknife support for nodes is given above the branches, Bremer support below. Note that in ITS support values marked with (*) are significantly higher if potential hybrid sequences (taxa highlighted by grey boxes) are excluded (see Figure 4.4 for the 44-taxa ITS tree). See text for discussion. Node numbers (in circles) are congruent with numbers in other figures and in Table 4.5 and Table **4.6**.

character states occurring at a given polymorphic site often also occur alone in one or more other individuals sequenced and thus suggest hypotheses on potential parents. In this *Anecphya* ITS dataset, 14 samples (Table 4.3) show between two and twelve polymorphisms at parsimony informative sites, all of which are consistent in suggesting the same parentage.

The process of concerted evolution results in either intermediate sequences, in which characters from both parents have been fixed in all paralogues, or in only one parental sequence dominating the rDNA repeats within a genome. Both versions can have strong, possibly misleading effects on phylogenetic inference if putative hybrids are included in the data set (Álvarez & Wendel 2003). Fuertes Aguilar & Nieto Feliner (2003) attribute the large polytomies in their trees of species relationships in the genus *Armeria* to disruptive effects of reticulation and concerted evolution. In the present data set the inclusion of putative hybrids also had adverse effects on the trees inferred from ITS. As obvious from Figure 4.9 (left) the resolution of the tree is reduced and, more remarkably, the Jackknife support is reduced significantly for the clades LS–I (node 2), SS–I (node 5) and SS–II (node 6). Thus, the exclusion of putative hybrid sequences from phylogeny reconstruction seems justified and is in line with the general approach in other studies (e.g., Sang *et al*. 1995, Fuertes Aguilar & Nieto Feliner 2003, Koontz *et al*. 2004, Marhold *et al*. 2004). If included, all putative hybrids – except NY380 – are depicted in one of the assumed parental clades, which points to ongoing concerted evolution towards one of the parental genomes in the respective taxa (Sang *et al*. 1995). On the other hand, this might also be an indication of introgression of another genome and subsequent backcrossings with one parent (Fuertes Aguilar & Nieto Feliner 2003). Evidence from other sources besides the sequences will be necessary to decide whether ancient hybridisation or introgression is the dominant underlying process. Such evidence could come from chromosome numbers and ploidy levels, geographical distribution, morphology, or unlinked gene sequences (Andreasen & Baldwin 2003, Vriesendorp & Bakker 2005).

Whereas the occurrence of polymorphic sites in ITS of some *Anecphya* individuals indicates rather recent hybridisation events, the incongruence of the trees obtained from nuclear and chloroplast markers points to more ancient hybridisation (and subsequent concerted evolution in the nuclear rDNA). Figure 4.9 shows that in the present data set incongruence is mainly restricted to inferred relationships among the small–seeded members of *Anecphya*. This group comprises the morphologically very variable *Nymphaea violacea*, but also the morphologically distinct and geographically restricted species *N. elleniae*, *N. hastifolia* and *Ondinea purpurea*. Nevertheless, based on molecular data in combination with additional information from biogeography and from morphological characteristics of the plants, some hypotheses on relationships, even within the small–seeded *Anecphya* clade, can be established. These will be outlined in the following paragraphs.

4.4.3 Phylogenetic relationships and reticulate evolution in *Nymphaea* subg. *Anecphya*

Phylogenetic analysis of the ITS data set clearly reveals a major subdivision of *Nymphaea* subg. *Anecphya* into a large–seeded (LS) and a small–seeded (SS) group of species. As mentioned earlier, the LS species share several morphological synapomorphies, such as rather large seeds, toothed leaf margins and a distinctive gap between petals and stamens, whereas the SS–group is characterized by relatively small seeds, entire leaf margins and petals grading into stamens (Jacobs 1992, 1994). Thus, molecular data are consistent with morphology regarding the two major clades in *Nymphaea* subg. *Anecphya*.

Large–seeded clade of subg. Anecphya

Both, the nuclear and the chloroplast data reveal a close affinity of *N. immutabilis* and *N. atrans*, which are resolved in a well–supported clade in all trees (node 3, see also Table 4.5 and Table 4.6). Hybrids between these two species with intermediate character states and reduced fertility have been observed in areas where both species occur sympatrically (Jacobs 1992). No indication for hybrids between *N. immutabilis* and *N. atrans* could be observed in the sampled sequences, because there were no parsimony informative sites that would distinguish between the two species.

However, one of the samples of *N. immutabilis* (NY427), from the east coast of Queensland (Figure 4.10), was designated as a putative hybrid between the LS–II (*N. immutabilis*, *N. atrans*) and the LS–I group (*N. gigantea*, *N. macrosperma*, *N. carpentariae*) based on frequent nucleotide polymorphisms. The maternal parent of NY427 seems to be a member of the LS–I group since the NY427 is clearly depicted in this clade in the chloroplast tree. Also the ITS sequences show some affinities to sequences of the LS–I clade, pointing to ongoing concerted evolution towards this paternal genome or to continuing gene flow through backcrossing with the paternal parent. Considering the distribution of species in both LS groups, NY427 presumably is a hybrid between *N. immutabilis* and *N. gigantea*. Jacobs (1992) reports frequent intergradation between the two species in the areas of overlapping distribution along the Queensland coast. In fact, the population from which the NY427 sample was taken differs from "normal" *N. immutabilis* in the shape of the leaf margin and the flower colour ("much bluer throughout", pers. communication B. Hellquist).

Another sample of the large–seeded members of *Anecphya* that stands out through the occurrence of both additive nucleotide polymorphisms and length polymorphisms is *N. immutabilis* ssp. *kimberleyensis* (NY380) from Western Australia (map in Figure 4.10). This subspecies is known only from a single population in a seasonally water–filled lagoon in the Kimberley region, and differs from *N. immutabilis* ssp. *immutabilis* in having considerably fewer stamens (up to 200 vs. up to 400), glabrous seeds, anthers with a purple gland at the base, and blue petals slightly fading with age (Jacobs 1992). The fact that only immature seeds could be found in the field hints on reduced fertility in this population. The sample NY380 is somehow exceptional in the present data set because it apparently is the outcome of a hybridisation event between large–seeded and small–seeded Anecphya clades. At the polymorphic sites in ITS character states from the the SS–I group and the LS–I group are present. Given the distribution of species in the SS–I clade (*N. violacea* p.p., *N. elleniae*), *N. violacea* must be considered as the paternal parent of this taxon. The well–supported

position of NY380 in the LS–I clade, as resolved from the chloroplast marker, is somehow puzzling since this would infer *N. macrosperma* to be the maternal parent (considering the distribution of species in the LS–I group) and not *N. immutabilis*, as expected from the phenology. However, since the present molecular data does not allow a final conclusion on the origin this taxon, it might also be that NY380 is a product of geneflow between all three species, *N. immutabilis*, *N. macrosperma* and *N. violacea*. A careful re–examination of morphological characters and a more detailed analysis using additional molecular markers would be necessary to unravel the history and affinities of *N. immutabilis* ssp. *kimberleyensis*.

Small–seeded clade of subg. Anecphya

Neither in trees obtained from ITS nor from *trnT-trnF*, the individuals sampled from *Nymphaea violacea* form a monophyletic group. Instead *N. violacea* samples appear in both subclades, SS–I and SS–II, together with either *N. elleniae* (SS–I) or *N. hastifolia* and *Ondinea purpurea* (SS–II). Apparently, at least two different nuclear genomes occur in *N. violacea*, and in five samples there is evidence for admixture of these different types (marked as possible violacea–hybrids with a pink triangle in Figure 4.4 and Figure 4.11). However, a general conclusion on the relationships within the small–seeded group is not possible,

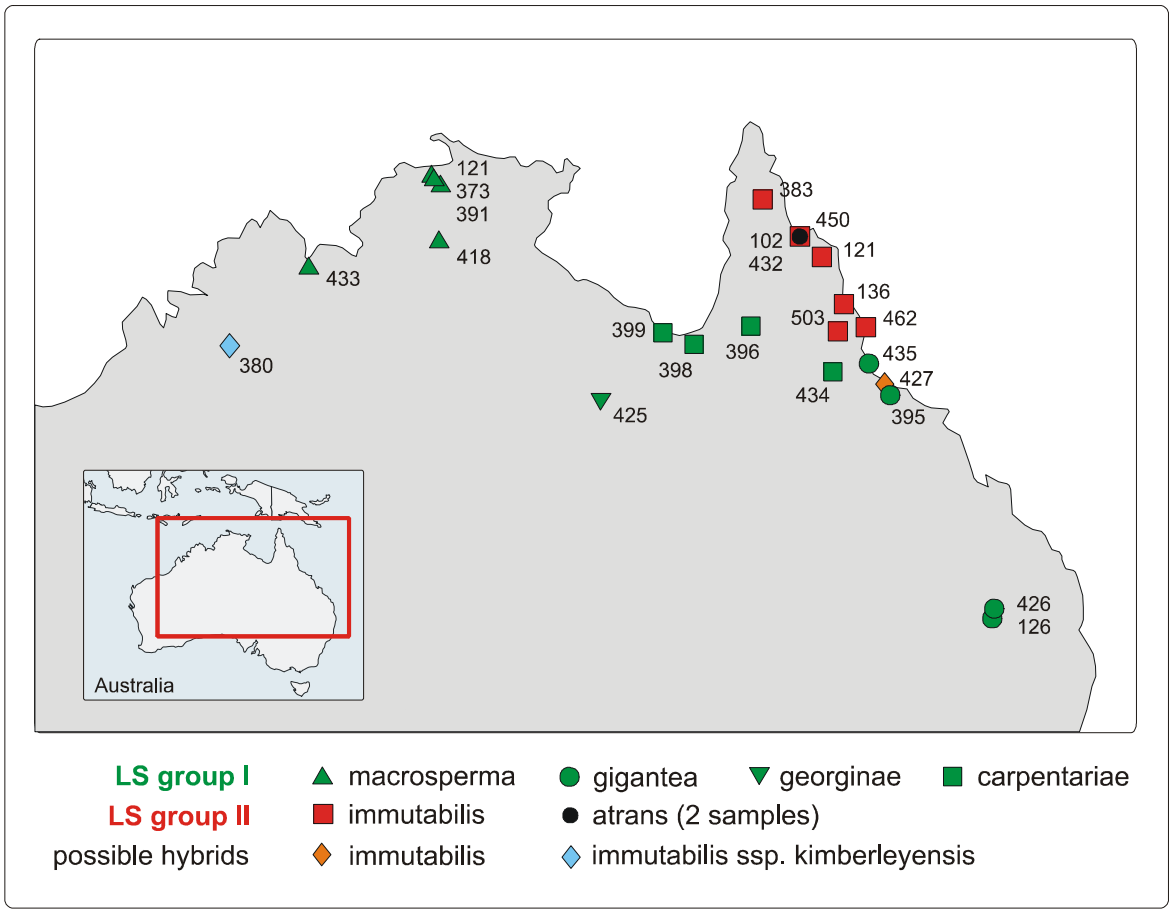


Figure 4.10 — Field localities of *Nymphaea* subg. *Anecphya* samples of the "Large–seeded group". Numbers in the figure are the DNA sample numbers given in Table 4.1. Shape and colour of symbols correspond to those used in the MP trees inferred from ITS and *trnT–trnF* (Figure 4.4 and Figure 4.6).

because of the complex pattern observed (see Figure 4.9). It appears impossible to decide of which subclade's (SS–I or SS–II) members are maternal or paternal ancestors of these hybrids. If the 5 putative hybrids are excluded from consideration, the discordant patterns revealed by the nuclear and the plastid marker might still be explained by lineage sorting (thereby eliminating the necessity for hybridisation or introgression as underlying processes). However, in view of the biogeographic data some interesting phenomena become obvious and will be discussed in the following.

The two *N. violacea* samples NY448 and NY110, both from nearby populations at the Cape York Peninsula (see map in Figure 4.11), apparently share the chloroplast genome with *N. elleniae*, but emerge distantly from *N. elleniae* in the ITS tree (*N. elleniae* in SS–I, NY448 and NY110 in SS–II). A close relationship of these two taxa to *N. elleniae*, possibly originated from gene flow in one or the other direction, appears plausible in view of the morphological distinctiveness and scent similar to *N. elleniae* apparent especially in NY448 (B. Hellquist, pers. communication). A similar situtation can be found in the two *N. violacea* samples NY405 and NY407. Although they possess different nuclear genomes (NY405 in SS–II, NY407 in SS–I), they share several apomorphies in the chloroplast DNA. The close proximity of the respective populations (Figure 4.11), suggests ongoing introgression or hybridisation.

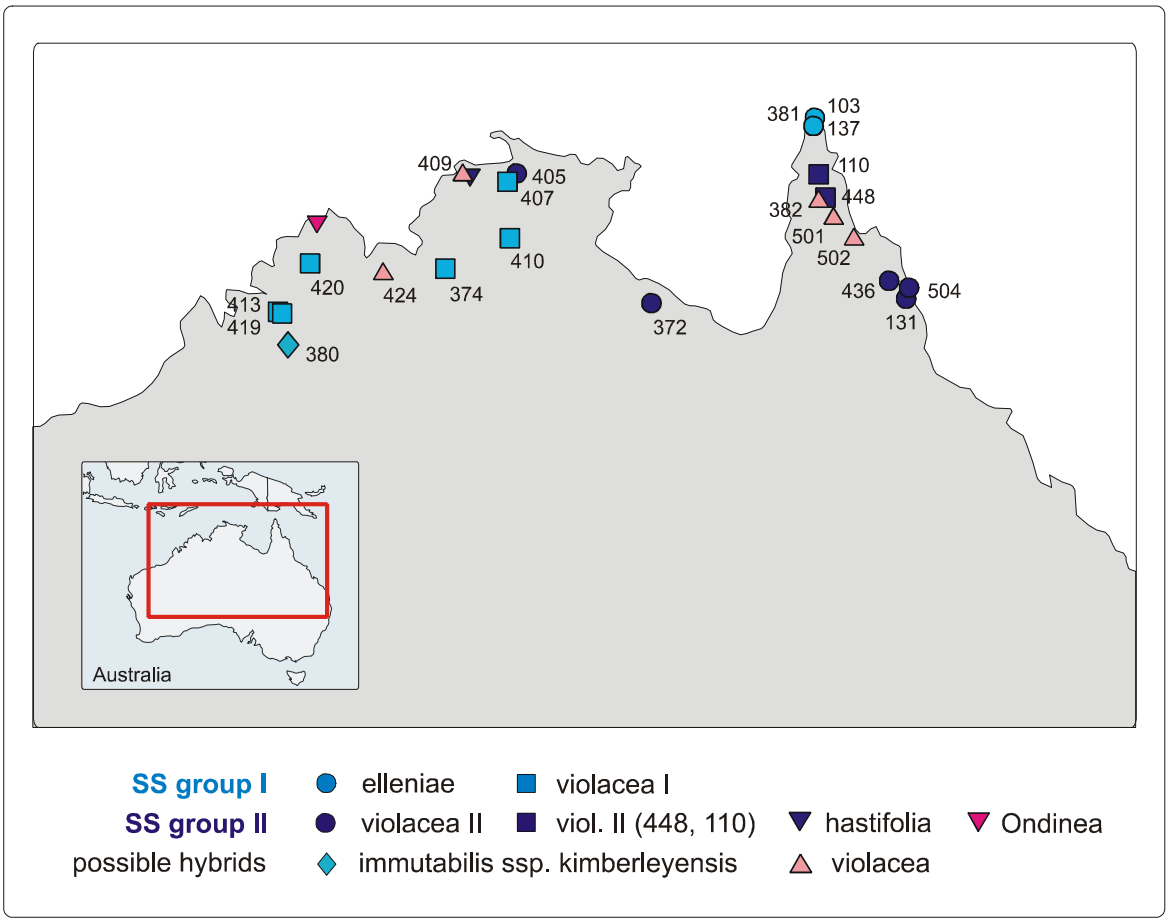


Figure 4.11 — Field localities of *Nymphaea* subg. *Anecphya* samples of the "Small–seeded group". Numbers in the figure are the DNA sample numbers given in Table 4.1. Shape and colour of symbols correspond to those used in the MP trees (Figure 4.4 and Figure 4.6).

Those *N. violacea* samples, in which polymorphic sites in ITS could be observed (NY382, NY409, NY424, NY501, NY502) all have the same chloroplast haplotype and appear in the *trnT–trnF* tree in a clade together with six other *N. violacea* samples (4 from SS–II, 2 from SS-I). Given their scattered geographic distribution and given the noticable phenological differences between them, they are probably the result of independent events of hybridisation or introgression. However, despite the observed polymorphism in the nuclear rDNA, the term "hybrid" might not be appropiate for these taxa because the possible parents belong different lineages of the same species and "hybridisation" is commonly used to describe interbreeding of members of different species (Vriesendorp & Bakker 2005). Another interesting group of samples of *N. violacea* are the three samples from the Kimberley region (NY413, NY419, and NY420). Morphologically they are quite distinct from the rest of *N. violacea*, as evident from the presence of large stipular sheaths, much larger seeds and striking white flowers with blue– or magenta–tipped petals (B. Hellquist, unpublished data). These samples are depicted within the SS–I group from ITS, but are resolved as a monophylum in the *trnT–trnF* tree, supported even by a synapomorphic indel (see Figure 4.8, last indel).

4.4.4 The position of *Ondinea purpurea*

Previous analyses of Nymphaeales using a large set of cpDNA markers (see Chapter 3) and a detailed examination of relationships within the genus *Nymphaea* based on a dense taxon sampling and *trnT–trnF* sequences (Borsch *et al.* subm.) revealed a close affinity of the *Ondinea* plastid sequences to those of the Australian water lilies *Nymphaea* subg. *Anecphya*, and within this subgenus to the samples of the small–seeded group. Based on these well–supported results a possible ancient hybrid origin of *Ondinea*, with a member of the small–seeded *Anecphya* clade being the maternal parent, was hypothesized (see Chapter 3.4.2.). The present study confirms the previous findings for the maternally inherited plastid marker and adds further evidence in support of this hypothesis from nuclear DNA. The ITS tree (Figure 4.4 and Figure 4.5) provides high support for the close affinity of *Ondinea* and the small–seeded *Anecphya* clade. In fact, the sequence of *Ondinea* shares several synapomorphies with that of *N. hastifolia* (NY134), which is a member of the SS–II clade. A close affinity between these two taxa and another set of samples from the small–seeded group is also suggested from the *trnT–trnF* data set (Figure 4.6 and Figure 4.7).

The inferred close relationship of *Ondinea* and the small–seeded species of *Anecphya*, especially *Nymphaea hastifolia*, raises a lot of new, interesting questions on the evolutionary history of these taxa. A close relationship of *Ondinea* to the genus *Nymphaea* has been proposed long ago based on several shared characters from morphology, anatomy and ecology, e.g. floral anatomy and venation patterns (Williamson & Moseley 1989, Schneider *et al.* 1995), floral organs grading from petals into stamens (Den Hartog 1970, Kenneally & Schneider 1983), the morphology of the gynoecial cup and stigmatic papillae (Schneider 1983, Schneider & Williamson 1993), pollen morphology (Müller 1970), fruit and seed anatomy (Schneider & Ford 1978), as well as similar pollination and seed dispersal syndromes (Schneider 1983, Schneider *et al.* 1984). However, so far no one has assumed that *Ondinea* might be more closely related to a specific taxon within the genus *Nymphaea*.

Instead, close relationships of *Ondinea* and *Barclaya* have been proposed alternatively (Kenneally & Schneider 1983, Schneider 1983, Jacobs & Porter in prep.).

In view of the geographic distribution, a close affinity between *Ondinea* and *N. hastifolia*, however, seems reasonable. Both species occur sympatrically in the Kimberley region of Western Australia (Figure 4.12), and both are characteristic for ephemeral habitats. *N. hastifolia* occurs in temporally water–filled pools on floodplains, whereas *Ondinea* is typical for streaming water in ephemeral creek systems. The presence of glabrous seeds and the grading transition from petals (if present) into stamens observable in *Ondinea* might add further evidence for a close affinity to the small–seeded species of *Anecphya*. To further elucidate the origin of *Ondinea*, in addition to more molecular data, an assessment of ploidy levels would be helpful to reveal possible scenarios, including allopolyploidisation and hybridisation.

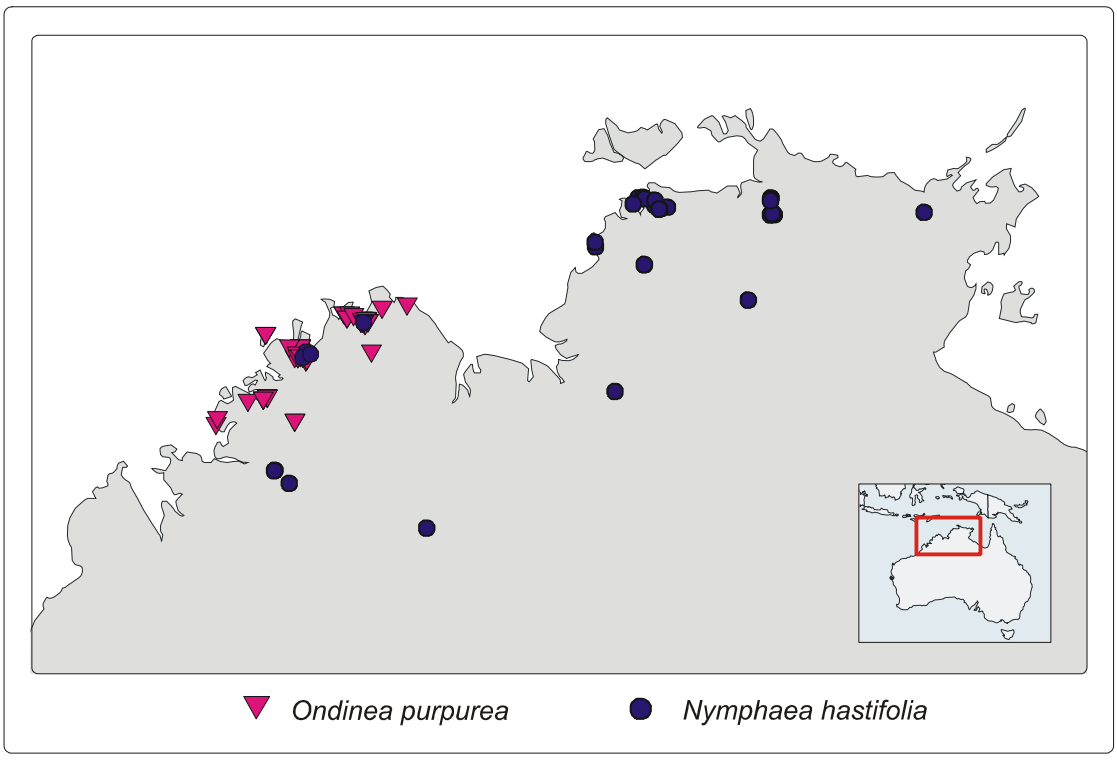


Figure 4.12 — Distribution of *Nymphaea hastifolia* Domin and *Ondinea purpurea* Hartog in Northern Australia. Circles and triangles indicate field localities of specimens deposited in the herbaria of Australia (New South Wales, Northern Territory, Western Australia, Tasmania, Canberra, Victoria, Queensland, and South Australia). Data was taken from the online resource of Australia's Virtual Herbarium (http://www.chah.gov.au/avh/), providing 37 records for *Nymphaea hastifolia* and 64 records for *Ondinea purpurea* (date of data accession: December, 2004).

4.5 Conclusions and suggestions for further studies

The present phylogenetic analysis of the Australian water lilies in *Nymphaea* subg. *Anecphya* is a first comprehensive study, including a broad taxon sampling of this subgenus and evidence from the nuclear and the chloroplast genome. Molecular data confirm the subdivision of subg. *Anecphya* into two major clades, which corresponds to groupings based on several morphological characters such as seed size (Jacobs & Porter in prep.). Among the large–seeded group, there is one subclade comprising *N. gigantea*, *N. macrosperma*, and *N. carpentariae* (LS–I) and another subclade containing *N. atrans* and *N. immutabilis* (LS–II). Relationships within the small–seeded group are less clear, since the trees obtained from the chloroplast marker *trnT–trnF* and from the nuclear marker ITS are strongly incongruent. The observed pattern of polymorphic sites and homogenous ITS copies further hints past and current reticulate evolution.

Clear hypotheses could be established with the present data, which need to be tested in future studies. Especially the results from the ITS data set require confirmation by other nuclear sequence data, mainly because of the special modes of molecular evolution of the nuclear rDNA. Concerted evolution levels differences between different allels over time and, therefore, evidence for further ancient hybridisation or introgression might be obscured. Additional information from low–copy or single–copy nuclear genes will be essential to obtain a reliable nuclear consensus topology (Álvarez & Wendel 2003, Bailey *et al.* 2003, Small *et al.* 2004) and to evaluate how representative the rDNA is for the nuclear genome. Additional variable plastid sequences are likely to provide better resolution in the subclades, especially among the specimens that group in the LS–I and LS–II clades.

Since reticulation is difficult to demonstrate conclusively, evidence from other sources should be used additionally, e.g. dominant markers such as ISSR, RAPD, or RFLP (Fuertes Aguilar & Nieto Feliner 2003), but also evidence from geographical distribution, chromosome numbers, mode of reproduction, or morphology (Andreasen & Baldwin 2003). The assessment of chromosome numbers and ploidy levels in the respective populations would be essential to decide whether putative hybridisation events were accompanied by alloploidisation. So far, no estimations of ploidy levels in species of *Nymphaea* subg. *Anecphya* have been made, but since polyploidy is common in other subgenera of Nymphaea (Langlet & Söderberg 1927, Gupta 1978, 1980, Okada & Tamura 1981, Wiersema 1987), it might be expected for this subgenus, too. Further studies involving different molecular markers would also be necessary to dispel the remaining doubts in the inferred close relationship of *Ondinea purpurea* and the small–seeded clade in *Nymphaea* subg. *Anecphya*. However, addititional investigations on the reproductive biology of Ondinea and presumed relatives from the small–seeded Anecphya–clade, including crossing experiments between *N. hastifolia* and *Ondinea*, might provide further valuable insights into the complex evolutionary pattern in Australian water lilies.

Chapter 5

Biogeography of Nymphaeales: Extant patterns and historical events

Contents

5.1 Introduction

The Nymphaeales are a monophyletic group comprising two currently recognized families: the Cabombaceae (consisting of *Cabomba* and *Brasenia*) and the Nymphaeaceae (consisting of *Nuphar*, *Barclaya*, *Victoria*, *Euryale*, *Nymphaea* and *Ondinea*). Recent molecular studies on angiosperm phylogeny inferred the Nymphaeales to be part of the basal grade of angiosperms, together with the monotypic *Amborella* and Austrobaileyales (Chapter 1.2.4). Alternative hypotheses for the branching order within this basal grade place *Amborella* either as the first branch followed by Nymphaeales and Austrobaileyales as successive sister lineages (e.g., Parkinson *et al.* 1999, Mathews & Donoghue 2000, Soltis *et al.* 2000, Borsch *et al.* 2003, Hilu *et al.* 2003) or depict a clade consisting of *Amborella* and Nymphaeales as sister to Austrobaileyales and all other angiosperms (e.g., Barkman *et al.* 2000, Zanis *et al.* 2002). Thus, Nymphaeales play an important role in understanding the early evolution of flowering plants.

With respect to the diversity in the order and the distribution of its species and genera, Nymphaeales may be regarded as the first globally diverse clade among extant angiosperms. Members of this basal angiosperm order occur in freshwater habitats all over the world – except Antarctica – with *Brasenia* and *Nymphaea* being the most widespread genera. Other genera in Nymphaeales have more restricted ranges with *Ondinea* representing the other extreme as a local endemic in the Kimberley region of NW Australia. The fact, that closely related taxa in Nymphaeales can occupy quite distant ranges, e.g. *Victoria* in South America and *Euryale* in East Asia, leads to questions about the evolutionary and biogeographic history of this lineage. Do extant patterns reflect ancient vicariance or are they the result of long–distance dispersal?

Traditionally, such disjunct distributional patterns in evolutionary rather old lineages were thought to reflect relicts of an ancient Gondwanan distribution (see McGlone 2005 for review). However, recent analyses of historical biogeographic relationships in different plant lineages have provided some new hypotheses. Examples from several widespread angiosperm families, such as Magnoliaceae (Azuma *et al.* 2001), Annonaceae (Richardson *et al.* 2004), Lauraceae (Chanderbali *et al.* 2001) and Malpighiaceae (Davis *et al.* 2002a, Davis *et al.* 2004), indicate post–Gondwanan diversification of the crown groups and migrations via Laurasia during warmer periods in the Tertiary. Additionally, there is increasing evidence for tropical disjunctions originating from long–distance dispersal across oceanic barriers (Givnish & Renner 2004). Muñoz *et al.* (2004) could show the importance of long–distance dispersal through wind in the southern hemisphere, and Renner (2004c) provided strong arguments for sea currents as agents for trans–Atlantic dispersals. The growing realization of the importance of long–distance dispersal in historical biogeography has mainly been triggered by accumulating molecular data and recent progress in dating phylogenies (De Queiros 2005). In many cases, those studies revealed relatively young crown group ages even of very basal angiosperms lineages (e.g., Zhang & Renner 2003).

With the development of cladistic methods, several approaches and techniques for analysing historical biogeography have been developed (e.g., Bremer 1992, Ronquist 1994, Bremer 1995, Ronquist 1997, Hausdorf 1998, see Morrone 2005 for overview), all acknowledging that individual geographic histories of lineages might be independent from general area–

cladograms. Dispersal–vicariance analysis (Ronquist 1996, 1997), a method that optimizes ancestral areas on a cladogram by minimizing the number of historical events, has become the most commonly used approach in historical biogeography (e.g., Sanmartín *et al*. 2001, Xiang & Soltis 2001, Davis *et al*. 2002b, Austin *et al*. 2003, Sanmartín & Ronquist 2004, Near & Keck 2005). However, Cook & Crisp (2005) argued that tree–based biogeographic inference can lead to wrong conclusions if the likelihood of particular vicariance or dispersal events at a given point of time — or asymmetry in the direction of dispersals — is ignored completely. Several Bayesian (Huelsenbeck *et al*. 2001, Pagel *et al*. 2004, Ronquist 2004) and maximum likelihood (Mooers & Schluter 1999, Nepokroeff *et al*. 2003) approaches have been applied in historical biogeography. However, all algorithms developed so far are not *per se* sufficient to reconstruct ancestral areas convincingly, but a careful interpretation of data and results considering the geological and biological background is always essential (Cook & Crisp 2005).

A new likelihood approach introduced by Ree *et al*. (2005) appears to be a promising tool in historical biogeography, since branch lengths, diversification times and the emergence and demise of barriers over a period of time can be regarded. However, the technical implementation of this method needs further refinement in order to allow computation of larger data sets with higher numbers of areas and uncertainties in divergence times or phylogenetic relationships (Ree *et al*. 2005). Thus, dispersal–vicariance analysis still seems to be the best applicable approach to date, but a careful consideration of the geographic history of the respective areas is necessary for a meaningful interpretation of the results. For inferring the historical biogeography of a specific lineage it is therefore essential to obtain robust estimates of phylogenetic relationships and the ages of relevant clades, as well as to consider the geological time sequence for formation of dispersal routes and barriers (Sytsma *et al*. 2004). For this purpose, the following two paragraphs provide a summary of the data from phylogenetic studies (Chapter 3 and Borsch *et al*. subm.) and from the fossil record of Nymphaeales. Furthermore, a short outline of the geological and climatic conditions during the period of earth history relevant for the evolution of Nymphaeales will be presented.

5.1.1 Phylogeny and evolution of Nymphaeales

The first analysis including DNA sequence information (Les *et al*. 1999) was based on a comprehensive data set of morphological and molecular characters, but taxon sampling was rather low. By representing each genus with one species and using *Cabomba* and *Brasenia* as outgroup taxa, Les *et al*. (1999) inferred the following branching order for Nymphaeaceae: (*Nuphar*, (*Barclaya*, (*Ondinea*, (*Nymphaea*, (*Victoria*, *Euryale*))))). However, the ideas about phylogenetic relationships in Nymphaeaceae changed significantly due to a recent study with an extensive taxon sampling reflecting the species diversity in the genus *Nymphaea* (Borsch *et al*. subm., 65 taxa, trnT–trnF). This analysis revealed three clades within the genus *Nymphaea*: one clade comprising subg. *Hydrocallis* and subg. *Lotos* as sister–groups, a second clade comprising subg. *Brachyceras* and subg. *Anecphya*, and, third, the isolated subg. *Nymphaea*, which appeared sister to the other two clades. Borsch et al. (subm.) showed that *Ondinea* is nested within a well–supported *Nymphaea* subg. *Anecphya* clade. The *trnT–trnF* data in addition indicated that *Nymphaea* subg. *Brachyceras* might be paraphyletic with respect to subg. *Anecphya*. Furthermore, the African water lily *Nymphaea*

petersiana, currently classified under subg. *Brachyceras*, was depicted as closely related to *N. lotus* and *N. pubescens*, which both are members of subg. *Lotos*.

Based on the results of the *trnT–trnF* study, Löhne *et al.* (subm., Chapter 3) designed a follow–up study that aimed at clarifying those relationships in Nymphaeales, which were only weakly supported by *trnT–trnF*, i.e. the monophyly of Nymphaeaceae and the monophyly of the genus *Nymphaea*. For this purpose, 29 representative taxa were selected from the data set of Borsch *et al.* (subm.) and sequenced for several additional markers. The results of Löhne *et al.* (subm.) corroborated the monophyly of Nymphaeaceae (JK = 99, PP = 0.96), with *Nuphar* branching first, followed by *Barclaya* and core Nymphaeaceae (*Nymphaea*, *Ondinea*, *Victoria*, *Euryale*). In congruence to the findings of Borsch *et al.* (subm.), Löhne *et al.* (subm.) revealed three major clades within the genus *Nymphaea* (*Hydrocallis–Lotos*, *Anecphya–Brachyceras*, and subgenus *Nymphaea*) and confirmed the position of *Ondinea* within *N.* subg. *Anecphya*. However, in contrast to Borsch *et al.* (subm.), the three *Nymphaea*–lineages did not constitute a monophyletic group. Instead, a clade comprising *Victoria* and *Euryale* was depicted sister to the *Hydrocallis–Lotos* clade, but with medium support (JK = 88, PP = 0.52). Furthermore, *N.* subg. *Brachyceras* emerges as a monophylum in the tree of Löhne *et al.* (subm.), albeit with medium support (JK = 78, PP = 0.98). Further studies will be necessary to discriminate between the alternative hypotheses obtained from the data sets of Borsch *et al.* (subm.) and Löhne *et al.* (subm.). For the present study on historical biogeography, both hypotheses will be regarded and their effects on the inference of ancestral areas will be tested.

A first attempt on dating the phylogeny of Nymphaeales has recently been made by Yoo *et al.* (2005). Using the proposed age of the angiosperms (131.8 Ma, based on fossil pollen, see Chapter 1.3.1) as a calibration point, they inferred a rather recent diversification of the Nymphaeales crown group, starting in the Eocene with the split between Cabombaceae and Nymphaeaceae around 44.6 Ma (± 7.9 Ma). However, the analysis of Yoo *et al.* (2005) was based on the data set and the phylogenetic hypothesis of Les *et al.* (1999). In view of the refined hypotheses on the complex phylogenetic history of Nymphaeales (Borsch *et al.* subm., Löhne *et al.* subm., Chapter 3) a new dating analysis that also involves a more extensive and judicious taxon sampling seems necessary.

5.1.2 The fossil record of Nymphaeales

A comprehensive synopsis of the fossil record of Nymphaeales has been compiled by Borsch (2000). Additional information on the fossil record of *Nuphar* is given by Chen et al. (2004). Most of the fossil remains that have been assigned to Nymphaeales are seeds or pollen. Morphologically, Nymphaeaceae and Cabombaceae seeds and pollen can easily be recognized (Collinson 1980 for review of seeds, Borsch 2000, Sampson 2000, and Hesse & Zetter 2005 for pollen). There are also numerous fossil leaves that have been assigned to Nymphaeaceae because of their subpeltate to peltate shape and an actinodromous venation. However, such leaves are known from several, not closely related lineages of flowering plants, e.g. *Nelumbo*, and might simply reflect an optimal structure for floating on the surface of water. Thus, the assignment of a fossil to Nymphaeales based only on leave

characters may be disputable, and will have to be based on a fine analysis of venation patterns and other anatomical details.

According to Borsch (2000), the fossil record of Cabombaceae is predominantly consisting of seeds of *Brasenia*, mainly from Oligocene through Pleistocene deposits of Europe and Russia, with the oldest fossil from the Upper Eocene of England (*B. spinosa* Chandler, Collinson 1980) and the youngest from the last interglacial (Eemian) of Middle and North Eastern Europe (e.g., Velichkevich 1994, Bosch *et al.* 2000). The nowadays monotypic genus *Brasenia* is likely to have been more diverse in the past as evident from a variety of described fossil species (20–30, see e.g., Dorofeev 1963, Collinson 1980, Borsch 2000). Fossils of *Cabomba* have not been described yet. The oldest fossils that can be unambiguously assigned to Nymphaeaceae may be the dispersed pollen grains from the Middle Eocene of Stolzenbach / Germany (Hesse & Zetter 2005). The oldest record for the genus *Nymphaea* is represented by seeds from the Upper Eocene or Lower Oligocene of England (*N. liminis* Collinson, Collinson 1980). Several fossil species of *Nymphaea* have been described from the Eocene through Pliocene of Europe and Russia based on preserved seeds, leaves and rhizomes (see Borsch 2000 for overview). Remarkably, there are Upper Miocene seeds from Europe that could be assigned to subgenera of *Nymphaea* (*N.* subgg. *Nymphaea* and *Lotos*) based on their characteristic seeds (Mai 1995). Recently reported fossil leaves of *Nymphaea* from the Upper Miocene of Argentina (Anzotegui 2004) reflect the first appearance of water lilies outside the holarctic region, although this documentation should be taken with caution because of the above–mentioned uncertainties in assigning fossil leaves to Nymphaeaceae. According to Chen *et al.* (2004) the oldest fossils of *Nuphar* are from the Middle Eocene of North America (Wehr & Manchester 1996) and from the Oligocene of Western Siberia (Dorofeev 1963, 1974). However, Chen *et al.* (2004) recently described fossil seeds of *Nuphar* from the Early Eocene of China (*N. wutuensis* Chen, Manchester & Chen) and mentioned a very similar specimen from the Late Palaeocene of North Dakota (*N. cf. wutuensis*). With an approximated age of 52 Ma (or 56 Ma in the case of the North Dakota fossil) these *Nuphar* seeds resemble the oldest fossils that are assignable to any of extant genera of Nymphaeales.

The existence of several morphologically deviating organ remains from that have been described as genera of Nymphaeales, such as *Allenbya* Cevalloz–Ferriz et Stockey, *Sabrenia* Collinson, *Dusembaya* Dorofeev, *Irtyshenia* Dorofeev, *Tavdenia* Dorofeev, *Barclayopis* Knobloch & Mai, or *Anoectomeria* Saporta, hints on the presence of several nowadays extinct lineages of Nymphaeales during the Tertiary. In most cases, these fossils can not be clearly assigned to any extant genera. However, Cevalloz–Ferriz & Stockey (1989) emphasize the close affinity of the Middle Eocene *Allenbya* seeds to *Victoria*. Furthermore, the close affinity of *Eoeuryale* Miki from the Miocene of Japan, Russia and Europe, *Pseudoeuryale* Dorofeev from the Pliocene of Russia and Europe, and *Paleoeuryale* Dorofeev from the Pleistocene of Russia to the extant genus *Euryale* is generally recognized (see Borsch 2000 for review).

Recently, a coalified tiny flower with associated monosulcate–reticulate pollen from the Barremian–Aptian of Portugal has been assigned to Nymphaeales by Friis *et al.* (2001). However, this assignment was based on very few characters that are not unique to Nymphaeales but also allow placing the fossil near Illiciaceae or other angiosperm families

(Gandolfo *et al*. 2004). Gandolfo *et al*. (2004) described a charcoalified flower bud from the Turonian of New Jersey / USA (*Microvictoria svitkoana* Nixon, Gandolfo & Crepet), for which they postulated a close affinity to the extant genus *Victoria*. However, the observed floral characters are not unequivocal and the result of the cladistic analysis of Gandolfo *et al*. (2004) might be biased by the fact that no taxa from outside the Nymphaeales were regarded.

5.1.3 Geological settings

In order to understand the spatial diversification of the Nymphaeales crown group an overview on palaeogeographic conditions during the respective period of earth history is necessary. At the Late Cretaceous–Early Tertiary boundary the break–up of the super–continents Gondwana and Laurasia had been in progress and the continental plates were separating from each other. Thus, it appears necessary to estimate time frames for migration routes or barriers that might have been important for the biogeographic history of Nymphaeales by screening the geological record.

Northern hemisphere

The northern hemisphere was characterized by a more or less continuous land mass in the late Cretaceous, comprising present–day North America, Greenland, Europe and Asia. However extensive epicontinental seaways, which existed periodically due to higher sea levels, impaired transcontinental migrations. The most important barrier of that kind in the Tertiary was the **Turgai Strait**, separating Europe and Asia from Mid–Palaeocene on until its ultimate demise in the Early Oligocene (Tiffney & Manchester 2001). However, although it evidently separated European and Asian faunas during the whole period of its presence (Legendre & Hartenberger 1992), there is evidence for continuous land floras across the southern end of the Turgai strait from the Early Eocene (e.g., Vickulin 1999). From Late Eocene on the Turgai Strait retreated southward, thereby allowing a large–scale floral and faunal interchange between Europe and Asia (Tiffney & Manchester 2001).

Asia and North America were directly connected during the whole Tertiary via the **Bering Land Bridge**. This connection got interrupted rather lately by the opening of the Bering Strait from Late Miocene to Pliocene (7.4 to 4.8 Ma ago, see Tiffney & Manchester 2001). However, given the position of the Bering Land Bridge at very high latitudes (65° today, and 70–80° during the Tertiary, see Smith *et al*. 1994, ODSN 1999), the presence of climatic limitations on migration of plants must be assumed. Tropical elements might have able to use this land bridge only during the Late Palaeocene–Early Eocene climatic optimum (see also Zachos *et al*. 2001), although evidence for subtropical to warm–temperate floras in Alaska and northern Siberia exists throughout the Eocene (see Tiffney & Manchester 2001 and citations therein). In the Oligocene a rapid floristic shift to deciduous, cold–temperate angiosperms and gymnosperms has been observed, which is in accordance with the drastic climatic deterioration during the Oligocene. However, as another — even more important — limiting factor for plant migrations across the Bering Land Bridge the restricted availability of light in these latitudes is discussed. Even though world climates might have been warm enough during the Eocene, long winter darkness might have acted as a filter, allowing only plants adopted to seasonality to cross this barrier (Tiffney 1985, Tiffney & Manchester 2001).

The **North Atlantic Land Bridge** was a direct connection between Europe and North America via southern Greenland and the British Isles that existed from Late Palaeocene to Early Eocene and was located at much lower latitudes than the Bering Strait (approximately 45–50° N, Smith *et al*. 1994, ODSN 1999). Additionally, northern Greenland stayed connected to Fennoscandia and northern Canada until the Late Eocene, but similarly to the Bering Land Bridge this migration route might have been restricted due to limitations in winter–daylight. Although the North Atlantic Land Bridge started to break up in the Eocene, there is fossil evidence for continued floristic exchange between North America and Europe during the Oligocene and Miocene (Axelrod 1975, Hably *et al*. 2000, Denk *et al*. 2005), which might have been possible through island hopping or drops in sea level (Tiffney 1985, Tiffney & Manchester 2001). Both, the Bering Land Bridge and the North Atlantic Land Bridge provided important dispersal routes for thermophilic angiosperms at least during the Late Palaeocene / Early Eocene climatic optimum and allowed the establishment of the so–called boreotropical vegetation (Wolfe 1975) all across the northern hemisphere. Further palaeontological and geological studies are needed to determine if these routes persisted during the whole Palaeogene or even into the Neogene (Collinson & Hooker 2003). However, the drastic climatic cooling at the Eocene–Oligocene boundary put an end to the existence of the boreotropical flora, and its elements either had to escape southward or faced extinction (Morley 2003).

Southern hemisphere

In the southern hemisphere, the most important plant dispersal route during the Cretaceous and the Tertiary ranged from South America across Antarctica to Australia. During the Late Palaeocene / Early Eocene climatic optimum this southern passage probably had been open even for subtropical to tropical elements (Morley 2003). A shallow seaway between East Antarctica and Australia had begun to form in the Palaeogene, but faunal and floral migration was still possible until the complete separation of the continents by Middle to Late Eocene (McLoughlin 2001). Concomitant with successive climatic cooling from the Middle Eocene on, the first ice sheets appeared in Antarctica by the end of the Eocene. The separation of South America from West Antarctica through the opening of the Drake Passage in the Oligocene (32–28 Ma, Lawver & Gahagan 1998) definitely terminated this southern Gondwanan connection. This event allowed the establishment of circumpolar sea currents, which caused drastic cooling of Antarctica and development of extensive ice sheets.

Africa became separated from Eastern Gondwana, i.e. Antarctica, Australia and India, relatively early by the opening of the Indian Ocean in the Late Jurassic (165 Ma, McLoughlin 2001). Direct connections between Africa and South America persisted until the Albian (105 Ma ago, McLoughlin 2001), and migration via volcanic islands on mid–oceanic ridges might have been possible until 95 Ma before present (Raven & Axelrod 1972). There is also some indication for dispersal events from Africa to India via Madagascar until the Late Cretaceous (Morley 2003).

Connections between northern and southern hemisphere

There is evidence for several connections between the Gondwanan continents and Laurasia. A direct land connection between Africa and Europe might have been existed until the

Lower Cretaceous (Smith *et al*. 1994), but occurrence of the common Laurasian *Normapolles* fossils in North Africa hints on enduring floristic interactions across the Tethys until the Late Cretaceous (Herngreen *et al*. 1996). However, this trans-Tethyan dispersal route was severed from at least Palaeocene until Late Eocene (Morley 2003). According to Smith *et al*. (1994) the establishment of new migration pathways from Europe to Africa or vice versa can be assumed for Late Eocene to Miocene via the Iberian peninsula, and also via Arabia since the Middle or Late Miocene. Evidence for plant dispersal from Africa to Europe in the Late Eocene comes from fossil deposits of Caesalpiniaceae in Spain (Cavagnetto & Anadón 1996). However, the uplift the European mountain ranges from Eocene to Miocene probably impaired migrations from Northern Europe to Africa and vice versa (Tiffney & Manchester 2001).

The Indian subcontinent played an important role in floristic interactions between the southern and northern hemisphere. After separating from Madagascar in the Late Cretaceous, it started to move rapidly northward and collided with Asia during Middle Eocene (50-39 Ma, Morley 2003). At the time of collision the SE Asian palynoflora underwent drastic changes as evident from the sudden appearance of Indian and Gondwanan elements and concurrent disappearance of ancient SE Asian pollen types (Morley 1998). Similarly to India, the Australian continent moved north into middle latitudes after its separation from East Antarctica in the Cretaceous and Palaeocene. During the Oligocene, the Australian plate collided with the Philippine and Asian plates (Hall 1998), coinciding with global cooling and drier periods in the SE Asian region (Morley 1998). The uplift of New Guinea and its associated islands in the Miocene (Morley 2003) provided a dispersal pathway for tropical elements from South East Asia, that invaded New Guinea and later dispersed into northern Australia. Also there is evidence for several dispersal events from Australia into South East Asia across the Wallace line during Miocene (Morley 1998). In fact, a complex intermingling of South East Asian and Australian floras can be observed (Wiffin 2002).

In the western hemisphere the connections between North and South America were most important for development and dispersal of biota (Pennington & Dick 2004). According to Morley (2003), there is evidence for biotic exchange during Late Cretaceous and Palaeocene from several plant lineages and animals. By that time the eastern edge of the Caribbean plate had formed an island arc or land bridge between Yucatan and Northern South America. This dispersal route was open until the Middle Eocene when the Caribbean plate drifted further east with the leading arc forming the modern Caribbean islands. However, the Caribbean arc is thought to have formed a land mass during the Eocene and Oligocene with persistent connection to South America and a short–lived connection to North America at the Eocene–Oligocene boundary (35–33 Ma ago, Iturralde-Vinent & MacPhee 1999). Due to further eastward movement of the Caribbean plate, the island arc at its western edge collided with the southern end of North America during the Middle Tertiary and with South America in the Early Miocene (Pindall *et al*. 1988). A few dispersals via this island arc might have happened in the Miocene (Bermingham & Martin 1998), but major faunal and floral migrations took place in the Late Pliocene when a continuous land bridge was formed due to subsequent uplift. The Pliocene was the time of the "Great American Interchange", which included mainly mammal faunas (e.g., Webb 1991), but there is also evidence from several plant lineages (Morley 2003).

5.1.4 Aims of this study

Based on hypotheses on the phylogeny of the Nymphaeales, and considering earth history as well as the fossil record, the following questions will be addressed: Is it possible to determine the origin of the Nymphaeales crown group geographically? Is the radiation of core Nymphaeaceae (*Nymphaea*, *Ondinea*, *Victoria*, *and Euryale*) correlated with geological or climatic events; and where did this radiation start? How can the disjunct ranges of closely related taxa (e.g., *Victoria* – *Euryale*, *N.* subg. *Hydrocallis* – subg. *Lotos*) be explained? Dispersal–vicariance analysis (Ronquist 1996, 1997) is used to infer the potential distribution at ancestral nodes in the Nymphaeales phylogeny. Due to recent results from comprehensive phylogenetic studies (Borsch *et al.* subm., Löhne *et al.* subm. and Chapter 3), it is — for the first time — possible to address the complex phylogenetic and biogeographic patterns in core Nymphaeaceae. The approach of analysing alternative topologies allows assessing the effect of taxon sampling and phylogenetic uncertainties on the inference of biogeographic history.

5.2 Materials and Methods

5.2.1 Taxon sampling and phylogenetic hypotheses

The present study of historical biogeographic relationships covers all major lineages of Nymphaeales. Analyses were conducted using two different taxon sets (16 vs. 33 taxa) and corresponding topologies A and B. With this approach the results of two most recent studies on Nymphaeales (Borsch *et al.* subm., Löhne *et al.* subm.) are applied, and the effect of the alternative phylogenetic hypotheses on reconstruction of ancestral areas can be evaluated.

The taxonomic units used in the first set of analyses (16 taxa, topology A) are the genera *Cabomba*, *Brasenia*, *Nuphar*, *Barclaya*, *and Euryale*, *Victoria*. In the case of the species–rich genus *Nymphaea*, its five subgenera have been regarded as operational units. *Ondinea*, which is nested among species of *Nymphaea* subg. *Anecphya* (Borsch *et al.* subm., Löhne *et al.* subm.), is not included as a distinct taxon here but is subsumed under *Nymphaea* subg. *Anecphya*. Representatives of Austrobaileyales (*Austrobaileya*, *Illicium*, *Schisandra* and *Kadsura*) and *Amborella*, all members of the basal grade of angiosperms (see description in Chapter 1.2.4), were chosen as outgroup taxa. The topology of Nymphaeales found in Chapter 3 (Löhne *et al.* subm.) was used for the 16–taxa constraint tree (topology A, Figure 5.1). In topology A, the genus *Nymphaea* is not monophyletic but the *Victoria–Euryale* clade is sister to a *Hydrocallis–Lotos* clade.

Topology B (Figure 5.2), which is based on the phylogenetic reconstruction of Borsch *et al.* (subm.), differs from topology A mainly in the position of the *Victoria–Euryale* clade, which is here sister to a clade comprising all lineages of *Nymphaea*. Furthermore, *Nymphaea* subg. *Brachyceras* is paraphyletic with respect to subg. *Anecphya*. Taxon sampling of the ingroup was increased to 33 in this second set of analyses in order to study possible effects of taxon sampling on the reconstruction of ancestral areas. Four species for *Nuphar* and two to six

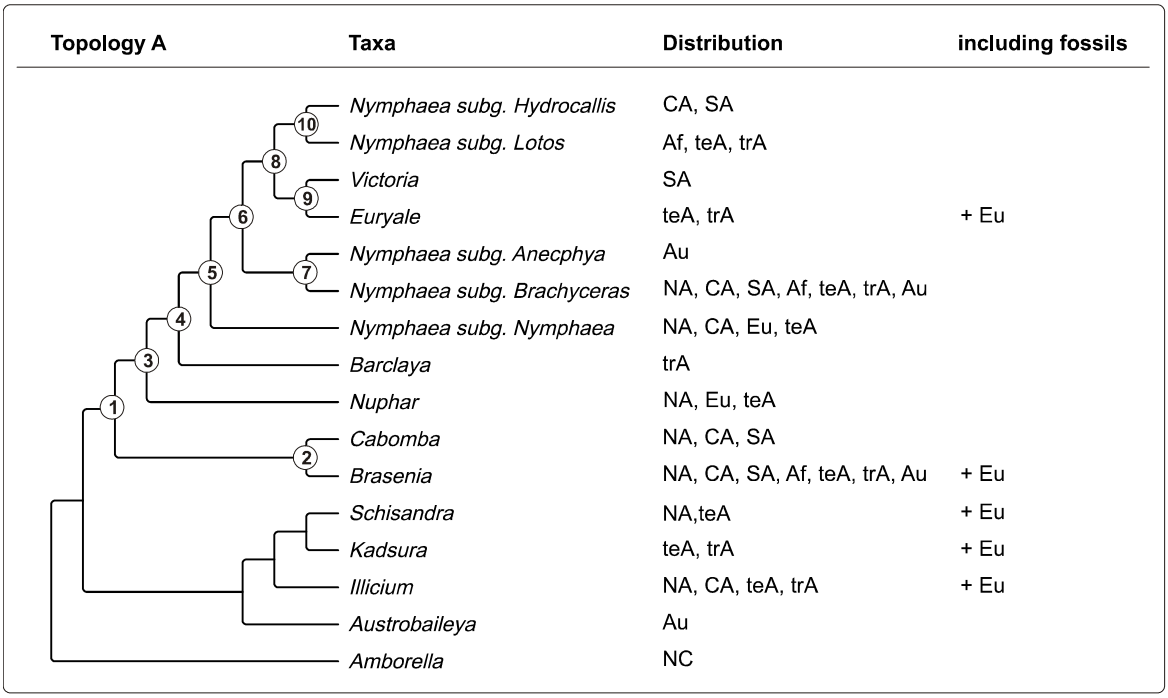


Figure 5.1 — Simplified phylogeny of Nymphaeales used as constraint topology A (16–taxa). This constraint tree was drawn following the results of the phylogenetic analysis presented in Chapter 3 (Löhne *et al.* subm.). Relevant nodes are numbered (in circles). The distribution of taxa, as coded for dispersal–vicariance analysis, is given behind the taxon names. A second matrix was constructed that regarded also information from the fossil record (thereby adding Europe to the distribution of *Euryale*, *Brasenia*, *Kadsura*, *Schisandra* and *Illicium*). NA = North America, CA = Central America, SA = South America, Eu = Europe, Af = Africa, teA = temperate Asia, trA = tropical Asia, Au = Australia, NC = New Caledonia.

species for each subgenus of *Nymphaea* were included. This second data set represents a subset of the 65–taxa data set of Borsch *et al.* (subm.). Exclusion of taxa was necessary because the tree of Borsch *et al.* (subm.) was not fully resolved but only bifurcations are allowed for DIVA analyses (see below). Species were chosen to represent the different ranges present in *Nuphar* or the subgenera of *Nymphaea*, respectively.

5.2.2 Estimation of divergence times

Molecular dating analyses were conducted following the methods described by Yoo *et al.* (2005). Analyses were run with the 29–taxa *matK* data set of Löhne et al. (subm.). The following four sequences were added to the alignment*: Larix gmelini* (Rupr.) Rupr. AF143433 (Wang *et al.* 2000), *Taxus wallichiana* Zucc. var. *mairei* (Lemée et H. Lév.) L.K. Fu & Nan Li AB024001 (Cheng *et al.* 2000), *Ginkgo biloba* L. AF456370 (Quinn *et al.* 2002), *Nuphar variegata* Engelm. ex Durand AF092979 (Les *et al.* 1999). The inclusion of gymnosperms as outgroup taxa allowed using the estimated age of angiosperms (131.8 Ma, see Soltis *et al.* 2002b, Yoo *et al.* 2005 and Chapter 1.3.1 for explanation and references) as a calibration point. Branch lengths were optimized under Maximum Parsimony (MP) as well as under Maximum Likelihood (ML) criteria in PAUP* (Swofford 2002), applying the GTR+G model of molecular evolution (as estimated in Chapter 3.2.4, page 58 f.) to the ML tree. Non–parametric rate smoothing (NRPS, Sanderson 1997) was conducted for both trees using TreeEdit v1.0 alpha 10 (Rambaut & Charleston 2002).

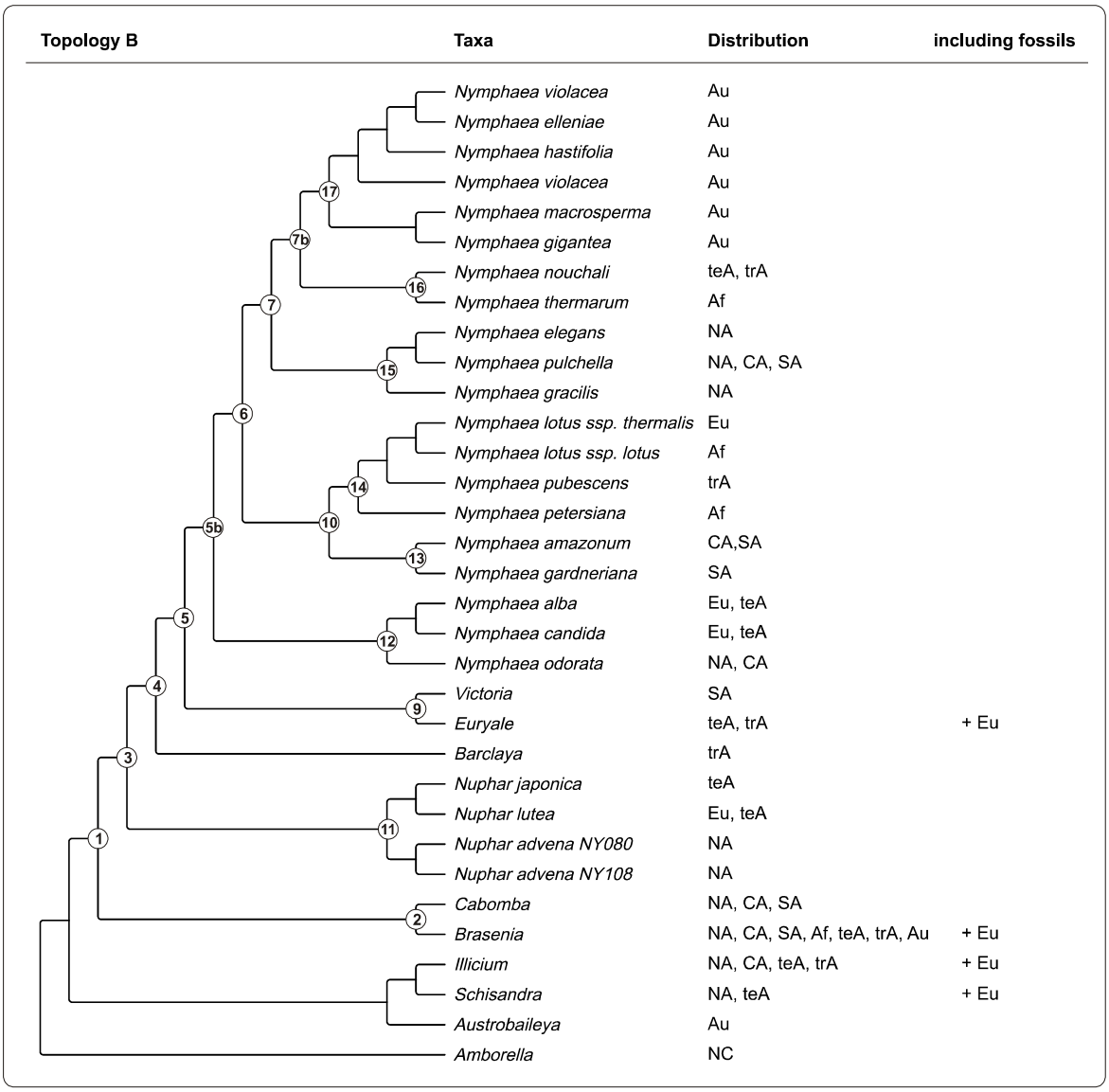


Figure 5.2 — Alternative tree of Nymphaeales used as constraint topology B (33–taxa). This tree is based on the *trnT–trnF* data set of Borsch *et al.* (subm.), but with reduced taxon sampling. Relevant nodes are numbered (in circles). The distribution of taxa, as coded for dispersal–vicariance analysis, is given behind the taxon names. Similar to the analysis of topology A, a second matrix was constructed regarding additional information from the fossil record (last column in the Figure). See Figure 5.1 for abbreviations.

5.2.3 Extant and fossil distribution

Distribution data from 69 species (Appendix 6) were compiled from floras (Appendix 7). As an additional major source of pre–compiled distribution data the online database of the Germplasm Resources Information Network (USDA, ARS) was used (http://www.ars-grin.gov/cgi-bin/npgs/html/gnlist.pl?185 for Cabombaceae, and http://www.ars-grin.gov/cgi-bin/npgs/html/gnlist.pl?779 for Nymphaeaceae). Species ranges were digitized using the geographic information software ArcView 3.2. To obtain an overview about the global diversity of Nymphaeales, presence and absence of every single species was projected to a 1° x 1° grid (using the "intersect" option of the Geoprocessing Wizard in ArcView 3.2.),

resulting in a map of 11,824 grid squares with 40,153 positive grid records within the natural range of the Nymphaeales.

To produce distribution maps for each genus of Cabombaceae and Nymphaeaceae as well as for the subgenera of *Nymphaea*, maps of the respective species were merged. *Nymphaea petersiana*, currently classified under subg. *Brachyceras*, was here subsumed under subg. *Lotos* because recent molecular studies revealed a close affinity of *N. petersiana* to *N. lotus* and *N. pubescens* (Bosch *et al.* 2000, Borsch *et al.* subm., Löhne *et al.* subm., see also Chapter 3). If there were fossils that could be clearly assigned to the extant genera, their localities have been marked on the maps in order to illustrate ancient distribution of the respective taxon. Extinct genera or fossil species of uncertain affinity were not regarded here.

5.2.4 Dispersal–vicariance analysis

In order to reconstruct the ancestral distribution and to identify possible dispersal events in the biogeographic history of Nymphaeales, dispersal-vicariance analyses (Ronquist 1997) were conducted as implemented in the computer programme DIVA version 1.1 (Ronquist 1996). DIVA reconstructs ancestral distributions in a given phylogeny without any prior assumptions about relationships between unit areas, thereby allowing reticulate as well as hierarchical relations among areas over a given period of time (Ronquist 1997). The algorithm searches for the optimal reconstruction of ancestral distributions by applying a three-dimensional cost matrix with the following assumptions: (1) Events of speciation by vicariance, i.e. separating a wider distribution into two mutually exclusive sets of areas, receive the cost of zero. (2) Duplication events, i.e. speciation within a single area, have the cost of zero. (3) Dispersal events, i.e. addition of new areas to the range of a taxon, have the cost of 1. (4) Extinction events, i.e. deletion of areas from the range of a taxon, also receive the cost of 1. Thus, the optimal solutions given by DIVA minimize the number of implied dispersal and extinction events and favour vicariance. Optimal ancestral distributions also have to fulfil the following two conditions: they cannot contain an area not occupied by any descendant, and they must include at least one unit area from each of the descendants (Ronquist 1996).

However, although DIVA seems to be the best currently available approach in historical biogeography it also has its weaknesses: extinction events are effectively not inferred under standard settings (Austin *et al.* 2003), and the reconstructed distribution at the root node is the least reliable, with the root ancestor usually being inferred as widely distributed due to the abovementioned conditions (Ronquist 1996). Since one objective of this study is to identify the geographic origin of the Nymphaeales crown group and of major nymphaealean lineages, it was attempted to solve this problem by analysing the data sets with and without constraints on the number of unit areas allowed in ancestral distributions. Using the "maxareas" option of the "optimize" command in DIVA, the number of unit areas was set to 9 (unconstraint), 5 and 2 for each of the two data set.

Nine unit areas were defined, that cover the ranges of all genera of Nymphaeales as well as the distribution of the outgroup taxa:

(A) *North America* (NA): from Alaska to Mexico north of the Isthmus of Tehuantepec;

(B) *Central America* (CA): from the Isthmus of Tehuantepec (Mexico) to Panama, including the Caribbean Islands;

(C) *South America* (SA): America south of Panama;

(D) *Europe* (Eu): including the European parts of Turkey and Russia;

(E) *Africa* (Af): including Madagascar;

(F) *temperate Asia* (teA): from the Arabian peninsula and the Caucasus region to China and Russian far east;

(G) *tropical Asia* (trA): from east Pakistan and India south of the Himalayas to South East Asia, covering all islands of the Indo-Malayan region including New Guinea;

(H) *Australia* (Au);

(I) *New Caledonia* (NC).

Based on the distribution maps shown in Figures 5.3–5.5, all taxa were coded for presence or absence in each unit area. However, to avoid too many widespread taxa and wrong conclusions, a taxon was not coded as present in an area if this area is only marginally touched by the natural range of this taxon. This was the case for *Nuphar* in Africa (*N. lutea* in northern Algeria) and Central America (*N. advena* in Cuba), *Nymphaea* subg. *Anecphya* in tropical Asia (some species occur also in New Guinea), and *Nymphaea* subg. *Hydrocallis* in North America (*N. jamesoniana* in Florida and, together with a few other widespread species, along the coast of Mexico). Two data sets were created by using the two topologies A and B, which are shown in Figure 5.1 and Figure 5.2 together with the distributions assigned to each terminal taxon. As mentioned before, some genera of Nymphaeales were present in Europe during the Tertiary but went extinct during the glaciation period. To account for this fossil distribution additional data sets were analysed for each topology, with those taxa coded as present in Europe where unambiguous fossils are known (see last columns in Figure 5.1 and Figure 5.2). Regarding the outgroup taxa there are fossils of *Illicium* (Mai 1970, 1995), and of *Kadsura* and *Schisandra* (Mai 1995, Denk & Oh 2006) known from the Tertiary of Europe. Thus, their fossil distribution has also been considered in the second set of analyses. All original data files can be obtained from the author.

To illustrate possible historical scenarios on the evolution and biogeography of Nymphaeales, maps were created that show the respective position of continental plates at different time frames. For this purpose the "Plate Tectonic Reconstruction Service" of the Ocean Drilling Stratigraphic Network (run by the GEOMAR Research Center for Marine Geoscience / Kiel and the Geological Sciences Department of Bremen University) was used, which is available online at www.odsn.de. This software is based on the plate tectonic data sets of Hay *et al.* (1999).

5.3 Results

5.3.1 Extant distribution

Maps of the extant distribution of the genera of Cabombaceae and Nymphaeaceae, including each subgenus of *Nymphaea*, are shown in Figures 5.3–5.5. Additionally, fossil locations have been marked by stars for *Brasenia* (fossil seeds of the genus from Europe and Asia), *Nuphar* (fossil seeds from North America, Europe, Western Russia and Siberia), *Euryale* (*Eoeuryale* in Japan, Russia and Europe, *Pseudoeuryale* in Russia and Europe, and *Paleoeuryale* in Russia) and *Nymphaea* (fossil seeds, rhizomes and leaves from Europe, Russia and Siberia). The location of a recently described leave fossil that has been assigned to the genus *Nymphaea* is marked on the map (Argentina), although its affinity to extant the extant genus is not clear (see paragraph 5.1.2, page 112 f.).

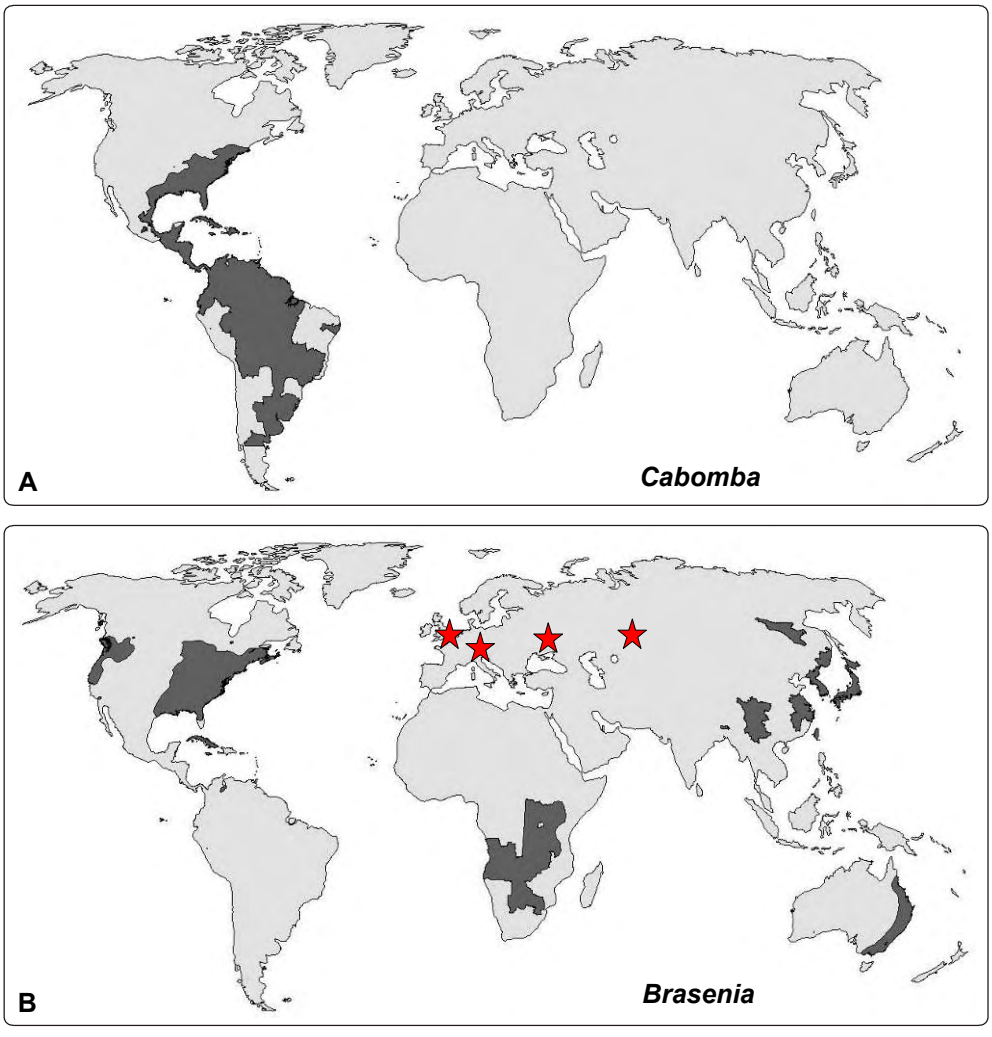


Figure 5.3 — Extant distribution of Nymphaeales. I, Cabombaceae. A) *Cabomba* (5 species), **B)** *Brasenia* (1 species). Extant distribution is shaded in dark grey; localities of fossils are marked by red stars. Note that several fossil localities of *Brasenia* are known for Europe and Russia, but for simplicity only single symbols have been inserted. No fossils are known from Cabomba.

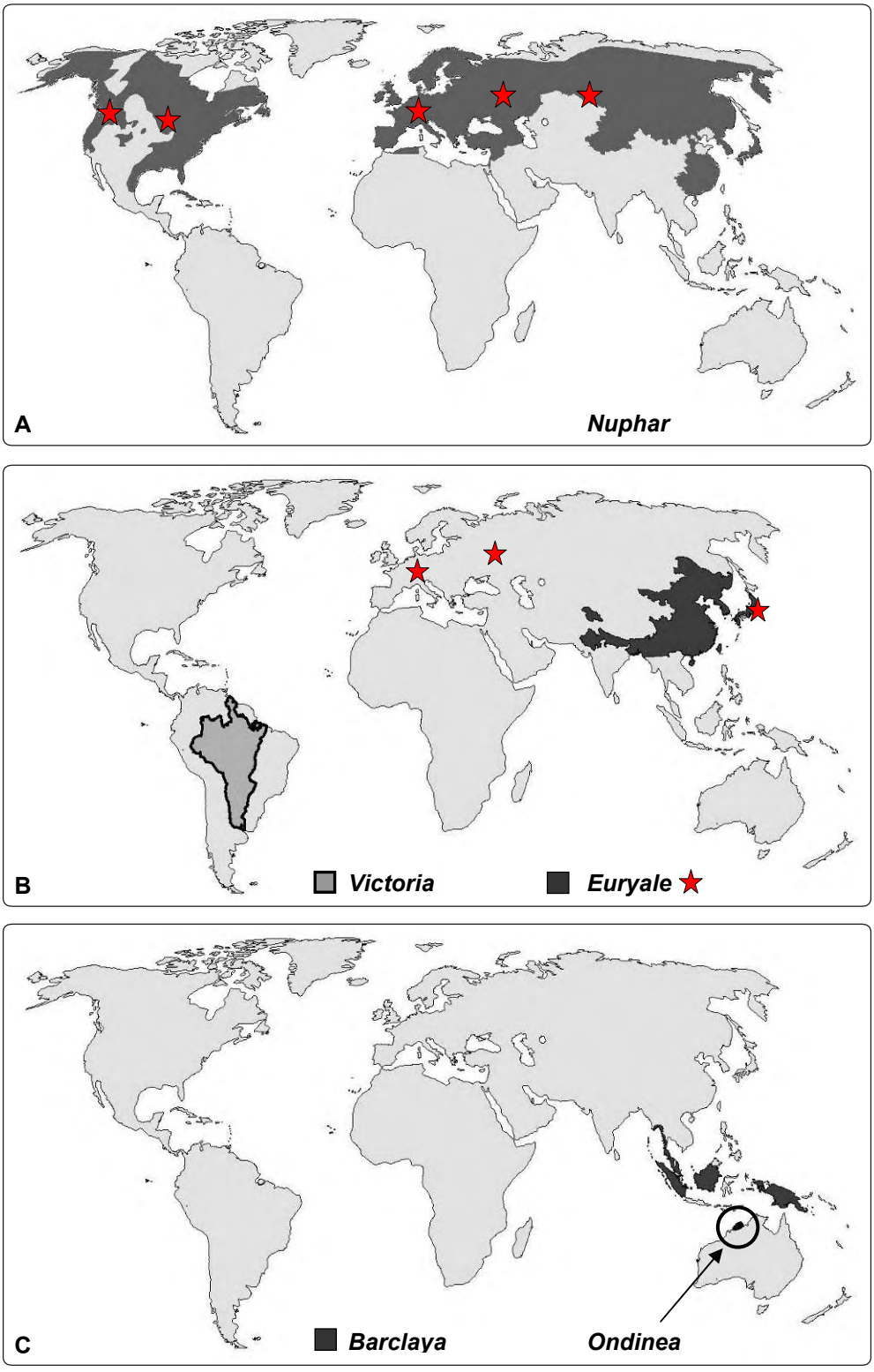


Figure 5.4 — Extant distribution of Nymphaeales. II, Nymphaeaceae. A) *Nuphar* (8 species), **B)** *Victoria* (2 species), *Euryale* (1 species), **C)** *Barclaya* (4 species), *Ondinea* (1 species). Extant distribution is shaded in grey; localities of fossils are marked by red stars. Note that several fossil localities of *Nuphar* and *Euryale* are known for Europe and Russia, but for simplicity only single symbols have been inserted. No fossils are known from *Victoria*, *Barclaya* and *Ondinea*.

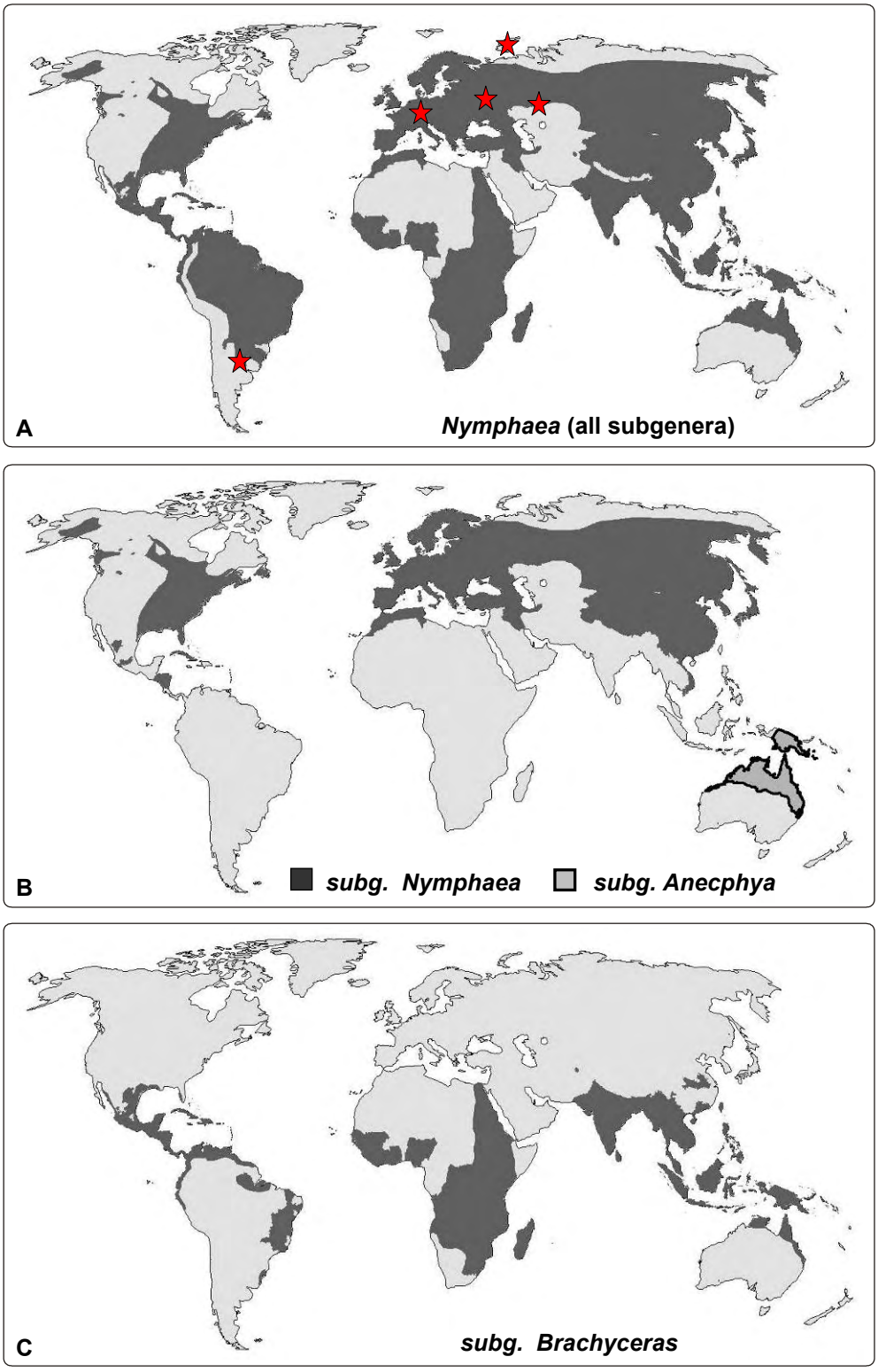


Figure 5.5 — Extant distribution of Nymphaeales. III, The genus *Nymphaea*. A) *Nymphaea* (all subgenera, 47 species), B) subg. Nymphaea (7 species), subg. *Anecphya* (7 species), **B)** subg. *Brachyceras* (16 species), **D, next page)** subg. *Hydrocallis* (14 species), subg. *Lotos* (3 species).

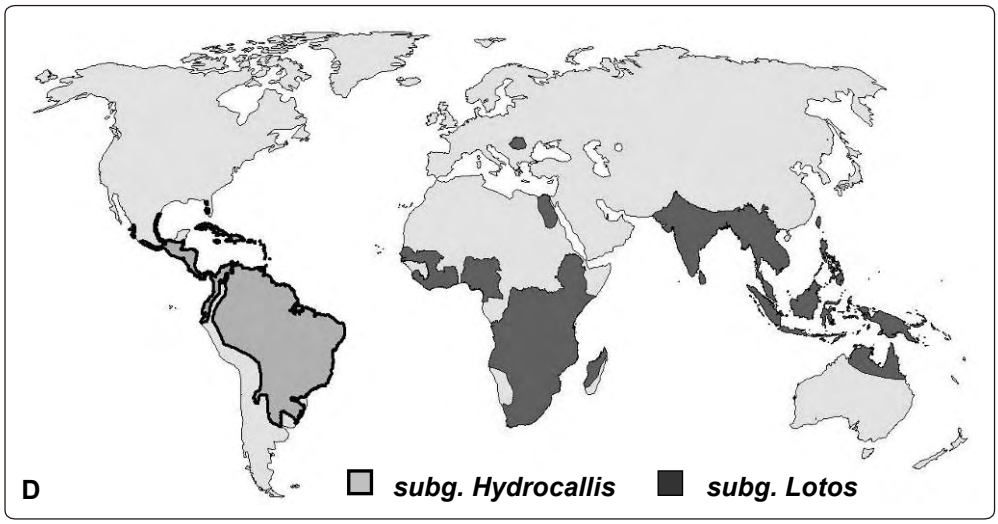


Figure 5.5 (continued) — D) subg. *Hydrocallis* (14 species), subg. *Lotos* (3 species).

Figure 5.6 shows the extant species diversity of Nymphaeales on a global scale. Northern South America, Central America, the Zambezian region of Africa (Botswana, Zambia, East Angola) and northern Australia appear as centres of diversity for Nymphaeales.

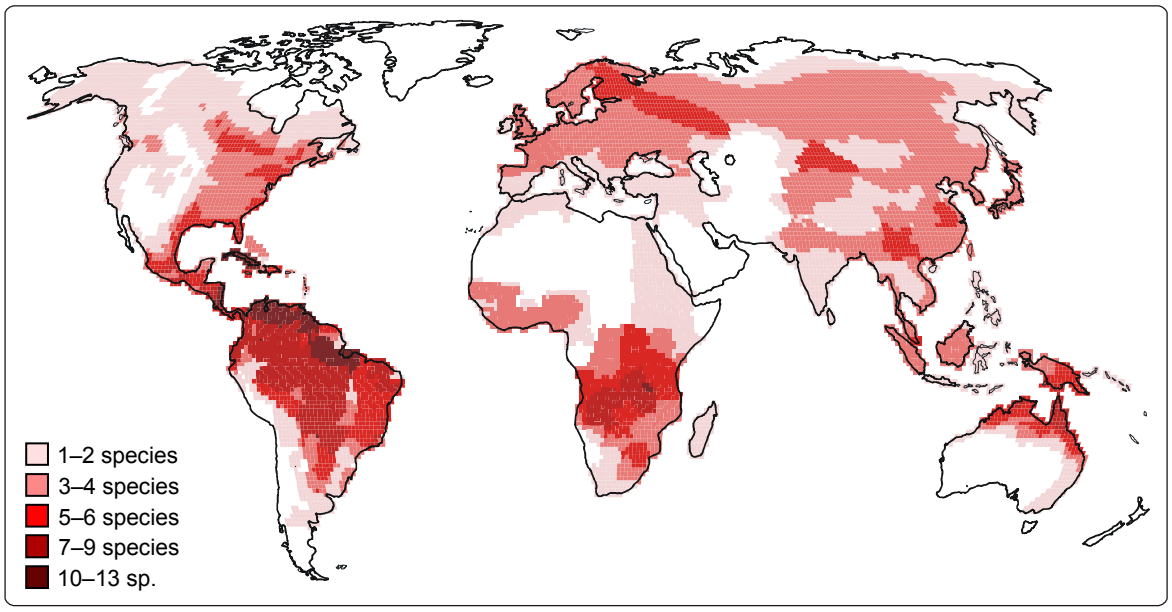


Figure 5.6 — Extant species diversity of Nymphaeales plotted on the world map. This figure is based on distribution data for each of the 69 currently accepted species of Nymphaeales, as interpolated from information in floras, the online database of the Germplasm Resources Information Network (USDA, ARS) and personal observations of J. Wiersema (see Chapter 5.2.3. for description of methods and resources).

5.3.2 Estimated divergence times

Figure 7 illustrates the calibrated tree of Nymphaeales with divergence times obtained from non–parametric rate smoothing (NRPS) using both MP and ML branch lengths. The age of the Nymphaeales crown group is estimated to 63.2–64.3 Ma which corresponds to the beginning of the Tertiary (65 Ma). Regarding the Cabombaceae crown group, the split between Cabomba and Brasenia is inferred to be between 39.2 and 39.7 years old. According to the present data set, *Nuphar* became separated from the remaining Nymphaeaceae rather shortly after the initial diversification of the Nymphaeales crown group (59.5 Ma from MP, 64.3 Ma from ML). The radiation of the genus *Nymphaea* and *Victoria–Euryale* is dated to Late Eocene / Early Oligocene (33.1–30.2 Ma from MP, 33.4 Ma from ML).

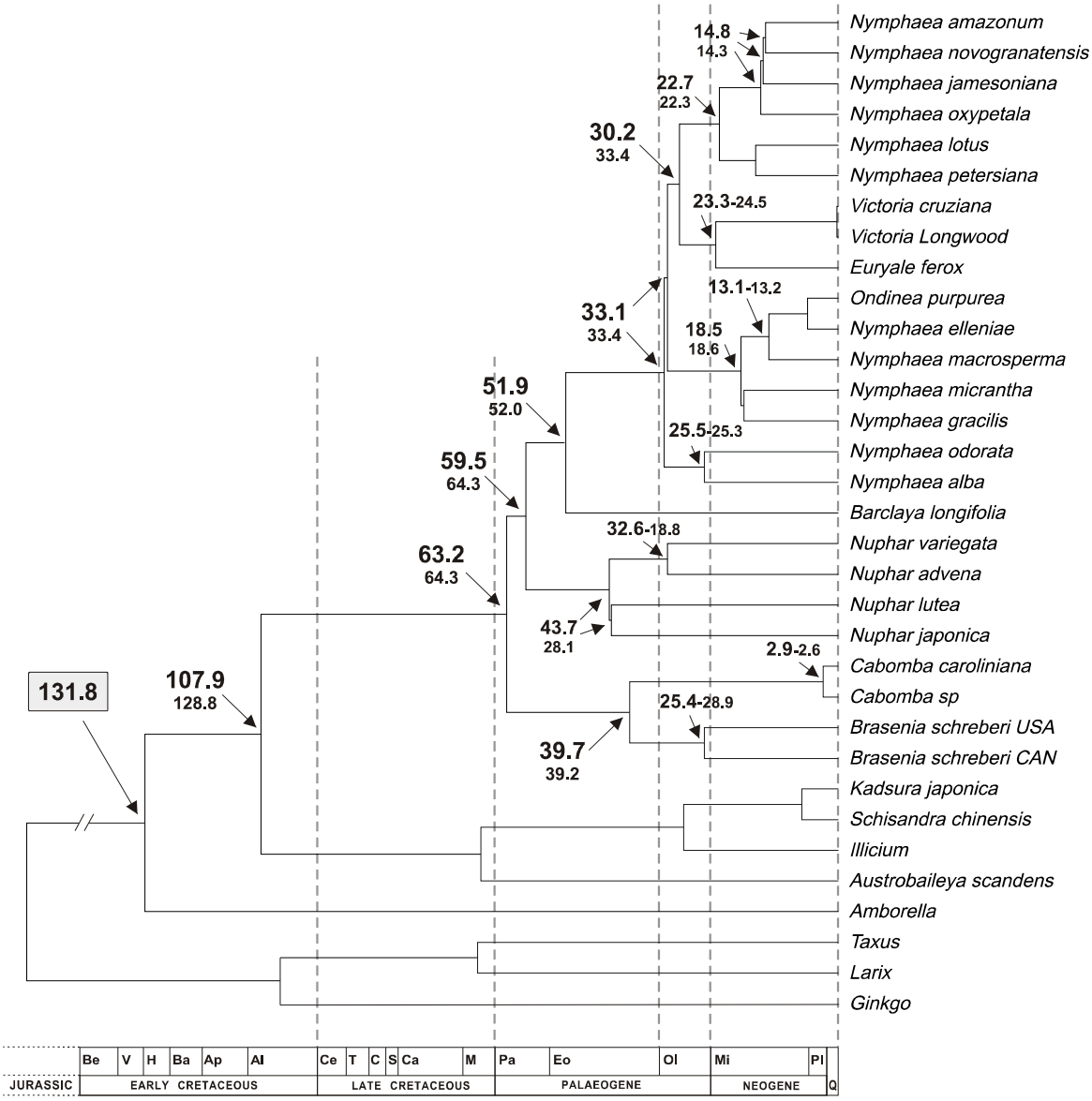


Figure 5.7 — Divergence times estimates for Nymphaeales. The phylogeny obtained from the multi–locus chloroplast data set by Löhne *et al.* (subm., see Chapter 3) was used to construct the constraint tree for this dating analysis. NPRS (Sanderson 1997) was conducted with MP– and ML–branch lengths, using the age of angiosperms (131.8 Ma) as a calibration point. Estimated ages from MP (upper / first values) and ML (lower / second values) are given for nodes in Nymphaeales.

For most of the nodes, calibrations from the tree with MP– or ML–branch lengths yield very similar ages (varying from 0.1 to 4.8 Ma). However, age estimations for the divergence of the Nymphaeales stem lineage from the rest of angiosperms differ noticeably between the two approaches: the node is inferred to be 107.9 Ma old from the MP tree, whereas the ML tree suggests a considerably older age of 128.8 Ma. This discrepancy is thought to result from the differences in the MP and the ML approaches in allocate character state changes between outgroup and ingroup either along the basal ingroup branches (MP) or the outgroup branches (MP). Another discrepancy between age estimates from MP and ML can be observed in the *Nuphar* crown group where ML estimates reveals considerably younger ages than MP (28.1 Ma from ML vs. 43.7 Ma from MP).

5.3.3 Inferred ancestral areas

Dispersal–vicariance analyses using DIVA have been conducted with four different data sets: 1) topology A with 16 terminal taxa, only extant distribution coded, 2) topology A with 16 terminals, extant and fossil distribution coded, 3) topology B with 33 terminals, only extant distribution coded, and 4) topology B with 33 terminals, extant and fossil distribution coded. The reconstructions are summarized in Table 5.1 for topology A and in Table 5.2 for topology B. All analyses resulted in multiple combinations of optimal solutions. The number of possible reconstructions on a node varied from one to 103, depending on whether the number of areas was constrained or not. Thus, for reasons of clarity, the tables are meant to give an overview about possible reconstructions. The complete output of all DIVA analyses is compiled in Appendix 8.

As predicted by the characteristics of the DIVA algorithm (see Materials and Methods), the ancestors are inferred to be widespread if the number of areas inhabited by the ancestor was not constrained. However, in analyses that consider the fossil distribution of extant taxa in Europe the reconstructions are much more precise for ancestral nodes at the backbone of the Nymphaeaceae phylogeny (nodes 3–6, and 8). In topology A (Table 5.1), the inferred ancestral areas at nodes 3, 5 and 8 are either Europe (D) or temperate Asia (F), irrespective of constraints on the number of ancestral areas. For node 4, i.e. all Nymphaeaceae above *Nuphar*, an ancestral distribution of Europe + tropical Asia (DG) or temperate Asia + tropical Asia (FG) was reconstructed when fossils were included. For node 6 (i.e. all subgenera of *Nymphaea* including *Victoria–Euryale*, except subg. *Nymphaea*) the inferred ancestral distribution is temperate Asia, temperate Asia + Australia (FH) or Europe + Australia (DH). However, inclusion of fossil distributions does not bring more clarity to the inferred ancestral distributions at node 1 (ancestor of Nymphaeales) and node 2 (ancestor of Cabombaceae). These ancestors are inferred to be extremely widespread in both analyses, if no constraints on the number of areas are invoked. Reduction of ancestral areas to two results in several combinations of northern hemispheric areas (North America, Europe, temperate Asia), and Central and South America, for nodes 1 and 2 (Table 5.1, column 3).

Table 5.1 — Results of DIVA analyses using topology A. The maximum number of areas in ancestral distributions was constrained to two (maxareas = 2) or four (maxareas = 5), or was left unconstrained (maxareas = 9). Ancestral areas are shown for all nodes in Nymphaeales (see Fig. 1), as inferred from the data set with only extant distribution coded ("extant", first line) and as inferred from data set with extant and fossil distributions coded ("*fossil*", second line, italics). Alternative reconstructions are separated by slashes. Note that, for reasons of clarity, only a summary of the solutions is given, if number of solutions was too high to be displayed here. The complete output of DIVA analyses is compiled in Appendix 8. "No. of dispersals" – number of inferred dispersal events (in each analysis). A = North America, B = Central America, C = South America, D = Europe, E = Africa, F = temperate Asia, G = tropical Asia, H = Australia.

		Inferred ancestral areas for topology A (using different constraints on maximum number of areas**)**		
node	**data set**	**maxareas = 2**	**maxareas = 5**	**maxareas = 9 (unconstrained)**
1	extant *fossil*	F / AF / BF / CF; *F / AF / BF / CF / D / AD/ BD / CD*	77 combinations from A to ABFGH; *AD / AF / BD / BF / CD / CF*	6 combinations from ABCDEG to ABCDEFGH; *same result*
2	extant *fossil*	A / B / C / AF / BF / CF; *A / B / C / AF / BF / CF / AD / BD / CD*	101 combinations from A to ABFGH, *A / B / C*	16 combinations from BCE to ABCEFGH; *8 combinations, all widespread*
3	extant *fossil*	F; *D / F*	13 combinations from A to ADFG; *D / F*	D / F / DF / DG / ADF / ADG / DFG / ADFG; *D / F*
4	extant fossil	FG; *DG / FG*	43 combinations from G to ADFGH; *DG / FG*	34 combinations from G to ABCDFGH; *DG / FG*
5	extant *fossil*	F; *D / F*	73 combinations from A to BDFGH; *D / F*	60 combinations from D to ABDFGH; *D / F*
6	extant *fossil*	F / FH; *F / FH / DH*	81 combinations from F to DEFGH; *F / DH / FH*	38 combinations from F to BCDEFGH; *F / DH / FH*
7	extant *fossil*	H / FH; *same result*	54 combinations from H to ABFGH; *27 combinations from H to ABFGH*	57 combinations from H to ABCEFGH; *33 combinations from H to ABCEFGH*
8	extant *fossil*	F; *D / F*	30 combinations from C or F to BDEFG; *D / F*	31 combinations from C or F to BCDEFG; *D / F*
9	extant *fossil*	CF; *CD / CF*	CF / CG / CFG; *CD / CF / CDF / CDG / CFG / CDFG*	CF / CG / CFG; *CD / CF / CDF / CDG / CFG / CDFG*
10	extant *fossil*	BF / CF; *BF / CF / BD / CD*	43 combinations from BD to BDEFG; *35 combination from BD to BDEFG*	44 combinations from BD to BCDEFG; *36 combination from BD to BCDEFG*
No. of dispersals		35 – extant, 40 – fos.	34 – extant, 38 – fos.	31 – extant, 35 – fos.

Table 5.2 — Results of DIVA analyses using topology B. The maximum number of areas in ancestral distributions was constrained to two (maxareas = 2) or four (maxareas = 5), or was left unconstrained (maxareas = 9). Ancestral areas are shown for relevant nodes in Nymphaeales, as inferred from the data set with only extant distribution coded ("extant", first line) and as inferred from data set with extant and fossil distributions coded ("fossil", second line, italics). Alternative reconstructions are separated by slashes. Note that, for reasons of clarity, only a summary of the solutions is given if the number of solutions was too high to be displayed here. The complete output of DIVA analyses is compiled in Appendix 8. No. disp. refers to the number of inferred dispersal events. A = North America, B = Central America, C = South America, D = Europe, E = Africa, F = temperate Asia, G = tropical Asia, H = Australia.

		Inferred ancestral areas for topology B (using different constraints on maximum number of areas**)**		
node	**data set**	**maxareas = 2**	**maxareas = 5**	**maxareas = 9 (unconstrained)**
1	extant *fossil*	A / AF; *A / F / AF / BF / CF*	56 combinations from A to ABFGH; *AF / BF / CF*	ABCEFG / ABCDEFG / BCEFGH /ABCEFGH; *6 combinations, all widespread*
2	extant *fossil*	A; *A / B / C / AF / BF / CF*	103 combinations from A to ABFGH; *A / B / C*	9 combinations from A to ABCEFGH; *12 combinations from ABCEG to ABCDEFGH*
3	extant *fossil*	A / F / AF / AG; *A / F / AF / AG / AD*	A / F / AF / AG / FG / AFG; *F*	F / ABDEFG / ACDEFG / BCDEFG / ABCDEFG; *F*
4	extant *fossil*	G / AG / FG; *G / AG / FG / DG*	28 combinations from G to ABDFG; *FG*	13 combinations from FG to ABCDEFG; *FG*
5	extant *fossil*	F / G / AC / AF / AG / BG / DG / EG / FG; *same combinations, plus D / AD*	80 combinations from F to ADEFG; *F*	23 combinations from F to ABCDEFG; *F*
5 b	extant *fossil*	A / F / AE / BE / DE / FE / AG / BG / DG / FG; *same results*	79 combinations from A to DEFGH; *F / EF / AEF / BEF / ABEF*	F / DE / EF / ADE / AEF / BDE / BEF / ABDE / ABEF; *F / EF / AEF / BEF / ABEF*
6	extant *fossil*	E / G / AC / AE / CF / CG / EF / EG; *same results*	43 combinations from E to AEFGH; *24 combinations from E to AEFGH*	25 combinations from E to ACEFGH; *same results*
7	extant *fossil*	AE / AF / AG / AH; *same results*	15 combinations from AE to AEFGH; *12 combinations from AE to A*EFGH	12 combinations from AE and AF to AEFGH; *same results*
7 b	extant *fossil*	EH / FH / GH; *same results*	EH / FH / GH / EFH / EGH / FGH / EFGH; *EH / FH / EFH / EGH / FGH / EFGH*	EH / FH / EFH / EGH / FGH / EFGH; *same results*
9	extant *fossil*	CF / CG; *CF / CG / CD*	CF / CG / CFG; *CF / CDF / CFG / CDFG*	CF / CG / CFG; *CF / CDF / CFG / CDFG*

Table 5.2 (continued)

node	data set	**Inferred ancestral areas for topology B** (using different constraints on maximum number of areas)		
		maxareas = 2	**maxareas = 5**	**maxareas = 9 (unconstrained)**
10	extant *fossil*	CE / CG; *same results*	CE / CG / CEG; *CE*	CE; *same result*
11	extant *fossil*	AF; *same result*	AF; *same result*	AF; *same result*
12	extant *fossil*	AD / BD / AF / BF; *same results*	AD / AF / BD / BF / ABD / ABF; *AF / BF / ABF*	AD / AF / BD / BF / ABD / ABF; *AF / BF / ABF*
13	extant *fossil*	C; *same result*	C; *same result*	C; *same result*
14	extant *fossil*	E / EG; *same results*	E / EG; *E*	E; *same result*
15	extant *fossil*	A; *same result*	A; *same result*	A; *same result*
16	extant *fossil*	EF / EG; *same results*	EF / EG / EFG; *same results*	EF / EG / EFG; *same results*
17	extant *fossil*	H; *same result*	H; *same result*	H; *same result*
No. of dispersals		33 – ext., 38 – fos.	32 – ext., 36 – fos.	29 – ext., 33 – fos.

Ancestral distributions at the basal nodes (nodes 1–5) inferred from topology B are congruent to those inferred from topology A. Again, ancestors are inferred to be extremely widespread if no constraints on the maximum number of areas are invoked in topology B. With the maximum number of areas set to two, reconstructions for node 1 and 2 include different combinations of North, Central and South America with Europe and Asia (Table 5.2, column 3), which is comparable to the results of topology A. Also for nodes 3–5 similar combinations have been reconstructed. The inclusion of fossil data reduced the inferred ancestral areas for nodes 3–5 to temperate Asia or temperate + tropical Asia (Table 5.2, column 5). Ancestral areas involving South America (AC) for node 5 (core Nymphaeaceae), as inferred under the maxareas = 2 assumption, are regarded to be unlikely since South America was not connected to the northern hemisphere by the time of diversification of the core Nymphaeaceae (Late Eocene / Early Oligocene). Ancestral distributions for the common ancestor of *Victoria* and *Euryale* (node 9) deduced from both topologies include South America and tropical or temperate Asia or a combination of all three areas. Europe is added

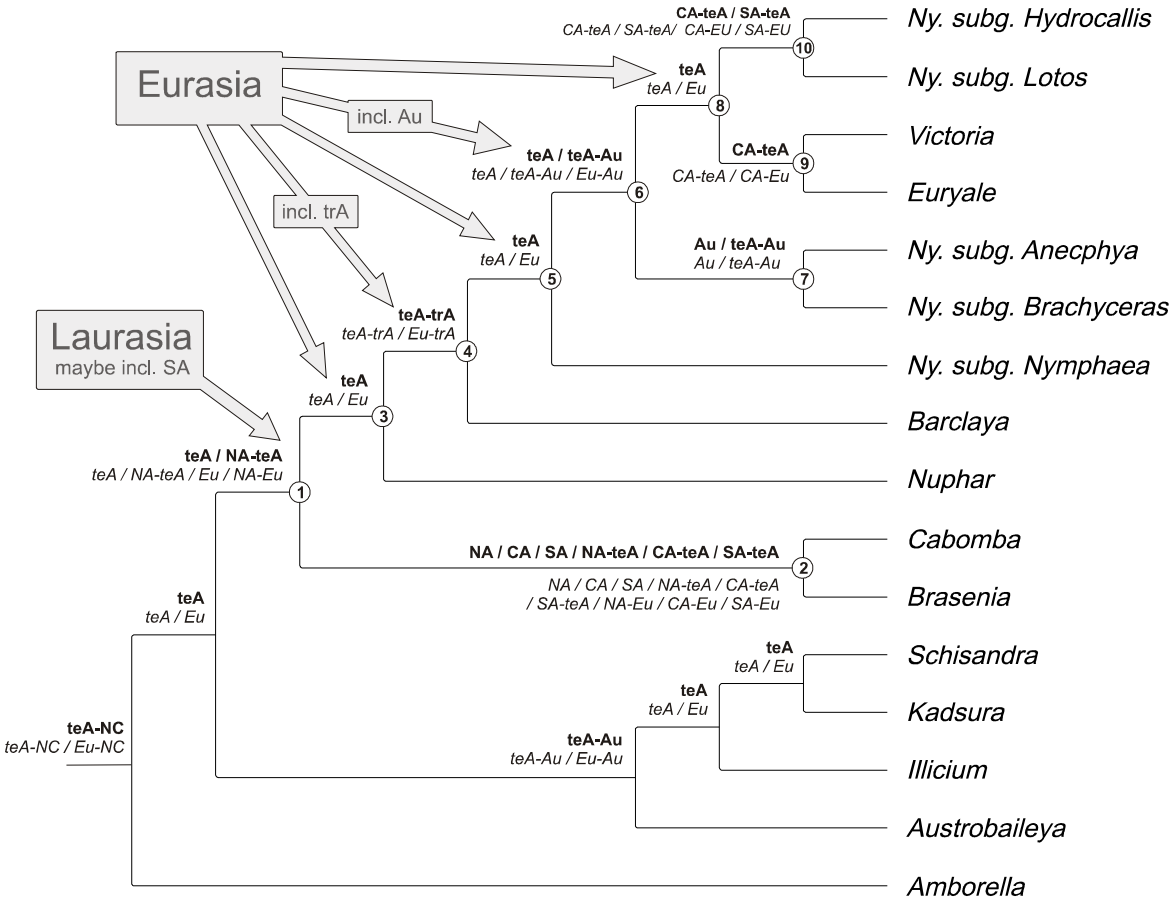


Figure 5.8 — Inferred ancestral areas in Nymphaeales. Presented are the results of dispersal–vicariance analyses with topology A and the maximum number of areas at ancestral nodes constrained to 2. Inferred ancestral areas for each node are given above the branches, with results from analysis regarding only the present distribution in bold face, and results from analysis regarding extant and fossil distribution in italics. Alternative solutions are separated by slashes, multiple–area reconstructions are indicated by hyphens (e.g., NA–teA). NA = North America, CA = Central America, SA = South America, Eu = Europe, Af = Africa, teA = temperate Asia, trA = tropical Asia, Au = Australia, NC = New Caledonia.

as another possible area if fossils are regarded. However, all the inferred combinations include areas that were clearly disjunct by the inferred time of the splitting between *Victoria* and *Euryale*. Therefore a more complex explanation for the present distribution, involving either extinction in other adjacent areas or long–distance dispersal, has to be found. The same applies to the ancestor of subgenera *Hydrocallis* and *Lotos*, for which all possible distributions include clearly disjunct areas (e.g., South America and Africa, or South America and temperate Asia, see node 10 in Tables 5.1 and 5.2).

The broader taxon sampling in topology B (Table 5.2) allows inferring ancestral areas for *Nuphar* (North America and temperate Asia) and for each subgeneric lineage in *Nymphaea* (see nodes 12–17 in Table 5.2). A major difference between topology A and B lies in the relationship of subg. *Brachyceras* and subg. *Anecphya* to each other and, accordingly, in the inferred distribution of their common ancestor (node 7 in both trees). Whereas Australia (H) or temperate Asia + Australia (FH) are the ancestral areas in topology A (Table 1, Figure 8), ancestral distribution for node 7 in tree B (Table 5.2) involves North America in all combinations. However, in both trees the ancestor of *Brachyceras* and *Anecphya* is inferred as rather widespread if no constraints on the number of areas are invoked.

5.4 Discussion

5.4.1 Extant patterns of biodiversity in Nymphaeales

The maps presented in Figures 3–5 illustrate the differentiated patterns of distribution in Nymphaeales. As indicated before, it is obvious that most of the genera, subgenera and species are restricted to continents and climatic zones, and that only the combination of those restricted ranges makes up the almost worldwide distribution of the Nymphaeales. Even taxa that seem to be more widespread at first sight, like the temperate clades *Nuphar* and *Nymphaea* subg. *Nymphaea*, or the pantropical water lily subgenus *Brachyceras*, reveal distinct patterns at the species level. In *Nuphar*, for instance, there two distinct clades that can be recognized by molecular and morphological data (Padgett *et al.* 1999), a lineage comprising only North American species and a second lineage of Eurasian species. The only exception is *N. microphylla*, which is consistently depicted within the Eurasian clade but occurs only in eastern Canada and the north eastern United States.

A similar pattern can be observed in the temperate water lily clade *Nymphaea* subg. *Nymphaea*. According to (Borsch *et al.* subm.) the Eurasian species *N. alba*, *N. candida*, *N. tetragona* (N. pygmaea was not included in this study) form a well–supported clade, as well as the North American taxa *N. odorata* and *N. leibergii* (Borsch, Wiersema, Hellquist, unpublished data). *Nymphaea mexicana* seems to represent a separate – possibly basal – lineage in the temperate clade (Borsch *et al.* subm.), although there is evidence for hybridisation and introgression occurring between *N. mexicana* and *N. odorata* that may have led to the evolution of *N. odorata* subsp. *tuberosa* (Woods *et al.* 2005a, Woods *et al.* 2005b). In *Nymphaea* subg. *Brachyceras* the New World species again form a well–supported clade (0.98 Bayesian posterior probability from the *trnT–trnF* data set, Borsch *et al.* subm.) that is sister to at least two distinct clades of old world species of this subgenus.

The present study indicates that most taxa in Nymphaeales are characterized by rather small, continuous ranges, which is not conform to the general widespread patterns assumed for "typical" aquatic plants (Sculthorpe 1967). The only widespread species in Nymphaeales are *Nymphaea tetragona*, with a mainly Eurasian distribution but also occurring in Alaska and western Canada, as well as *Brasenia schreberi*, which sporadically occurs in Africa, Asia, Australia and the American continents. These exceptional phenomena need to be regarded when delineating possible historical biogeographic scenarios (see below).

Figure 6 reveals that the highest concentrations of different species of Nymphaeales can be found in northern South America and the Caribbean, the Zambezian region of Africa and northern Australia. These centres of diversity correlate with the main ranges of species in *Cabomba* and *Nymphaea* subg. *Hydrocallis* in South America, *N.* subg. *Brachyceras* in the Zambesian region of Africa, and *N.* subg. *Anecphya* in northern Australia. Species diversity in Nymphaeales is higher in the tropical than temperate regions, which reflects a general latitudinal gradient observable in plants and other organisms (e.g., Rohde 1992, Gaston 2000, Mutke *et al.* 2005). However, the pattern of diversity in Nymphaeales is not fully congruent with the major patterns in angiosperms. Barthlott *et al.* (1996a, 2005) found the highest diversity of vascular plants in regions of high geodiversity (Barthlott *et al.* 1996a), especially in mountain areas of the humid tropics and subtropics, but water lilies are generally missing from mountainous regions. The reasons for this discrepancy are obvious: mountains generally do not provide suitable habitats for aquatics like the water lilies. All members of Nymphaeales need rather large, still or slowly moving water bodies (Schneider & Williamson 1993, Williamson & Schneider 1993). Larger lakes exist, e.g., on the Andean altiplano but these areas are usually characterized by strong winds, permanent waves and high salt concentrations – all rather inappropriate conditions for water lilies.

5.4.2 Divergence times in Nymphaeales and Nymphaeaceae

Based on non–parametric rate smoothing the age of the Nymphaeales crown group was estimated to 63.2–64.2 Ma in the present study. Obviously, a first diversification into three major lineages, i.e. Cabombaceae, *Nuphar* and the remaining Nymphaeaceae, took place in a rather short period of time in the Palaeocene (between 64.3 and 59.5 Ma). A rapid early diversification has been assumed earlier by Löhne *et al.* (subm., Chapter 3) based on observed short branches at the base of the Nymphaeales phylogeny. A second major radiation phase can be observed in the Early Oligocene (33.4–30.2 Ma), when the extant lineages within the genus *Nymphaea*, including a *Victoria–Euryale* clade, differentiated from each other. In the previous analysis of Yoo et al. (2005) this clear identification of the two radiation phases was not possible because of limited taxon sampling. The extended taxon sampling in the present study also allows for the first time the estimation of crown group ages within Nymphaeales, such as *Nuphar* or the subgenera of *Nymphaea*.

These estimates of divergence times for Nymphaeales are solely based on a single calibration point, the proposed age of angiosperms of 131.8 Ma. Interpolations from a single calibration point should be taken with care because the results might be biased and misleading (Sanderson *et al.* 2004, Bell *et al.* 2005, Magallón & Sanderson 2005). However, in the case of the present data set, the estimated ages for nodes within Nymphaeales are fully consistent

with evidence from the fossil record, i.e. all lineages are inferred to be older than any fossils that can unambiguously be assigned to extant genera. Among those fossils are *Brasenia spinosa* from the Upper Eocene of England (Collinson 1980), setting the minimum age of the Cabombaceae crown group (node 2) to 37–33.7 Ma (Berggren *et al.* 1995), and *Nuphar wutuensis* from the Early Eocene of China (Chen *et al.* 2004), which sets the minimum age of the Nymphaeaceae crown group (node 3) to 52 Ma. Furthermore, the age of *Nymphaea liminis* (Upper Eocene to Lower Oligocene, Collinson 1980), a fossil that can be placed on the branch leading to node 6[3], is consistent with the age inferred from the molecular data (33.1–33.4, Lower Oligocene).

Despite the striking congruence of present molecular and fossil data, one has to keep in mind that information from the fossils can only provide minimum ages for lineages. Since the fossil record can not be regarded as complete (Springer 1995) it is always possible that older fossils of a given lineage will be discovered. However, care has to be taken when interpreting characters of new fossils and postulating close affinities to extant taxa. Taxonomic misidentification of fossils and their erroneous placement onto the tree can lead to wrong conclusions on the age of lineages (Benton & Ayala 2003). The study of Yoo *et al.* (2005) revealed a remarkable inconsistency between the actual age of two recently described Cretaceous fossils assigned to Nymphaeaceae (Friis *et al.* 2001, Gandolfo *et al.* 2004) and the molecular–based age estimates of Nymphaeales. In fact, the assignment of the "Nymphaeaceae" fossil described by Friis et al. (2001) is based on very few characters that are not unique to Nymphaeales but also allow placing the fossil near Illiciaceae or other families (Gandolfo *et al.* 2004). The postulated close affinity of the Turonian fossil *Microvictoria svitkoana* to extant *Victoria* (Gandolfo *et al.* 2004) might be biased by the fact that no taxa from outside the Nymphaeales were regarded in the cladistic analysis. Yoo *et al.* (2005) suggest a placement of *Microvictoria* on the stem lineage of Nymphaeales. Further research on morphological character evolution interpreting these fossils will be needed and may lead to the conclusion that both fossil flowers are in fact outside of Nymphaeales.

Besides the uncertainties in the fossil record there are other sources of error, i.e. special characteristics of the molecular data or the algorithms used for estimating divergence times. Yoo et al. (2005) report considerable differences in ages obtained from NRPS of trees with either ML or MP branch lengths; and some authors stress the weaknesses of the NRPS approach and its tendency to overestimate ages (Sanderson 2002, Magallón & Sanderson 2005, Renner 2005). However, differences in ages estimated from MP and ML branch lengths are minor in the present data set (Figure 5.6). Preliminary results obtained from Penalized Likelihood and a Bayesian relaxed clock approach (M.–J. Yoo, unpublished data) seem to confirm the findings from NRPS as discussed here.

Linder *et al.* (2005) report strong effects of taxon sampling on estimated divergence times. Based on their data set of Restionaceae, Linder *et al.* (2005) conclude that an increase in the number of taxa sampled generally leads to older estimates of divergence times. Thus, undersampling can cause a remarkable underestimation of divergence times. A comparison of the divergence times inferred by Yoo *et al.* (2005) from a data set comprising 8 taxa of

[3] Seeds with longitudinal rows of hairs indicate a close to *Nymphaea* subg. *Brachyceras* or subg. *Lotos* or ancestor thereof (Borsch 2000). Such rows are not present in *Nymphaea* subg. *Nymphaea*.

Nymphaeales with the results from the present analysis indicates a similar taxon sampling effect. Ages for nodes in Nymphaeales obtained from the present *matK* data set (including 30 ingroup taxa) are on average 8–24 Ma older than those from the *matK* data set used in the previous study. Thus, future studies in Nymphaeales, including an even more comprehensive taxon sampling than the present study, might further increase the inferred age of the Nymphaeales crown group and its major lineages.

5.4.3 Ancestral areas and implications for possible dispersal scenarios

The results of the dispersal–vicariance analyses are summarized in Tables 5.1 and 5.2 and in Figure 5.8. For most nodes the optimization process leads to multiple solutions, providing a variety of equally possible ancestral distributions. The inferred results for node 1 (common ancestor of extant Nymphaeales) exemplify the major disadvantages of dispersal–vicariance analysis: for the basal nodes DIVA generally infers a widespread distribution if the number of area units is not constrained. The history of areas, i.e. the rise and demise of migration barriers through time, can not be regarded in the algorithm. Therefore, a considerate interpretation of the results seems necessary. Indeed, the picture becomes much clearer if the history of the areas and the special ecological characteristics of the plant group are regarded for interpretation.

By the time of divergence of the major Nymphaeales lineages (Cabombaceae, *Nuphar* and the remaining Nymphaeaceae; nodes 1 and 3), i.e. in the Palaeocene, the northern areas were still connected and formed the super–continent Laurasia. Plant migration was not yet hindered by barriers such as the Turgai strait or the North Atlantic. Therefore, it does not seem reasonable to narrow down the ancestral distribution to a subunit of the northern super–continent. Instead, whole Laurasia must be assumed as the area of origin of extant Nymphaeales. However, a land connection from Laurasia to South America via the Caribbean arc still existed in the Palaeocene. South America should therefore be included in the proposed ancestral range of Nymphaeales. In fact, some area combinations inferred with DIVA include South America. Other Gondwanan continents that are nowadays inhabited by water lilies and relatives, i.e. Africa and Australia, were clearly separated from Laurasia during the Palaeocene and can thus be ruled out as possible areas of origin for Nymphaeales.

In the course of the Eocene the northern hemispheric areas subsequently became isolated due to the establishment of physical and climatic barriers (Chapter 5.1.3, page 114 f.). The Eocene / Oligocene boundary is the inferred time frame for the diversification of core Nymphaeaceae (nodes 5, 6 and 8 in Figure 5.1 and 5.2). Temperate Asia is consistently depicted as a possible area for the common ancestor of core Nymphaeaceae (node 5). It is the only solution for topology B if fossils are included and it is an alternative to Europe in topology A. However, since the Turgai strait, which separated Europe from Asia during the Eocene, was already retreating by that time, an ancestral distribution involving Asia and Europe seems possible. Physical connections between Asia and North America and between North America and Europe might still have existed, and therefore the ancestor of core Nymphaeaceae might have been more widespread. Ancestral ranges involving either Australia (topology A, node 6) or South America (topology B, node 5 and 6) can be regarded

as unlikely for that period of time since both continents were clearly isolated at the Eocene / Oligocene boundary.

A comparison of Table 5.1 and 5.2 reveals that the different topologies tested in this study (with the main difference being the position of *Victoria–Euryale*) have no significant effects on the inferred ancestral areas for the basal nodes in the Nymphaeales tree (nodes 1–5). However, the increased taxon sampling used for topology B allows conclusions on the ancestral distribution of the genus *Nuphar* and the subgenera of *Nymphaea* – information that can not be deduced from topology A, since these taxa were not represented by individual species. Furthermore, the distribution inferred for the common ancestor of subgg. Brachyceras and Anecphya differs between topology A and B. This difference is due to the basal position of the American lineage of subg. *Brachyceras* in topology B, which introduces North or Central America as potential ancestral areas. However, this apparent conflict can be disregarded because it refers to the weakest node in topology B (depicted only in the Bayesian analysis of *trnT–trnF*, with low support, Borsch *et al.* subm.). Based on the results of the present study and the just mentioned considerations of their significance it is now possible to design a plausible scenario for the spatial evolution of Nymphaeales during the Tertiary.

5.4.4 Historical biogeography of Nymphaeales — a synthesis

Diversification of the Nymphaeales crown group in the Early Tertiary

By the time the Nymphaeales crown group diversified, Gondwana had already broken up and the southern continents were quite remote from each other and from Laurasia. Based on the present data, a northern hemispheric origin of the Nymphaeales crown group seems most likely. However, an early range expansion to South America is possible since there was a continuous land connection during this period (Figure 5.9). Fossil seeds of *Nuphar* from the Late Palaeocene of North America and Early Eocene of China indicate that this lineage of Nymphaeales has already diverged in the Palaeocene and was widely distributed in the northern hemisphere, which was characterized by a stable warm–temperate to tropical climate by that time. Early Nymphaeales presumably were an element of the so–called boreotropical flora that covered large parts of the northern hemisphere during Palaeocene and Eocene.

Evidence for the early history of the Cabombaceae–lineage is, however, scarce. Dispersal–vicariance analyses failed to provide an explicit solution, which is mainly caused by the present widespread distribution of *Brasenia*. However, the New World seems to be favoured in the reconstructions if the number of ancestral areas is constrained to two or five. According to the molecular data the split between the extant genera *Cabomba* and *Brasenia* was dated to Middle Eocene (~40 Ma). Since South America was isolated by that time, it can be assumed that *Cabomba* invaded this continent after the re–connection to North America during the Miocene. The extant monotypic genus *Brasenia* obviously represents a relict regarding both, distribution and phylogenetic diversity. A variety of fossil species have been described from the Eocene to Pleistocene of Eurasia, and in fact, *Brasenia* used to be a common element of the European freshwater vegetation until its final extinction during the

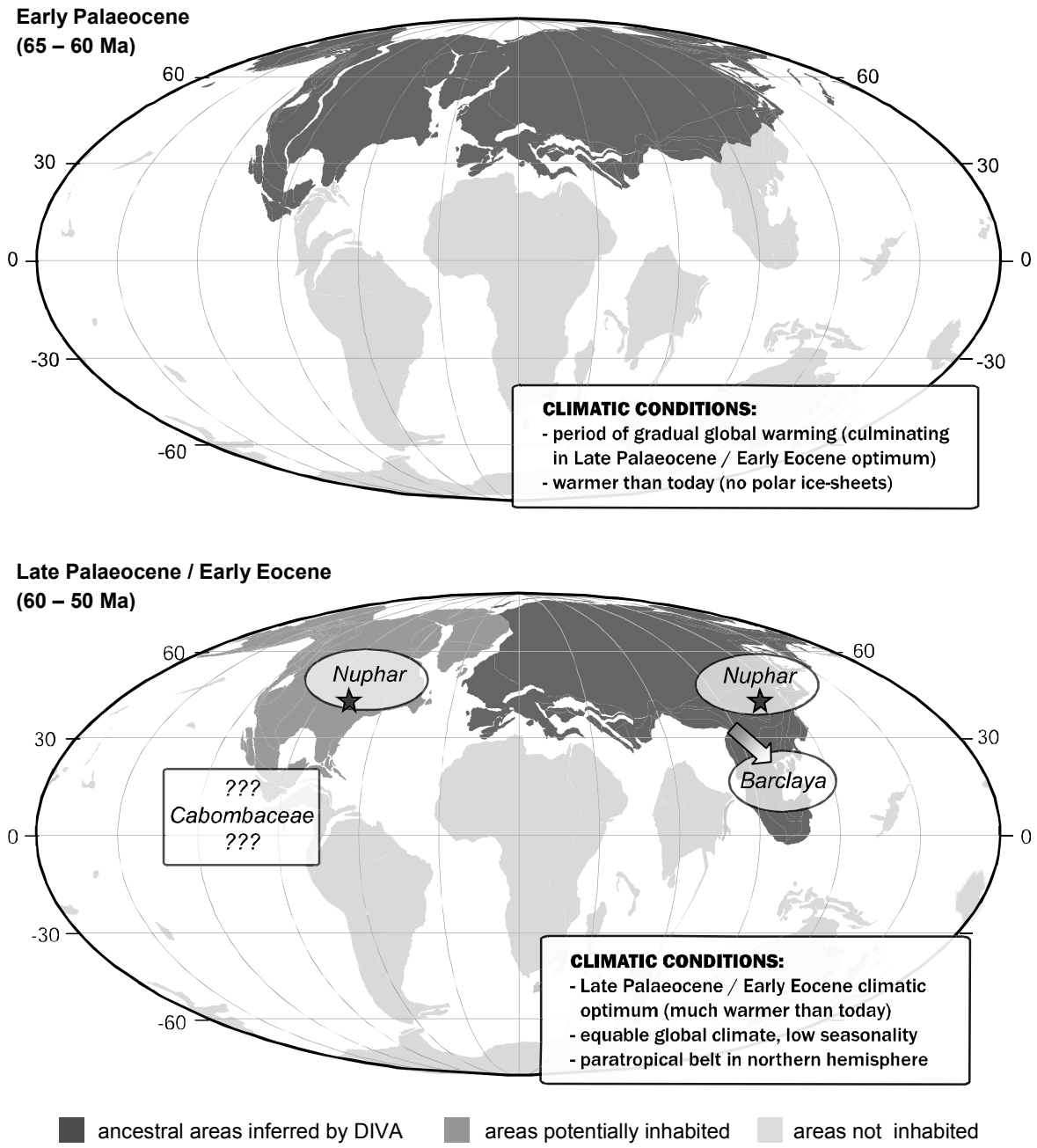


Figure 5.9 — Historical Biogeography of Nymphaeales in the Early Tertiary. The upper map illustrates the areas presumably inhabited by the common ancestor of the extant Nymphaeales (shaded in dark grey), as inferred from dispersal–vicariance analysis. SE Asia was directly connected to Laurasia and therefore and might have been inhabited as well, although this has not been inferred with DIVA (therefore marked as "potentially inhabited", shaded in lighter grey). The lower map shows the situation in Late Palaeocene / Early Eocene when some extant lineages of extant Nymphaeales already had diverged and were diversifying in the northern hemisphere. The ancestral area of Cabombaceae remains elusive, but a New World origin seems possible. Maps showing the position of continental plates at 65 Ma (top) and 50 Ma (bottom) were generated using the online "Plate Tectonic Reconstruction Service" of the Ocean Drilling Stratigraphic Network (see Materials and Methods). Note that maps show delimitations of continental plates and not the actual shore lines. Information on climatic conditions was obtained from Zachos *et al.* (2001).

last glacial maximum (Dorofeev 1984). However, a closer examination of genetic variability between the extant disjunct populations and areas would be worthwhile in order to decide if this pattern really reflects relict areas of a previously more widespread taxon or if it rather is the outcome of more recent long–distant dispersal. The differentiation of fossil species is almost exclusively based on seed characters, but so far no comparative analyses of specimens from disjunct extant location of *Brasenia schreberi* have been conducted. A detailed examination might reveal differences that are not obvious at first sight. There are several examples from other aquatic organisms — generally showing a reduction in morphological traits and high intraspecific variation — that were presumed to be widespread species but turned out to be composed of groups of sibling species, e.g. *Zannichellia* (Van Vierssen 1982), *Lemna* (Crawford *et al*. 1996) or *Wolffiella* (Crawford *et al*. 1997, see Santamaría 2002 for review).

Another lineage in extant Nymphaeales that branched relatively early is *Barclaya*, which at present is occurring with four species in tropical South East Asia. Since there are no fossils of *Barclaya* any conclusions on its history will be tentative, but in view of the present it seems plausible that the ancestor of extant *Barclaya* became separated from the Laurasian stock of Nymphaeales during the Early Eocene and evolved independently in South East Asia. This region was characterized by relatively stable wet–tropical climates throughout the whole Tertiary but was isolated from the northern boreotropical flora by an arid belt expanding from China to Middle Asia (Tiffney & Manchester 2001). A subsequent eastward range expansion of *Barclaya* obviously took place during the Miocene when the Australian plate collided with SE Asia and when New Guinea has been elevated (Hall 1998).

Radiation of core Nymphaeaceae during the Oligocene

The Early Oligocene was a period of major changes in the global climate and ocean circulation, which were reflected in significant turnovers in marine and terrestrial biota (Collinson 1992, Wolfe *et al*. 1992, Zachos *et al*. 2001). The Oligocene was initiated with a glaciation of Antarctica and, in general, it was considerably cooler and dryer than previous periods of earth history. The Early Oligocene climatic deterioration set the seal on the demise of the boreotropical flora, which had been declining since the middle Eocene (e.g., Wolfe 1992). The sea level was, at least in some short stages, 100–160 m lower than at present (Hallam 1992), which might have allowed migration of biota via new land bridges.

The initial radiation of core Nymphaeaceae temporally coincides with the climatic changes at the Eocene–Oligocene boundary. The break–up of the North Atlantic Land Bridge and the expansion of drier and cooler zones led to the fragmentation of Laurasia and the fragmentation of ranges of plants adopted to constantly warm and humid climates. In view of the tremendous impact of these climatic changes on the world's biota — especially in the northern hemisphere — a correlation with the evolution of core Nymphaeaceae seems plausible. Especially, the development and radiation of a temperate clade of *Nymphaea* most likely was associated with the expansion of the cold–temperate zone. A piece of evidence is the fact that the species within *Nymphaea* subg. *Nymphaea* clearly group into the Eurasian and the North American species (Borsch, Wiersema, Hellquist, unpublished data), which indicates that this clade originated in advance of the barrier formation and that the descendants diversified separately on both sides of the barrier (North Atlantic). The estimated

minimum age of segregation of the North American from the Eurasian clade (~25 Ma) seems to support this hypothesis. However, further studies including more taxa and more molecular data are necessary to confirm this age and this biogeographical scenario. In the case of *Nymphaea tetragona*, a rather recent range expansion over the Bering Land Bridge (or across the Bering Strait) seems to be the most plausible explanation for the occurrence of this "Old World"–species in Alaska and Northern Canada. The Bering area was an important refugium and migration pathway for several cold–adapted plant taxa during the Quarternary (reviewed in Abbott & Brochmann 2003).

The biogeographic history of the tropical subgenera of *Nymphaea* and their ancestors, however, seems more obscure. In view of the temperate subgenus *Nymphaea* as the basal lineage in the genus, one could assume that the water lilies originally evolved as a temperate lineage. However, there is weak indication that the neotropical *Nymphaea mexicana* may be basal in subg. *Nymphaea* (Borsch *et al.* subm., Borsch, Wiersema, Hellquist, unpublished data) suggesting that *Nymphaea* subg. *Nymphaea* might represent a lineage that separated early from tropical or subtropical stock and diversified in the northern temperate zone during the cooling periods of the Tertiary. Cold resistance obviously has been developed independently several times during the evolution of Nymphaeales (present in *Brasenia* and *Cabomba*, *Nuphar*, *Nymphaea* subg. *Nymphaea*). Nevertheless, the capacity for surviving in seasonal habitats seems to be a pre–adaptation in the whole clade. Some exclusively tropical taxa are rather short–lived perennials with a strong tendency towards annual lifeforms (*Victoria*, *Euryale*), others can endure drier periods with low water level or are even capable of growing in ephemeral water bodies due to the persisting rhizomes or tubers (especially in *Nymphaea* subg. *Anecphya* and *Ondinea*, Jacobs & Porter in press).

Climatic change is not the only factor that influenced and triggered the radiation of core Nymphaeaceae. Additionally, differential ecological adaptation — especially with regard to pollination — played an important role in the evolution of water lilies. A whole clade is characterized by adaptations to nocturnal rhinoceros beetles (Cyclocephalini) as pollinators, including night–flowering behaviour, increased temperature inside the flowers and the emission of strong scents (Prance & Arias 1975, Wiersema 1987, Hirthe & Porembski 2003). This clade comprises the neotropical subgenus *Hydrocallis*, *Victoria*, and the palaeotropical *Nymphaea lotus*. The existence of same pollination syndromes and same pollinators[4] in American and African water lilies led Ervik & Knudsen (2003) to the assumption that this association predates the opening of the Atlantic Ocean and that the extant distribution patterns are the result of continental drift. However, Gondwanan vicariance seems unlikely given the divergence times of Nymphaeales inferred in this study. A scenario involving long–distance dispersal across the Atlantic can also be rejected, because this would require a parallel dispersal of both the plant and its pollinator. Instead, the most probable explanation is the co–evolution of the common ancestor of Cyclocephalini beetles and the common ancestor of night–flowering water lilies in the Northern Hemisphere and a parallel separation of lineages due to the segregation of North America and Eurasia. Due to global cooling the American and the Eurasian lineages were dislocated further south. Whereas *Hydrocallis* and

[4] *Ruteloryctes morio*, the pollinator of *Nymphaea lotus*, is the only palaeotropical member of the otherwise neotropical tribe Cyclocephalini of the Dynastinae beetle family (Enrödi 1985). *Cyclocephala* beetles pollinate species of *Nymphaea* subg. *Hydrocallis* and *Victoria* (Cramer *et al.* 1975, Prance & Arias 1975).

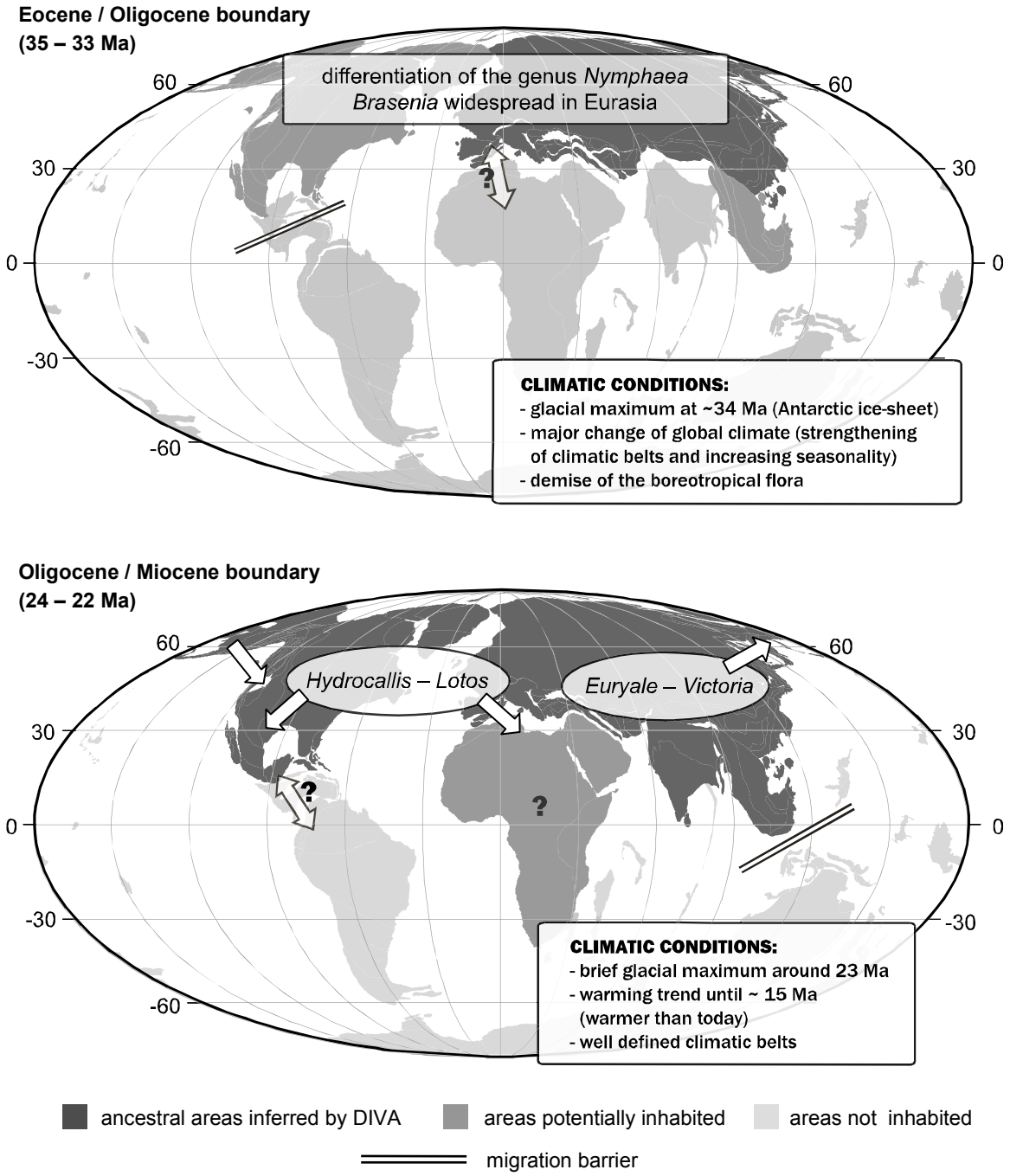


Figure 5.10 — Historical biogeography of Nymphaeales in the Mid–Tertiary. The upper map illustrates the biogeographic scenario for the Eocene/Oligocene boundary, when the radiation of major lineages of *Nymphaea* and *Victoria–Euryale* took place. There was no land connection to South America by that time. There is evidence from plant other plant groups for dispersal from Africa to Europe or vice versa across the Tethys Sea, but it is not possible to decide whether members of Nymphaeales invaded Africa by that time. The lower map shows the situation around 23 Ma b.p. when the diversification of the subgenera of *Nymphaea* proceeded and, e.g., *Victoria* and *Euryale* diverged from a common ancestor. *Victoria* or its ancestor might have migrated to America via the Bering Land Bridge, or optionally, by sweepstake dispersal over the North Atlantic. Dispersal to South America might have been possible through island hopping but is regarded as unlikely. Ancestors of extant members of *Nymphaea* subg. *Brachyceras* or subg. *Lotos* possibly inhabited Africa. Maps are showing the position of continental plates at 35 Ma (top) and 23 Ma (bottom). See previous figure for source and references.

the American Cyclocephalini underwent a parallel radiation in the neotropics, the African beetle–plant association remained as a relict. This scenario would be in congruence with the fossil record of both, Nymphaeales and the beetles. The are no fossils of Cyclocephalini (Ervik & Knudsen 2003), but the oldest fossil of the beetle subfamiliy Dynastinae, to which Cyclocephalini belong, has recently been found in the Middle Eocene of Oregon / USA (Ratcliffe *et al*. 2005).

The separation of the neotropical and palaeotropical sister–lineages (*Hydrocallis–Lotos*, *Victoria–Euryale*) was dated to Early Miocene (23–22 Ma) according to the present data set. By that time, plant migrations across the North Atlantic were still possible (Axelrod 1975, Tiffney & Manchester 2001), albeit probably limited due to enlarging water barriers. A recently published analysis of Miocene floras from Iceland indicates persisting floral exchange to both, North America and Eurasia until Late Miocene (Denk *et al*. 2005). Especially during the warm peak in Middle Miocene, Iceland was re–invaded by thermophilic Laurasian plants such as Magnoliaceae, and it seems likely that this migration pathway was open for water lilies as well. There is also strong evidence that the Miocene was warm enough to allow thermophilic vegetation as far north as eastern Siberia and Alaska (Dacrydium and Podocarpus pollen in late Miocene of Alaska, Reinink-Smith & Leopold 2005). Therefore, migration of thermophilic plants like water lilies might have been possible also via Beringia during the Miocene. Furthermore, it seems reasonable to assume a delayed response to climatic deterioration by water lilies and other aquatics because of the wider ecological tolerance of these plants (see previous page) as well as the relatively greater uniformity and long–term stability of aquatic habitats (Schenck 1886, Les 1986).

As mentioned earlier (Chapter 5.1.3, page 114), the most important restraint on plant growth in northern latitudes during warmer periods was the availability of light. However, given the general adaptation of Nymphaeales taxa to seasonal habitats, dark winters probably were not limiting their growth and range expansion in northern latitudes. Based on the present data, it does not seem possible to decide whether the North Atlantic Route or the Beringia route were more important for the spatial diversification of Nymphaeales. In view of Miocene fossils of *Nymphaea* subg. *Lotos* present in western Eurasia (Europe) but absent from eastern Eurasia it might be tempting to assume a North Atlantic disjunction. However, the absence of fossils can not be taken as evidence for the absence of certain taxa in that area (e.g., Heads 2005). Therefore, the absence of fossils of *Nymphaea* or the *Victoria–Euryale* clade from North America does not interfere with the hypothesis of a boreotropical origin and subsequent migration to the neotropics. In fact, Tiffney (1985) notes that the Tertiary fossil record of North America, especially Eastern North America, is extraordinarily poor, which can be attributed to tectonic settings unfavourable for the formation of deposits, and intensive erosion of the few fossil deposits during the Pleistocene by ice sheets in the north or extensive weathering in the south. In general, the present pattern of tropical disjunctions of closely related lineages in Nymphaeales can be explained as a result of range expansion in the northern hemisphere during the Tertiary (Eocene to Oligocene) and subsequent vicariance due to formation of climatic and oceanic barriers during Oligocene and Miocene. Climatic changes led to wide–ranging extinctions in the Northern hemisphere and dislocation of thermophilic Nymphaeales–lineages to Central and South America, Africa and tropical Asia. The importance of the northern hemisphere for the spatial evolution of extant

pantropical plant lineages has been shown for several other angiosperm clades, such as Lauraceae (Chanderbali *et al*. 2001), Magnoliales (Azuma *et al*. 2001, Doyle *et al*. 2004, Richardson *et al*. 2004), or Malpighiaceae (Davis *et al*. 2002b, Davis *et al*. 2004). Similarly to the results of the present study, divergence of thermophilic (tropical or subtropical) sister–lineages in the Old World and in the New World has been dated to precede the final closure of the North Atlantic pathway, e.g. in *Ocotea* (Lauraceae, Chanderbali *et al*. 2001) or Arbutoideae (Ericaceae, Hileman *et al*. 2001), providing evidence for vicariance instead of transoceanic long–distance dispersal.

Although transoceanic long–distance dispersal has been shown to be an important factor in historical biogeography of several plant groups, e.g. some lineages in Melastomataceae, Bromeliaceae or Legumes (Givnish *et al*. 2004, Lavin *et al*. 2004, Renner 2004b), it is regarded as rather unlikely and less important for the extant biogeographic patterns of Nymphaeales. An important additional reason for this conclusion lies in the biological and ecological characteristics of the water lilies and relatives. Propagation through seeds is the main way of reproduction and dispersal in Nymphaeales (in the case of *Cabomba* the dislocation of stem segments could also be important, Orgaard 1991). Usually, fruits of Nymphaeaceae and Cabombaceae ripe under water, and seeds are subsequently released by decomposition of the fruit. Thus, passive drifting within the water body or dislocation to other lakes and rivers by water birds is regularly assumed to play the major role in dispersal (Sculthorpe 1967, Sanders 1979, Orgaard 1991, Les *et al*. 2003). Since seeds of Nymphaeales are generally not desiccation tolerant (Tarver & Sanders 1977, Smits *et al*. 1989), endozoochorous transport seems to be the only way for long–distance dispersal in Nymphaeales. Indeed, water lily seeds are an important food source for ducks and other diving waterfowl (Tréca 1981, Hoppe *et al*. 1986, Green *et al*. 2002). However, Smits *et al*. (1989) could show that seeds of *Nymphaea* and *Nuphar* become completely digested and are not capable of germination after gut passage. Only transport via the crop of waterfowl (followed by regurgitation, Proctor 1968, Charalambidou & Santamaría 2002) seems possible, but this would be confined to rather short distances. Remarkably, water lilies are generally absent from remote oceanic islands. Furthermore, bird migration routes generally follow North–South directions (Berthold 1993); East–West migrations across continents have only been observed in the Artic region. These patterns rather support the scenario of a Laurasian origin and northern hemispheric migration of nymphaealean ancestors and subsequent allocation to the south.

Evolution of water lilies in the Miocene

The most important tectonic events in the Miocene were the re–connection of North and South America (via the western island arc of the Caribbean plate), the collision of Australia with the SE Asian plate and subsequent uplift of New Guinea, as well as the final closure of the Tethyan seaway. The Miocene was characterized by a rather warm climate but also by increasing aridity. Two major radiations of extant lineages of *Nymphaea* have been dated to the Middle Miocene: the diversification of subgenus *Hydrocallis* in South America and the diversification of subgenus *Anecphya* in Australasia. The radiation of the neotropical subgenus *Hydrocallis* clearly predates the final closure of the Isthmus of Panama in the Pliocene, but southward dispersal via the island arc, as it has been observed in several other

groups of plants (Bell & Donoghue 2005) and animals (Bermingham & Martin 1998), seems likely.

The timing of the radiation of Australian water lilies is fully congruent with geological evidence. The uplift of New Guinea and surrounding islands in Middle Miocene (18–15 Ma) opened a migration pathway between Australia and the Indomalayan archipelago (Hall 1998). Several Australian lineages dispersed rapidly towards SE Asia, such as Myrtaceae or Podocarpaceae (Morley 1998). In addition, Morley (2000) mentions members of Fabaceae (*Acacia*) or Convolvulaceae (*Merremia*) as possible candidates for dispersal into the opposite direction. The review of Crisp *et al.* (2004) reveals that there are several other plant groups that have their first appearance in Australia in the Miocene and show remarkable radiations at this continent, e.g. some lineages of Asteraceae, Poaceae, and Chenopodiaceae (see also Kadereit *et al.* 2005). Generally, the radiation of these lineages is associated with significant vegetational changes in the central parts of Australia due to strong aridification from Middle Miocene on. The radiation of the Australian water lilies *Nymphaea* subg. *Anecphya*, which was inferred in the present study, provides evidence that also the flora of the monsoonal north of Australia was significantly influenced by invasions during the Miocene.

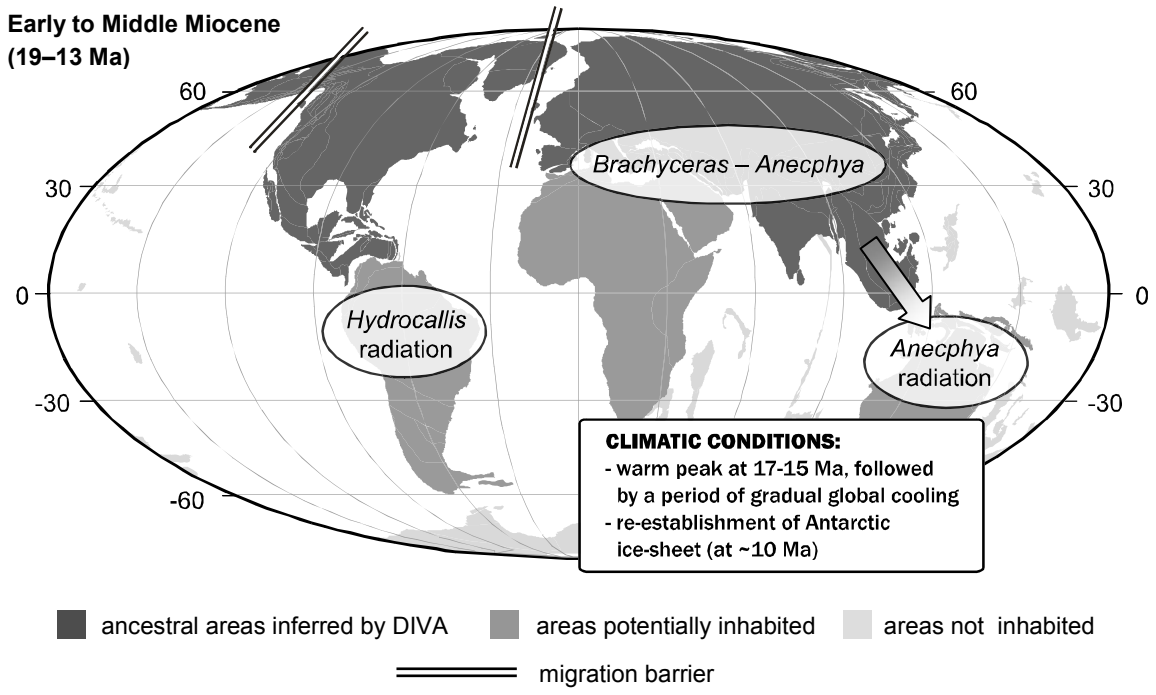


Figure 5.11 — Historical biogeography of Nymphaeales in the Late Tertiary. The map shows the position of plates at 14 Ma b.p. and the inferred distribution of Nymphaeales. the diversification of the water lily subgenera *Brachyceras* and *Anecphya* During Early Miocene (either as two clades separating from a common ancestor, or with the neotropical *Brachyceras*–lineage as the first branch, followed by the palaeotropical *Brachyceras*–lineages and *Anecphya*). The extant subgenus *Anecphya* started to diversify after the arrival of its ancestor in Australia in Middle Miocene. In parallel, the New World water lily subgenus *Hydrocallis* diversified in South America. Migration pathways in the northern hemisphere across the North Atlantic or the Bering Land Bridge were severed for not cold–adapted lineages of Nymphaeales. See caption of Figure 5.10 for source and references.

The picture of the biogeographic history of African water lilies, however, remains elusive. Based on the present data, a northern hemispheric origin of *Nymphaea* subgg. *Lotos* and *Brachyceras* and subsequent displacement to Africa during Oligocene or Miocene is assumed. However, this assumption can not be substantiated with evidence, because the reconstruction of ancestral areas is ambiguous for the respective nodes and there is no known fossil record for Nymphaeaceae in Africa. Morley (2000) and Jacobs & Herendeen (2004) state that, generally, Tertiary fossil data for Africa and the palaeotropics are extremely rare. Besides that, the phylogenetic relationships especially within the pantropical subgenus Brachyceras are not yet resolved. There is some weak indication that the neotropical lineage could be sister to a clade comprising the remaining, palaeotropical species of *Brachyceras* plus the Australian subgenus *Anecphya* (Borsch *et al*. subm., topology B in this study), which would have implications for the inferred ancestral distribution, but final conclusions will have to await further insights from phylogenetic studies. Similarly, the geographic history of the subgenus *Lotos* can also not be explained based on present distribution and species diversity alone. In the recent molecular analysis of Borsch *et al*. (subm.), the enigmatic *Nymphaea petersiana* from southern Africa is sister to a clade consisting of the African–European *N. lotus* and the Asian *N. pubescens*. Possible scenarios either include an African origin of the subgenus and subsequent range expansion to Europe and Asia, or a northern hemispheric origin followed by extinction and dislocation to southern Africa and Asia. The latter scenario is favoured here, considering the presence of Tertiary fossil seeds in Europe that can be assigned to this subgenus (Mai 1995).

5.5 Conclusion and prospect

The present study is the first that uses a comprehensive data set – including all genera of Nymphaeales and all subgenera of the genus *Nymphaea* – to track the evolution of this basal angiosperm order in time and space. The broad taxon sampling allows the identification of two major radiation phases in the evoltuion of the extant Nymphaeales: First, an initial diversification into three main lineages (Cabombaceae, *Nuphar*, and remaining Nymphaeaceae) during the Palaeocene and second, the radiation of core Nymphaeaceae (*Nymphaea*, *Ondinea*, *Victoria*, *Euryale*) in the Early Oligocene.

The diversification of the Nymphaeales crown group started in the northern hemisphere, at a time when the global climate was much warmer than today and when the northern super–continent Laurasia was widely covered with the so–called boreotropical flora. The radiation of core Nymphaeaceae in the Early Oligocene temporally coincides with strong global cooling, the demise of the boreotropical flora, and the separation of northern hemispheric continents. Therefore, it seems reasonable to assume climatic effects and the rise of migration barriers as the main causes for the rapid diversification of core Nymphaeaceae. However, this diversification was also accompanied by differential ecological adaptations, i.e. the development of different pollination strategies.

The disjunct distribution of extant sister–lineages in core Nymphaeaceae, e.g. *Euryale* – *Victoria* or *Nymphaea* subg. *Hydrocallis* – subg. *Lotos*, can be explained with a northern hemispheric origin and a subsequent spatial separation and southward dislocation of the

descendant lineages. Long–distance dispersal across wide oceanic barriers seems less likely, given the special biological and ecological characteristics of water lilies and relatives. Propagation through seeds is the main way of reproduction and dispersal in Nymphaeales, but seeds are not desiccation tolerant and are easily digested by waterfowl, which limits the probabilities for both, exo– and endozoochorous long–distance dispersal.

The extant patterns of distribution in Nymphaeales are not conform to the general pattern of extremely widespread aquatic plants such as *Ceratophyllum* — a fact, that can be attributed to the limited vagility of water lilies compared to other aquatics. A remarkable exception seems to be the monotypic genus *Brasenia*, which occurs sporadically on all continents except Europe and Antarctica. However, more data on the reproductive ecology, the dispersability of seeds, and the diversity among disjunct populations of *Brasenia* are needed for understanding the history of this genus. More information on the phylogenetic relationships within *Nymphaea* subg. *Brachyceras* are also needed to clarify the remaining questions on the history of Nymphaeales. An extended taxon sampling would help to elucidate the relation of subg. *Brachyceras* to subg. *Anecphya* (monophyletic vs. paraphyletic) and would provide additional information on the historical biogeography of this lineage.

The present study demonstrates the potential of combining well–supported phylogenetic hypotheses with evidence from extant distribution, the fossil record, and earth history for understanding the evolution of a lineage in space and time. A judicious taxon sampling, that reflects the actual diversity in a lineage, as well as a thorough consideration of biological and ecological factors, that influence the dispersability and distribution of the plants, are essential for the reconstruction of a plausible historical scenario.

Summary

Löhne, Cornelia (2006). ***Molecular phylogenetics and historical biogeography of basal angiosperms – a case study in Nymphaeales.*** Doctoral Thesis, Mathematisch-Naturwissenschaftliche Fakultät, Rheinische Friedrich–Wilhelms–Universität Bonn. 174 pp.

Scientific progress during the last two decades has greatly improved our knowledge on phylogenetic relationships among major lineages of flowering plants. Besides the two major groups of angiosperms, the eudicots and the monocots, there are several not closely related lineages that are generally referred to as the "basal angiosperms". Among those lineages, *Amborella*, Nymphaeales and Austrobaileyales are currently assumed to be successive sisters to the rest of angiosperms, thus forming a "basal grade". However, the phylogenetic relationships among the other angiosperm lineages are still not resolved, which is to a large extent due to the persisting need for efficient molecular markers at this taxonomic level.

In the present thesis, the molecular evolution of the *petD* intron was examined and its suitability as a new marker for resolving basal angiosperm relationships was proved. Furthermore, this study revealed the potential of microstructural changes as phylogenetic markers. *Amborella* was inferred to be sister to all other angiosperms in the *petD* analysis, but support at the backbone of the angiosperm was rather low. This is thought to be due to lower variability of group II introns compared to spacers and group I introns. However, the *petD* intron proved to be very useful for resolving major lineages of angiosperms and relationships within, such as the magnoliids, the eudicots or the Nymphaeales. Therefore, the *petD* intron was chosen as a marker, together with other fast evolving chloroplast regions (*rpl16* intron, *trnK* intron, *matK* gene, *trnT–trnL* spacer, *trnL* intron, *trnL–trnF* spacer) to elucidate details of phylogenetic relationships in Nymphaeales.

This first comprehensive analysis of Nymphaeales with regard to taxon sampling, i.e. including all genera but also all subgenera of *Nymphaea* as well as both sections of *Nuphar*, could confirm the monophyly of the Cabombaceae but does not provide convincing support for the monophyly of Nymphaeaceae with respect to *Nuphar*. Furthermore, the genus *Nymphaea* is inferred to be polyphyletic with respect to *Ondinea*, *Victoria* and *Euryale*. In fact, *Victoria* and *Euryale* are inferred to be closely related to a clade comprising all night blooming water lilies (*Nymphaea* subgenera *Hydrocallis* and *Lotos*). The Australian endemic *Ondinea* forms a highly supported clade with the Australian water lilies *Nymphaea* subg. *Anecphya*.

A detailed examination of relationships among Australian water lilies using chloroplast and nuclear markers (ITS, *trnT–trnF*) confirmed the close affinity of *Ondinea* to *N.* subg.

Anecphya, and within this subgenus especially to *Nymphaea hastifolia*. The ITS data set resolved two well supported clades in *Anecphya* which correspond to the small–seeded and large–seeded group defined by (Jacobs 1992, 1994). *Ondinea* and *N. hastifolia* are part of the small–seeded group. The striking incongruence of trees obtained from the chloroplast and from the nuclear marker with regard to relationships within the small–seeded group, points on recent hybridisation and introgression in this group.

The rather young radiation of water lilies in Australia gave rise to one of the centres of diversity in Nymphaeales. Other centres of diversity are northern South America and South–Central Africa, which correspond to likewise recent radiations in the water lily subgenera *Hydrocallis* and *Brachyceras*. The radiation of core Nymphaeaceae, i.e. *Nymphaea, Victoria, Euryale and Ondinea*, occurred in the Tertiary around the Eocene–Oligocene boundary and is correlated with strong global cooling, the demise of the boreotropical flora and the segregation of northern hemispheric continents. The northern hemisphere, and possibly also South America, is inferred to be the ancestral range of Nymphaeales. Other Gondwanan continents such as Africa or Australia have been invaded by water lilies rather recently.

Some findings of this thesis are of more general significance — beyond Nymphaeales or basal angiosperms: The present study provides another peace of evidence for the general utility of non–coding, fast–evolving chloroplast genomic regions and of microstructural changes as phylogenetic markers. Furthermore, the Nymphaeales study exemplifies the importance of judicious taxon sampling for correct phylogenetic inference. Only the combination of well–supported evidence from molecular phylogenetics, earth history and the fossil record with a thorough consideration of biological and ecological factors allows reasonable conclusions on the evolution of a lineage in space and time.

References

Abbott R. J. & C. Brochmann (2003). History and evolution of the arctic flora: in the footsteps of Eric Hultén. *Molecular Ecology* 12**:** 299-313.

Álvarez I. & J. F. Wendel (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29**:** 417-434.

Andreasen K. & B. G. Baldwin (2003). Nuclear ribosomal DNA sequence polymorphism and hybridization in checker mallows (*Sidalcea*, Malvaceae). *Molecular Phylogenetics and Evolution* 29**:** 563-581.

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

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

Anzotegui M. (2004). Megaflora of the chiquimil formation (Late Miocene) in the Santa Maria and Villavil valleys, Catamarca and Tucuman provinces, Argentina. *Ameghiniana* 41**:** 303-314.

Aoki S., K. Uehara, M. Imafuku, M. Hasebe, & M. Ito (2004). Phylogeny and divergence of basal angiosperms inferred from *APETALA3*- and *PISTILLATA*-like MADS-box genes. *Journal of Plant Research* 117**:** 229-244.

Arber E. A. N. & J. Parkin (1907). On the origin of angiosperms. *Journal of the Linnean Society* 38**:** 29-80.

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

Asmussen C. B. & M. W. Chase (2001). Coding and noncoding plastid DNA in palm systematics. *American Journal of Botany* 88**:** 1103-1117.

Austin J. D., S. C. Lougheed, P. E. Moler, & P. T. Boag (2003). Phylogenetics, zoogeography, and the role of dispersal and vicariance in the evolution of the *Rana catesbeiana* (Anura: Ranidae) species group. *Biological Journal of the Linnean Society* 80**:** 601-624.

Axelrod D. I. (1972). Edaphic aridity as a factor in angiosperm evolution. *The American Naturalist* 106**:** 311-320.

Axelrod D. I. (1975). Evolution and biogeography of Madrean-Tethyan sclerophyll vegetation. *Annals of the Missouri Botanical Garden* 62**:** 280-334.

Azuma H., J. G. García-Franco, V. Rico-Gray, & L. B. Thien (2001). Molecular phylogeny of the Magnoliaceae: The biogeography of tropical and temperate disjunctions. *American Journal of Botany* 88**:** 2275-2285.

Bailey C. D., T. G. Carr, S. A. Harris, & C. E. Hughes (2003). Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics & Evolution* 29**:** 435-455.

Bakker F., A. Culham, R. Gomez-Martinez, J. Carvalho, J. Compton, R. Dawtrey, & M. Gibby (2000). Patterns of nucleotide substitution in angiosperm cpDNA *trnL* (UAA)-*trnF* (GAA) regions. *Molecular Biology and Evolution* 17**:** 1146-1155.

Baldwin B. G., M. J. Sanderson, J. M. Porter, M. F. Wojciechowski, C. S. Campbell, & M. J. Donoghue (1995). The ITS region of nuclear ribosomal DNA: a valuable source of

evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82**:** 247-277.

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

Barthlott W., W. Lauer, & A. Placke (1996a). Global distribution of species diversity in vascular plants: towards a world map of phytodiversity. *Erdkunde* 50**:** 317-327.

Barthlott W., J. Mutke, D. Rafiqpoor, G. Kier, & H. Kreft (2005). Global centers of vascular plant diversity. *Nova Acta Leopoldina NF* 92**:** 61-83.

Barthlott W., C. Neinhuis, R. Jetter, T. Bourauel, & M. Riederer (1996b). Waterlily, poppy, or sycamore: on the systematic position of *Nelumbo*. *Flora* 191**:** 169-174.

Becker A. & G. Theissen (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics and Evolution* 29**:** 464-489.

Behnke H.-D. (1969). Die Siebröhren-Plastiden bei Monocotylen. *Naturwissenschaften* 55**:** 140–141.

Bell C. D. & M. J. Donoghue (2005). Phylogeny and biogeography of Valerianaceae (Dipsacales) with special reference to the South American valerians. *Organisms Diversity & Evolution* 5**:** 147-159.

Bell C. D., D. E. Soltis, & P. S. Soltis (2005). The age of angiosperms: A molecular timescale without a clock. *Evolution* 59**:** 1245-1258.

Benton M. J. & F. J. Ayala (2003). Dating the tree of life. *Science* 300**:** 1698-1700.

Berenbaum M. (1983). Coumarins and caterpillars: a case for coevolution? *Evolution* 37**:** 163-179.

Berggren W. A., D. V. Kent, C. C. Swisher, & P.-P. Aubry (1995). A revised Cenozoic geochronology and chronostratigraphy. *In:* Geochronology, time scales and global stratigraphic correlation (W. A. Berggren, D. V. Kent, C. C. Swisher, P.-P. Aubry, and J. Hardenbol, Eds.), pp. 129-212, SEPM Spec. Publ., Tulsa.

Bermingham E. & A. P. Martin (1998). Comparative mtDNA phylogeography of neotropical freshwater fishes: Testing shared history to infer the evolutionary landscape of lower Central America. *Molecular Ecology* 7**:** 499-517.

Berthold P. (1993). Bird migration. A general survey. 2nd ed. Oxford University Press, Oxford.

Bessey C. E. (1915). The phylogenetic taxonomy of flowering plants. *Annals of the Missouri Botanical Garden* 2**:** 109-164.

Bonen L. & J. Vogel (2001). The ins and outs of group II introns. *Trends in Genetics* 17**:** 322-331.

Borsch T. (2000). Phylogeny and evolution of the genus *Nymphaea* (Nymphaeaceae). PhD thesis. Botanical Institute, University of Bonn, Bonn.

Borsch T., K. W. Hilu, D. Quandt, V. Wilde, C. Neinhuis, & W. Barthlott (2003). Non-coding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* 16**:** 558-576.

Borsch T., K. W. Hilu, J. H. Wiersema, C. Löhne, W. Barthlott, & V. Wilde (subm.). Phylogeny of *Nymphaea* (Nymphaeaceae): *trnT-F* evidence for evolution of the first global lineage of extant angiosperms. *International Journal of Plant Science*.

Borsch T., C. Löhne, K. Müller, K. W. Hilu, S. Wanke, A. Worberg, W. Barthlott, C. Neinhuis, & D. Quandt (2005). Towards understanding basal angiosperm diversification: recent insights using rapidly evolving genomic regions. *Nova Acta Leopoldina NF* 92**:** 85-110.

Bosch J. H. A., P. Cleveringa, & T. Meijer (2000). The Eemain stage in the Netherlands: history, character and new research. *Netherlands Journal of Geosciences* 79**:** 135-145.

Bowe L. M., G. Coat, & C. W. DePamphilis (2000). Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Science* 97**:** 4092-4097.

Bramwell D. (2002). How many plant species are there? *Plant Talk* 28**:** 32-34.

Bremer K. (1992). Ancestral areas: a cladistic reinterpretation of the center of origin concept. *Systematic Biology* 41**:** 436-445.

Bremer K. (1995). Ancestral areas: optimization and probability. *Systematic Biology* 44**:** 255-259.

Bremer K., E. Friis, & B. Bremer (2004). Molecular phylogenetic dating of asterid flowering plants shows Early Cretaceous diversification. *Systematic Biology* 53**:** 496-505.

Brenner G. (1962). The spores and pollen of the Potomac Group of Maryland. *Bulletin of the Department of Geology Mines and Water Resources, State of Maryland, Board of Natural Resources* 27**:** 1-215.

Brenner G. & I. S. Bickoff (1992). Palynology and age fo the Lower Cretaceous Basal Kurnub group from the coastal plain to the northern Negev of Israel. *Palynology* 16**:** 137-185.

Bromham L. & D. Penny (2003). The modern molecular clock. *Nature Reviews Genetics* 4**:** 216-224.

Brown R. (1826). Character and description of *Kingia* [...] with observations [...] on female flowers of Cycadaceae and Coniferae. *In:* Narrative of a survey of the intertropical coasts of Western Australia (P. P. King, Ed.), pp. 538-565, Murray, London.

Buckler E. S., A. Ippolito, & T. P. Holtsford (1997). The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* 145.

Bukowiecki H., M. Furmanowa, & H. Oledzka (1972). The numerical taxonomy of Nymphaeaceae Bentham et Hooker. *Acta Poloniae Pharmaceutica* 29**:** 319-327.

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

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

Burleigh J. G. & S. Mathews (2004). Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. *American Journal of Botany* 91**:** 1599-1613.

Campagna M. L. & S. R. Downie (1998). The intron in chloroplast gene *rpl16* is missing from the flowering plant families Geraniaceae, Goodeniaceae, and Plumbaginaceae. *Transactions of the Illinois State Academy of Science* 91**:** 1-11.

Campbell C. S., M. F. Wojciechowski, B. G. Baldwin, L. A. Alice, & M. J. Donoghue (1997). Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). *Molecular Biology and Evolution* 14**:** 81-90.

Capperino M. E. & E. L. Schneider (1985). Floral biology of *Nymphaea mexicana* Zucc. (Nymphaeaceae). *Aquatic Botany* 23**:** 83-93.

Caspary R. (1865). Nymphaeaceae. *Ann. Mus. Bot. Lugduno-Batavum* 2**:** 241-253.

Caspary R. (1891). Nymphaeaceae. *In:* Die natürlichen Pflanzenfamilien, III. Teil Abt. 2, 2a (A. Engler, and K. Prantl, Eds.), pp. 1-10, Wilhelm Engelmann, Leipzig.

Castoe T., T. Doan, & C. Parkinson (2004). Data Partitions and Complex Models in Bayesian Analysis: The Phylogeny of Gymnophthalmid Lizards. *Systematic Biology* 53**:** 448-469.

Cavagnetto C. & P. Anadón (1996). Preliminary palynological data on floristic and climatic changes during the Middle Eocene-Early Oligocene of the eastern Ebro Basin, northeast Spain. *Review of Palaeobotany and Palynology* 92**:** 281-305.

Cevallos-Ferriz S. R. S. & R. A. Stockey (1989). Permineralized fruits and seeds from the Princeton chert (Middle Eocene) of British Columbia: Nymphaeaceae. *Botanical Gazette* 150**:** 207-217.

Chanderbali A. S., H. Van der Werff, & S. S. Renner (2001). Phylogeny and historical biogeography of Lauraceae: evidence from the chloroplast and nuclear genomes. *Annals of the Missouri Botanical Garden* 88**:** 104-134.

Charalambidou I. & L. Santamaría (2002). Waterbirds as endozoochorous dispersers of aquatic organisms: a review of experimental evidence. *Acta Oecologica* 23**:** 165-176.

Chase M. W. (2004). Monocot relationships: an overview. *American Journal of Botany* 91**:** 1645-1655.

Chase M. W., D. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. Price, H. G. Hills, Y.-L. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, J. W. Kress, K. G. Karol, W. D. Clark, M. Hédren, B. S. Gaut, R. K. Jansen, K. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, Q.-Y. Xiang, G. M. Plunkett, et al. (1993). Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80**:** 528–580.

Chaw S.-M., C.-C. Chang, H.-L. Chen, & W.-H. Li (2004). Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *Journal of Molecular Evolution* 58**:** 424-441.

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

Chaw S.-M., A. Zharkikh, H.-M. Sung, T.-C. Lau, & W.-H. Li (1997). Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Molecular Biology and Evolution* 14**:** 56-68.

Chen I., S. R. Manchester, & Z. Chen (2004). Anatomically preserved seeds of *Nuphar* (Nymphaeaceae) from the Early Eocene of Wutu, Shandong Province, China. *American Journal of Botany* 91**:** 1265-1272.

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

Chippindale P. T. & J. J. Wiens (1994). Weighting, partitioning, and combining characters in phylogenetic analysis. *Systematic Biology* 43**:** 278-287.

Clausing G. & S. S. Renner (2001). Molecular phylogenetics of Melastomataceae and Memecylaceae: implications for character evolution. *American Journal of Botany* 88**:** 486-498.

Collinson M. E. (1980). Recent and Tertiary seeds of the Nymphaeaceae sensu lato with a revision of *Brasenia ovula* (Brong.) Reid and Chandler. *Annals of Botany* 46**:** 603-632.

Collinson M. E. (1992). Vegetational and floristic changes around the Eocene/Oligocene boundary in western and central Europe. *In:* Eocene-Oligocene climatic and biotic evolution (D. A. Prothero, and W. A. Berggren, Eds.), pp. 437-450, Princeton University Press, Princeton, NJ.

Collinson M. E. & J. J. Hooker (2003). Paleogene vegetation of Eurasia: framework for mammalian faunas. *In:* Distribution and migration of Tertiary mammals in Eurasia. A volume in honour of Hans de Bruijn. DEINSEA 10 (J. W. F. Reumer, and W. Wessels, Eds.), pp. 41-83.

Conard H. S. (1905). The waterlilies. A monograph of the genus *Nymphaea*. Lord Baltimore Press, Baltimore.

Cook C. D. K. (1988). Wind pollination in aquatic angiosperms. *Annals of the Missouri Botanical Garden* 75**:** 768-777.

Cook L. G. & M. D. Crisp (2005). Directional asymmetry of long-distance dispersal and colonization could mislead reconstructions of biogeography. *Journal of Biogeography* 32**:** 741-754.

Costa M., F. Michel, & E. Westhof (2000). A three-dimensional perspective on exon binding by a group II self-splicing intron. *The EMBO Journal* 19**:** 5007-5018.

Cramer J. M., A. D. J. Meeuse, & P. A. Teunissen (1975). A note on the pollination of nocturnally flowering species of *Nymphaea*. *Acta Botanica Neerl.* 24**:** 489-490.

Crane P. R. (1985). Phylogenetic analysis of seed plants and the origin of angiosperms. *Annals of the Missouri Botanical Garden* 72**:** 716-793.

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

Crane P. R., P. Herendeen, & E. M. Friis (2004). Fossils and plant phylogeny. *American Journal of Botany* 91**:** 1683-1699.

Crane P. R., K. R. Pedersen, E. M. Friis, & A. N. Drinnan (1993). Early Cretaceous (early to middle Albian) platanoid inflorescences associated with Sapindopsis leaves from the Potomac Group of eastern North America. *Systematic Botany* 18**:** 328–344.

Crawford D. J., E. Landolt, & D. H. Les (1996). An allozyme study of two sibling species of Lemna (Lemnaceae) with comments on their morphology, ecology and distribution. *Bulletin of the Torrey Botanical Club* 123**:** 1-6.

Crawford D. J., E. Landolt, D. H. Les, & E. Tepe (1997). Allozyme variation and the taxonomy of Wolffiella. *Aquatic Botany* 58**:** 43-54.

Creer S., A. Malhotra, & R. S. Thorpe (2003). Assessing the phylogenetic utility of four mitochondrial genes and a nuclear intron in the Asian pit viper genus, *Trimeresurus*: Separate, simultaneous, and conditional data combination analyses. *Molecular Biology and Evolution* 20**:** 1240-1251.

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

Crepet W. L. & K. C. Nixon (1998a). Fossil Clusiaceae from the Late Cretaceous (Turonian) of New Jersey and implications regarding the history of bee pollination. *American Journal of Botany* 85**:** 1122–1133.

Crepet W. L. & K. C. Nixon (1998b). Two new fossils of magnoliid affinity from the Late Cretaceous of New Jersey. *American Journal of Botany* 85**:** 1273-1288.

Crepet W. L., K. C. Nixon, & M. A. Gandolfo (2004). Fossil evidence and phylogeny: the age of major angiosperm clades based on mesofossil and macrofossil evidence from Cretaceous deposits. *American Journal of Botany* 91**:** 1666-1682.

Crisp M., L. Cook, & D. Steane (2004). Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present-day communities? One contribution of 16 to a Discussion Meeting Issue 'Plant phylogeny and the origin of major biomes'. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 359**:** 1551 - 1571.

Crisp M. D. & L. Cook (2005). Do early branching lineages signify ancestral traits? *Trends in Ecology and Evolution* 20**:** 123-128.

Cronquist A. (1981). An integrated system of classification of flowering plants. Columbia University Press, New York.

Cronquist A. (1988). The evolution and classification of flowering plants. 2nd edn. The New York Botanical Garden, Bronx, NY.

Cunningham C. W. (1997). Can three incongruence tests predict when data should be combined? *Molecular Biology and Evolution* 14**:** 733-740.

Dahlgren R. (1983). General aspects of angiosperm evolution and macrosystematics. *Nordic Journal of Botany* 3**:** 119-149.

Dahlgren R., H. T. Clifford, & P. F. Yeo (1985). The families of monocotyledons. Structure, evolution and taxonomy. Springer, Berlin, Heidelberg, New York.

Dahlgren R., B. Hansen, K. Jakobsen, S. R. Jensen, K. Larsen, & B. J. Nielsen (1979-1980). Angiospermernes taxonomi. 2nd ed. Akademisk Forlag, Kopenhagen.

Darwin C. (1903). [Letter from 1879] to J. D. Hooker / [Letter] to Oswald Heer. *In:* More letters of Charles Darwin: A record of his work in a series of hitherto unpublished

papers. (F. Darwin, and A. C. Seward, Eds.), pp. pp. 20-22 / 239-240, J. Murray, London.

Davies J. T., T. G. Barraclough, M. W. Chase, P. S. Soltis, D. E. Soltis, & V. Savolainen (2004). Darwin's abominable mystery: Insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences* 101**:** 1904-1909.

Davis C. C., C. D. Bell, P. W. Fritsch, & S. Mathews (2002a). Phylogeny of *Acridocarphus-Brachylophon* (Malphighiceae): Implications for Tertiary tropical floras and Afroasian biogeography. *Evolution* 56**:** 2395-2405.

Davis C. C., C. D. Bell, S. Mathews, & M. J. Donoghue (2002b). Laurasian migration explains Gondwanan disjunctions: Evidence from Malpighiaceae. *Proceedings of the National Academy of Sciences* 99**:** 6833-6837.

Davis C. C., P. W. Fritsch, C. D. Bell, & S. Mathews (2004). High-latitude tertiary migrations of an exclusively tropical clade: Evidence from Malpighiaceae. *International Journal of Plant Sciences* 165**:** S107-S121.

De Bodt S., S. Maere, & Y. Van de Peer (2005). Genome duplication and the origin of angiosperms. *Trends in Ecology and Evolution* 20**:** 591-597.

De Candolle A.-P. (1813). Théorie élémentaire de la botanique, ou exposition des principes de la classification naturelle et de l'art de décrire et d'étudier les végétaux. Deterville, Paris.

De Candolle M. (1819). Sur les affinités naturelles de la famille des Nymphaeacées. *In:* Mémoires de la Société De Physique et D´Histoire Naturelle de Geneve, J.J. Paschoud, Geneve.

De Pinna M. C. C. (1991). Concepts and tests of homology in the cladistic paradigm. *Cladistics 7***:** 367-394.

De Queiros A. (2005). The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology and Evolution* 20**:** 68-73.

De Queiros A., M. J. Donoghue, & J. Kim (1996). Separate versus combined analysis of phylogenetic evidence. *Annual Reviews of Ecology and Systematics* 26**:** 657-681.

Den Hartog C. (1970). *Ondinea*, a new genus of Nymphaeaceae. *Blumea* 18**:** 413-417.

Denk T., F. Grímsson, & Z. Kvacek (2005). The Miocene floras of Iceland and their significance for late Cainozoic North Atlantic biogeography. *Botanical Journal of the Linnean Society* 149**:** 369-417.

Denk T. & I.-C. Oh (2006). Phylogeny of Schisandraceae based on morphological data: evidence from modern plants and the fossil record. *Plant Systematics and Evolution* 256**:** 113 - 145.

Dib-Hajj S. D., S. C. Boulanger, S. K. Hebbar, C. L. Peebles, J. S. Franzen, & P. S. Perlman (1993). Domain 5 interacts with domain 6 and influences the second transesterification reaction of group II intron self-splicing. *Nucleic Acids Research* 21**:** 1797-1804.

Dilcher D. & P. R. Crane (1984). Archaeanthus: an early angiosperm from the Cenomanian of the western interior of North America. *Annals of the Missouri Botanical Garden 71***:** 351-738.

Dilcher D. L. (1989). The occurrence of fruits with affinities to Ceratophyllaceae in Lower and mid-Cretaceous sediments. *American Journal of Botany* 76**:** 162 (Abstr.).

Dixit R., P. K. Trivedi, P. Nath, & R. V. Sane (1999). Organization and post-transcriptional processing of the *psbB* operon from chloroplasts of *Populus deltoides*. *Current Genetics* 36**:** 165-172.

Donoghue M. J. & J. A. Doyle (1989). Phylogenetic analysis of angiosperms and the relationships of Hamamelidae. *In:* Evolution, systematics, and fossil history of the Hamamelidae (P. R. Crane, and S. Blackmore, Eds.), pp. 17-45, Claredon Press, Oxford, UK.

Donoghue M. J. & J. A. Doyle (2000). Seed plant phylogeny: demise of the anthophyte hypothesis? *Current Biology* 10**:** R106–R109.

Donoghue M. J. & B. R. Moore (2003). Toward an integrative historical biogeography. *Integrative and Comparative Biology* 43**:** 261-270.

Dorofeev P. I. (1963). Tertiary floras of western Siberia. Komorov Botanical Institute, Academy Nauk SSSR, Leningrad, USSR.
Dorofeev P. I. (1974). Nymphaeales. *In:* Magnoliophyta Fossilia U.R.S.S. (A. L. Takhtajan, Ed.), pp. 52-88, Nauka, Leningrad, USSR.
Dorofeev P. I. (1984). The taxonomy and history of the genus *Brasenia* (Cabombaceae). *Bot. Zhurn. (Leningr.)* 69**:** 137-148.
Doyle J. A. (1992a). Revised palynological correlations of the Potomac Group (USA) and the Coco beach sequence of Gabon (Barremian-Aptian). *Cretaceous Research* 13**:** 337–349.
Doyle J. A. (1996). Seed plant phylogeny and the relationships of Gnetales. *International Journal of Plant Sciences* 157**:** S3-S39.
Doyle J. A. (1998). Molecules, morphology, fossils, and the relationship of angiosperms and Gnetales. *Molecular Phylogenetics and Evolution* 9**:** 448–462.
Doyle J. A. (2000). Paleobotany, relationships, and geographic history of Winteraceae. *Annals of the Missouri Botanical Garden* 87**:** 303-316.
Doyle J. A. & M. J. Donoghue (1986). Seed plant phylogeny and the origin of the angiosperms: an experimental cladistic approach. *The Botanical Review* 52**:** 321-431.
Doyle J. A. & M. J. Donoghue (1992). Fossils and seed plant phylogeny reanalyzed. *Brittonia* 44**:** 89-106.
Doyle J. A. & P. K. Endress (2000). Morphological phylogenetic analysis of basal angiosperms: Comparison and combination with molecular data. *International Journal of Plant Sciences* 161**:** S121-S153.
Doyle J. A. & L. J. Hickey (1976). Pollen and leaves from the mid-Cretaceous Potomac group and their bearing on early angiosperm evolution. *In:* Origin and early evolution of angiosperms (C. B. Beck, Ed.), pp. 139-206, Columbia University Press, New York.
Doyle J. A. & C. L. Hotton (1991). Diversification of early angiosperm pollen in a cladistic context. *In:* Pollen and spores (S. Blackmore, and S. H. Barnes, Eds.), pp. 169-195, Clarendon, Oxford.
Doyle J. A., H. Sauquet, T. Scharaschkin, & A. Le Thomas (2004). Phylogeny, molecular and fossil dating, and the biogeographic history of Annonaceae and Myristicaceae (Magnoliales). *International Journal of Plant Sciences* 165**:** S55-S67.
Doyle J. J. (1992b). Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany* 17**:** 144-163.
Drinnan A. N., P. R. Crane, E. M. Friis, & K. R. Pedersen (1990). Lauraceous flowers from the Potomac Group (mid-Cretaceous) of eastern North America. *Botanical Gazette* 151**:** 370-384.
Duvall M., G. Learn, Jr, L. Eguiarte, & M. Clegg (1993). Phylognetic analysis of *rbcL* sequences identifies *Acorus calamus* as the primal extant monocotyledon. *Proceedings of the National Academy of Sciences* 90**:** 4641-4644.
Eklund H. (2004). First Cretaceous flowers from Antarctica. *Review of Palaeobotany and Palynology* 127**:** 187-217.
Eklund H., J. A. Doyle, & P. S. Herendeen (2004). Morphological phylogenetic analysis of living and fossil Chloranthaceae. *International Journal of Plant Sciences* 165**:** 107-151.
Elder J. F. & B. J. Turner (1995). Concerted evolution of repetitive DNA sequences in Eukaryotes. *The Quarterly Review of Biology* 70**:** 297-320.
Emboden W. A. (1978). The sacred narcotic lily of the Nile: *Nymphaea caerulea*. *Economic Botany* 32**:** 395-407.
Emboden W. A. (1979). Narcotic Plants. Revised edition. Macmillan, New York.
Endress P. K. (2001). The flowers in extant basal angiosperm and inferences on ancestral flowers. *International Journal of Plant Sciences* 162**:** 1111-1140.
Endress P. K. (2004). L.A.S. Johnson Review No. 3. Structure and relationships of basal relictual angiosperms. *Australian Systematic Botany* 17**:** 343-366.

Endress P. K. & A. Igersheim (2000). Gynoceum structure and evolution in basal angiosperms. *International Journal of Plant Sciences* 161**:** S211-S223.

Engler A. (1886). Führer durch den Königlich Botanischen Garten der Universität zu Breslau. J. U. Kerns, Breslau.

Engler A. (1892). Syllabus der Vorlesungen über die specielle und medicinisch-pharmaceutische Botanik: eine Übersicht über das gesammte Pflanzensystem mit Berücksichtigung der Medicinal-und Nutzpflanzen. Berlin.

Engler A. (1897). Übersicht über die Unterabteilungen, Klassen, Reihen, Unterreihen und Familien der Embryophyta (Siphonogama). *In:* Die natürlichen Pflanzenfamilien. Nachträge zu II-IV (A. Engler, and K. Prantl, Eds.), pp. 341-380, Wilhelm Engelmann, Leipzig.

Enrödi S. (1985). The Dynastinae of the world. Dr W. Junk Publishers, Dordrecht.

Ervik F. & J. T. Knudsen (2003). Water lilies and scarabs: faithful partners for 100 million years? *Biological Journal of the Linnean Society* 80**:** 539-543.

Farrell B. D. (1998). "Inordinate fondness" explained: Why are there so many beetles? *Science* 281**:** 553-557.

Farris J. S. (1989). The retention index and the rescaled consistency index. *Cladistics* 5**:** 417-419.

Federova O., T. Mitros, & A. M. Pyle (2003). Domains 2 and 3 interact to form critical elements of the group II intron active site. *Journal of Molecular Biology* 330**:** 197-209.

Feild T. S. & N. C. Arens (2005). Form, function and environments of the early angiosperms: merging extant phylogeny and ecophysiology with fossils. *New Phytologist* 166**:** 383-408.

Feild T. S., N. C. Arens, & T. E. Dawson (2003). The ancestral ecology of angiosperms: emerging perspectives from extant basal lineages. *International Journal of Plant Sciences* 164**:** S129-S142.

Feild T. S., N. C. Arens, J. A. Doyle, T. E. Dawson, & M. J. Donoghue (2004). Dark and disturbed: a new image of early angiosperm ecology. *Paleobiology* 30**:** 82-107.

Feild T. S., T. J. Brodribb, & N. M. Holbrook (2002). Hardly a relict: Freezing and the evolution of vesselless wood in Winteraceae. *Evolution* 56**:** 464-478.

Forbis T. A., S. K. Floyd, & A. De Queiros (2002). The evolution of embryo size in angiosperms and other seed plants: implications for the evolution of seed dormancy. *Evolution* 56**:** 2112-2125.

Fossen T. & O. M. Andersen (1999). Delphinidin 3-galloyl-galactosids from blue flowers of *Nymphaea caerula*. *Phytochemistry* 50**:** 1185-1188.

Fossen T., A. Larsen, & O. M. Andersen (1998). Anthocyanins from flowers and leaves of Nymphaea x marliacea cultivars. *Phytochemistry* 48**:** 823-827.

Friedman W. E. & J. H. Williams (2004). Developmental evolution of the sexual process in ancientflowering plant lineages. *Plant Cell* 16**:** S119-132.

Friis E. M., J. A. Doyle, P. K. Endress, & Q. Leng (2003). *Archaefructus* - angiosperm precursor or specialized early angiosperm? *Trends in Plant Science* 8**:** 369–373.

Friis E. M., H. Eklund, K. R. Pedersen, & P. R. Crane (1994a). *Virginianthus calycanthoides* gen. et sp. nov. - a calycanthaceous flower from the Potomac Group (Early Cretaceous) of eastern North America. *International Journal of Plant Sciences* 155**:** 772-785.

Friis E. M., K. R. Pedersen, & P. R. Crane (1994b). Angiosperm floral structures from the Early Cretaceous of Portugal. *Plant Systematics and Evolution* Suppl. 8**:** 31-49.

Friis E. M., K. R. Pedersen, & P. R. Crane (1999). Early angiosperm diversification: the diversity of pollen associated with angiosperm reproductive structures in Early Cretaceous floras from Portugal. *Annals of the Missouri Botanical Garden* 86**:** 259–296.

Friis E. M., K. R. Pedersen, & P. R. Crane (2001). Fossil evidence of water lilies (Nymphaeales) in the Early Cretaceous. *Nature* 410**:** 357-360.

Friis E. M., K. R. Pedersen, & P. R. Crane (2004). Araceae from the Early Cretaceous of Portugal: Evidence on the emergence of monocotyledons. *Proceedings of the National Academy of Sciences* 101**:** 16565-16570.

Frohlich M. & D. S. Parker (2000). The mostly male theory of flower evolutionary origins: from genes to fossils. *Systematic Botany* 25**:** 155-170.

Frohne D. & U. Jensen (1998). Systematik des Pflanzenreichs (unter besonderer Berücksichtigung chemischer Merkmale und pflanzlicher Drogen). 5th ed. Wissenschaftliche Verlagsgesellschaft mbH, Stuttgart.

Frumin S. & E. M. Friis (1999). Magnoliid reproductive organs from the Cenomanian-Turonian of north-western Kazakhstan: Magnoliaceae and Illiciaceae. *Plant Systematics and Evolution* 216**:** 265-288.

Fuertes Aguilar J. & G. Nieto Feliner (2003). Additive polymorphisms and reticulation in an ITS phylogeny of thrifts (Armeria, Plumbaginaceae). *Molecular Phylogenetics and Evolution* 28**:** 430–447.

Fuertes Aguilar J., J. A. Rosselló, & G. Nieto Feliner (1999). nrDNA concerted evolution in natural and artificial hybrids of *Armeria* (Plumbaginaceae). *Molecular Ecology* 8**:** 1341-1346.

Gandolfo M. A., K. C. Nixon, & W. L. Crepet (1998a). *Tylerianthus crossmanensis* gen. et sp. nov. (aff. Hydrangeaceae) from the Upper Cretaceous of New Jersey. *American Journal of Botany* 85**:** 376-386.

Gandolfo M. A., K. C. Nixon, & W. L. Crepet (2004). Cretaceous flowers of Nymphaeaceae and implications for complex insect entrapment pollination mechanisms in early Angiosperms. *Proceedings of the National Academy of Sciences* 101**:** 8056-8060.

Gandolfo M. A., K. C. Nixon, W. L. Crepet, D. W. Stevenson, & E. M. Friis (1998b). Oldest known fossils of monocotyledons. *Nature* 394**:** 532-533.

Gaston K. J. (2000). Global patterns in biodiversity. *Nature* 405**:** 220-227.

Giribert G. & W. Wheeler (1999). On gaps. *Cladistics* 13**:** 132-143.

Givnish T. J., K. C. Millam, T. M. Evans, J. C. Hall, J. C. Pires, P. E. Berry, & K. J. Sytsma (2004). Ancient vicariance or recent long-distance dispersal? Inferences about phylogeny and South American–African disjunctions in Rapateaceae and Bromeliaceae based on *ndhF* sequence data. *International Journal of Plant Sciences* 165**:** S35-S54.

Givnish T. J. & S. S. Renner (2004). Tropical intercontinental disjunctions: Gondwana breakup, immigration from the boreotropics, and transoceanic dispersal. *International Journal of Plant Sciences* 165**:** S1-S6.

Golenberg E. M., M. T. Clegg, M. L. Durbin, J. Doebley, & D. P. Ma (1993). Evolution of a non-coding region of the chloroplast genome. *Molecular Phylogenetics & Evolution* 2**:** 52-64.

Goremykin V. V., V. Bobrova, J. Pahnke, J. Troitsky, A. S. Antonov, & W. Martin (1996). Noncoding sequences from the slowly evolving chloroplast inverted repeat in addition to the *rbcL* data do not support gnetalean affinities of angiosperms. *Molecular Biology and Evolution* 13**:** 383-396.

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

Goremykin V. V., K. I. Hirsch-Ernst, S. Wölfl, & F. H. Hellwig (2004). The chloroplast genome of *Nymphaea alba*: Whole-genome analyses and the problem of identifying the most basal angiosperm. *Molecular Biology and Evolution* 21**:** 1445-1454.

Govaerts R. (2001). How many species of seed plants are there? *Taxon* 50**:** 1085-1090.

Gradstein F. M., F. P. Agterberg, J. G. Ogg, J. Hardenbol, P. Van Veen, J. Thierry, & Z. Huang (1994). A Mesozoic time scale. *Journal of Geophysical Research, Washington* 99**:** 24051-24074.

Graham S. W. & R. G. Olmstead (2000a). Evolutionary significance of an unusual chloroplast DNA inversion found in two basal angiosperm lineages. *Current Genetics* 37**:** 183-188.

Graham S. W. & R. G. Olmstead (2000b). Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87**:** 1712-1730.

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

Graur D. & W. Martin (2004). Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends in Genetics* 20**:** 80-86.

Green A. J., J. Figuerola, & M. I. Sánchez (2002). Implications of waterbird ecology for the dispersal of aquatic organisms. *Acta Oecologica* 23**:** 177-189.

Grüß J. (1927a). Die Haustoren der Nymphaeaceen. *Berichte der Deutschen Botanischen Gesellschaft* XLV**:** 459-466.

Grüß J. (1927b). Die Luftblätter der Nymphaeaceen. *Berichte der Deutschen Botanischen Gesellschaft* XLV**:** 454-458.

Gu X. & W. H. Li (1995). The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment. *Journal of Molecular Evolution* 40**:** 464-473.

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

Gupta P. P. (1978). Cytogenetics of aquatic ornamentals. II. Cytology of *Nymphaeas*. *Cytologia* 43**:** 477-484.

Gupta P. P. (1980). Cytogenetics of aquatic ornamentals. VI. Evolutionary trends and relationships in the genus *Nymphaea*. *Cytologia* 45**:** 307-314.

Hably L., Z. Kvacek, & S. R. Manchester (2000). Shared taxa of land plants in the Oligocene of Europe and North America in context of Holarctic phytogeography. *Acta Univ Carol Geol* 44**:** 59–74.

Hall B. K. (1998). The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. *In:* Biogeography and geological evolution of SE Asia (B. K. Hall, and J. D. Holloway, Eds.), Backhuys, Leiden.

Hall T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41**:** 95-98.

Hallam A. (1992). Phanerozoic sea level changes. Columbia University Press, New York.

Hamby R. K. & E. A. Zimmer (1992). Ribosomal RNA as a phylogenetic tool in plant systematics. *In:* Molecular systematics of plants (P. S. Soltis, D. E. Soltis, and J. J. Doyle, Eds.), Chapman and Hall, New York, USA.

Hamilton M. B., J. M. Braverman, & D. F. Soria-Hernanz (2003). Patterns and relative rates of nucleotide and insertion/deletion evolution at six chloroplast intergenic regions in new world species of the Lecythidaceae. *Molecular Biology and Evolution* 20**:** 1710-1721.

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

Hasebe M., R. Kofuki, M. Ito, M. Kato, & K. U. Iwatsuki, K. (1992). Phylogeny of gymnosperms inferred from *rbcL* gene sequences. *Botanical Magazine, Tokyo* 105**:** 673-679.

Hausdorf B. (1998). Weighted ancestral area analysis and a solution of the redundant distribution problem. *Systematic Biology* 47**:** 445-456.

Hay W. W., R. DeConto, C. N. Wold, K. M. Wilson, S. Voigt, M. Schulz, A. Wold-Rossby, W.-C. Dullo, A. B. Ronov, A. N. Balukhovsky, & E. Soeding (1999). Alternative global Cretaceous paleogeography. *In:* The Evolution of Cretaceous Ocean/Climate Systems, Geological Society of America Special Paper 332 (E. Barrera, and C. Johnson, Eds.), pp. 1-47.

Heads M. (2005). Dating nodes on molecular phylogenies: a critique of molecular biogeography. *Cladistics* 21**:** 62-78.
Hedges S. B. & S. Kumar (2004). Precision of molecular time estimates. *Trends in Genetics* 20**:** 242-247.
Hennig W. (1950). Grundzüge einer Theorie der phylogenetischen Systematik. Deutscher Zentralverlag, Berlin.
Hennig W. (1965). Phylogenetic systematics. *Annual Review of Entomology* 10**:** 97-116.
Hennig W. (1966). Phylogenetic systematics. University of Illinois Press, Urbana.
Hennig W. (1974). Cladistic analysis or cladistic classification? A reply to Ernst Mayr. *Systematic Zoology* 24**:** 244-256.
Herendeen P. S., D. H. Les, & D. L. Dilcher (1990). Fossil *Ceratophyllum* (Ceratophyllaceae) from the Tertiary of North America. *American Journal of Botany* 77**:** 7-16.
Herngreen G. F. W., M. Kedves, L. V. Rovnina, & S. B. Smirnova (1996). Cretaceous palynological provinces: a review. *In:* Palynology: Principles and Applications (J. Jansonius, and D. C. McGregor, Eds.), pp. 1157-1188, American Association of Stratigraphic Palynologists Foundation.
Hesse M. & R. Zetter (2005). Ultrastructure and diversity of recent and fossil zona-aperturate pollen grains. *Plant Systematics and Evolution* 255**:** 145-176.
Hileman L. C., M. C. Vasey, & V. T. Parker (2001). Phylogeny and biogeography of the Arbutoideae (Ericaceae): Implications for the Madrean-Tethyan hypothesis. *Systematic Botany* 26**:** 131-143.
Hillis D. M., M. W. Allard, & M. M. Miyamoto (1993). Analysis of DNA sequence data: phylogenetic inference. *Methods in Enzymology* 224**:** 456-490.
Hillis D. M., C. Moritz, C. A. Porter, & R. J. Baker (1991). Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* 251**:** 308-310.
Hilu K. W., T. Borsch, K. Müller, D. E. Soltis, P. S. Soltis, V. Savolainen, M. W. Chase, M. P. Powell, L. A. Alice, R. Evans, H. Sauquet, C. Neinhuis, T. A. B. Slotta, J. G. Rohwer, C. S. Campbell, & L. W. Chatrou (2003). Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90**:** 1758-1776.
Hilu K. W. & H. Liang (1997). The *matK* gene: sequence variation and application in plant systematics. *American Journal of Botany* 84**:** 830-839.
Hirthe G. & S. Porembski (2003). Pollination of *Nymphaea lotus* (Nymphaeaceae) by rhinoceros beetles and bees in the northeastern ivory coast. *Plant Biology* 5**:** 670-675.
Hoot S. B. & A. W. Douglas (1998). Phylogeny of the Proteaceae based on *atpB* and *atpB-rbcL* intergenic spacer region sequences. *Australian Systematic Botany* 11**:** 301-320.
Hoppe R. T., L. M. Smith, & D. B. Webster (1986). Foods of wintering diving ducks in South Carolina. *Journal of Field Ornithology* 57**:** 126-134.
Huelsenbeck J. P., B. Larget, & D. Swofford (2000). A compound Poisson process for relaxing the molecular clock. *Genetics* 154**:** 1879-1892.
Huelsenbeck J. P., R. Nielsen, & J. P. Bollback (2001). Empirical and hierarchical Bayesian estimation of ancestral states. *Systematic Biology* 50**:** 673-688.
Huelsenbeck J. P. & F. Ronquist (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17**:** 754-755.
Hughes N. F. (1994). The enigma of angiosperm origins. Cambridge University Press, Cambridge, UK.
Hughes N. F., A. B. McDougall, & J. L. Chapman (1991). Exceptional new record of Cretaceous Hauterivian angiospermoid pollen from Southern England. *Journal of Micropalaeontology* 10**:** 75-82.
Ito M. (1987). Phylogenetic Systematics of Nymphaeales. *The Botanical Magazin, Tokyo* 100**:** 17-35.
Iturralde-Vinent M. A. & R. D. E. MacPhee (1999). Paleogeography of the Caribbean region: Implications for Cenozoic biogeography. *Bulletin of the American Museum of Natural History* 238**:** 1–95.

Jacobs B. F. & P. S. Herendeen (2004). Eocene dry climate and woodland vegetation in tropical Africa reconstructed from fossil leaves from northern Tanzania. *Palaeogeography, Palaeoclimatology, Palaeoecology* 213**:** 115– 123.

Jacobs S. W. L. (1992). New species, lectotypes and synonyms of Australasian *Nymphaea* (Nymphaeaceae). *Telopea* 4**:** 635-641.

Jacobs S. W. L. (1994). Further notes on Nymphaea (Nymphaeaceae) in Australasia. *Telopea* 5**:** 403-706.

Jacobs S. W. L. & C. B. Hellquist (in press.). Three new species of *Nymphaea* (Nymphaeaceae) in Australia. *Telopea*.

Jacobs S. W. L. & C. L. Porter (in prep.). Nymphaeaceae. *In:* Flora of Australia [draft manuscript] (A. S. George, Ed.).

Jacobs S. W. L. & C. L. Porter (in press). Nymphaeaceae. *In:* Flora of Australia Vol. 2 (A. Wilson, Ed.).

Ji Q., H. Li, L. M. Bowe, Y. Liu, & D. W. Taylor (2004). Early Cretaceous *Archaefructus eoflora* sp. nov. with bisexual flowers from Beipiao, Western Liaoning, China. *Acta Geologica Sinica* 78**:** 883-896.

Johnson L. A. & D. E. Soltis (1995). Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *matK* sequences. *Annals of the Missouri Botanical Garden* 82**:** 149-175.

Judd W. S., C. D. Campbell, E. A. Kellogg, P. F. Stevens, & M. J. Donoghue (2002). Plant systematics: A phylogenetic approach. 2nd ed. Sinauer Associates, Inc., Sunderland, MA.

Judd W. S. & R. G. Olmstead (2004). A survey of tricolpate (eudicot) phylogenetic relationships. *American Journal of Botany* 91**:** 1627-1644.

Kadereit G., D. Gotzek, S. W. L. Jacobs, & H. Freitag (2005). Origin and age of Australian Chenopodiaceae. *Organisms Diversity & Evolution* 5**:** 59-80.

Kadono Y. & E. L. Schneider (1987). The life history of *Euryale ferox* Salisb. in southwestern Japan with special reference to reproductive ecology. *Plant Species Biology* 2**:** 109-115.

Kak A. M. & S. Durani (1986). A contribution to the seed anatomy of *Nelumbium nuciferum* Gaertn. *Journal of Plant Anatomy and Morphology* 3**:** 59-64.

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

Källersjö M., J. S. Farris, M. W. Chase, B. Bremer, M. Fay, C. J. Humphries, G. Petersen, O. Seberg, & K. Bremer (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 Systematics and Evolution* 213**:** 259-287.

Kelchner S. A. (2000). The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* 87**:** 482-498.

Kelchner S. A. (2002). Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. *American Journal of Botany* 89**:** 1651-1669.

Kelchner S. A. & L. G. Clark (1997). Molecular evolution and phylogenetic utility of the chloroplast rpl16 intron in *Chusquea* and the Bambusoideae (Poaceae). *Molecular Phylogenetics & Evolution* 8**:** 385-397.

Kelchner S. A. & J. F. Wendel (1996). Hairpins create minute inversions in non-coding regions of chloroplast DNA. *Current Genetics* 30**:** 259-262.

Kenneally K. F. & E. L. Schneider (1983). The genus *Ondinea* (Nymphaeaceae) including a new subspecies from the Kimberley region, Western Australia. *Nuytsia* 4**:** 359-365.

Kennedy E. M. (2003). Discovery of a Cretaceous angiosperm reproductive structure from New Zealand. *New Zealand Journal of Geology and Geophysics* 46**:** 519-522.

Kerner von Marilaun A. (1891). Pflanzenleben, Bd. 2 Geschichte der Pflanzen. Verlag des Bibliographischen Instituts, Leipzig.

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

Kim S., M.-J. Yoo, V. A. Albert, J. S. Farris, P. S. Soltis, & D. E. Soltis (2004b). Phylogeny and diversification of B-function MADS-box genes in angiospermes: Evolutionary and functional implications of a 260-million-year-old duplication. *American Journal of Botany* 91**:** 2102-2118.

Kishino H. & M. Hasegawa (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of Molecular Evolution* 29**:** 170-179.

Kitching I. J., P. L. Forey, C. J. Humphries, & D. M. Williams (1998). Cladistics. 2nd edition. Oxford University Pressh.

Kjer K. M. (1995). Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: An example of alignment and data presentation from the frogs. *Molecular Phylogenetics & Evolution* 4**:** 314-330.

Kluge A. G. & J. S. Farris (1969). Quantitative phyletics and the evolution of anurans. *Systematic Zoology* 18**:** 1-32.

Knoop V. & A. Brennicke (1993). Group II introns in plant mitochondria - Trans-splicing, RNA editing, evolution and promiscuity. *In:* Plant Mitochondria (A. Brennicke, and U. Kück, Eds.), pp. 221-232, VCH Verlag, Weinheim, Germany.

Koontz J. A., P. S. Soltis, & D. E. Soltis (2004). Using phylogeny reconstruction to test hypotheses of hybrid origin in *Delphinium* section *Diedropetala* (Ranunculaceae). *Systematic Botany* 29**:** 345-357.

Kress J. W., L. M. Prince, W. H. Hahn, & E. A. Zimmer (2001). Unraveling the evolutionary radiation of the families of the Zingiberales using morphological and molecular evidence. *Systematic Biology* 50**:** 926-944.

Kumar S., K. Tamura, I. B. Jakobsen, & M. Nei (2001). MEGA 2.1 Molecular evolutionary genetics analysis software, Arizona State University, Tempe, Arizona, USA.

Labandeira C. C. (2002). The history of associations between plants and animals. *In:* Plant-Animal Interactions: An Evolutionary Approach (C. M. Herrera, and O. Pellmyr, Eds.), pp. 26-74, 248-61, Blackwell Science, Oxford, UK.

Labandeira C. C., D. L. Dilcher, D. R. Davis, & D. L. Wagner (1994). Ninety-seven million years of angiosperm-insect association: paleobiological insights into the meaning of coevolution. *Proceedings of the National Academy of Science* 91**:** 12278-12282.

Lambkin C. L. (2004). Partitioned Bremer support localises significant conflict in bee flies (Diptera: Bombyliidae: Anthracinae). *Invertebrate Systematics* 18**:** 351-360.

Langlet O. & E. Söderberg (1927). Über die Chromosomenzahlen einiger Nymphaeaceen. *Acta Horti Berg.* 9**:** 85-104.

Laroche J. & J. Bousquet (1999). Evolution of the mitochondrial *rps3* intron in perennial and annual angiosperms and homology to *nad5* intron 1. *Molecular Biology and Evolution* 16**:** 441-452.

Lavin M., P. S. Herendeen, & M. F. Wojciechowski (2005). Evolutionary rate analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Biology* 54**:** 575-594.

Lavin M., B. P. Schrire, G. Lewis, R. T. Pennington, A. Delgado-Salinas, M. Thulin, C. E. Hughes, A. B. Matos, & M. F. Wojciechowski (2004). Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. One contribution of 16 to a Discussion Meeting Issue 'Plant phylogeny and the origin of major biomes'. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 359**:** 1509 - 1522.

Lavin M., M. F. Wojciechowski, A. Richman, J. Rotella, M. J. Sanderson, & A. B. Matos (2001). Identifying tertiary radiations of Fabaceae in the Greater Antilles: Alternatives to cladistic vicariance analysis. *International Journal of Plant Sciences* 162**:** S53-S76.

Lawrence G. H. L. (1951). Taxonomy of vascular plants. Macmillan, New York.

Lawver L. A. & L. M. Gahagan (1998). Opening of the Drake Passage and its impact on Cenozoic ocean circulation. *In:* Tectonic boundary conditions for climate reconstructions. *Oxford Monographs on Geology and Geophysics* (T. J. Crowley, and K. C. Burke, Eds.), pp. 212-223.

Learn G. H. J., J. S. Shore, G. R. Furnier, G. Zurawski, & M. T. Clegg (1992). Constraints on the evolution of plastid introns: The Group II intron in the gene encoding tRNA-Val (UAC). *Molecular Biology and Evolution* 9**:** 856-871.

Leebens-Mack J., L. A. Raubeson, L. Cui, J. V. Kuehl, M. H. Fourcade, T. W. Chumley, J. L. Boore, R. K. Jansen, & C. W. dePamphilis (2005). Identifying the Basal Angiosperm Node in Chloroplast Genome Phylogenies: Sampling One's Way Out of the Felsenstein Zone. *Molecular Biology and Evolution* 22**:** 1948-1963.

Legendre S. & J.-L. Hartenberger (1992). Evolution of mammalian faunas in Europe during the Eocene and Oligocene. *In:* Eocene-Oligocene climatic and biotic evolution (D. A. Prothero, and W. A. Berggren, Eds.), pp. 516-528, Princeton University Press, Princeton, N.J.

Les D. H. (1986). The phytogeography of *Ceratophyllum demersum* and *C. echinatum* (Ceratophyllaceae) in glaciated North America. *Canadian Journal of Botany* 64**:** 498-509.

Les D. H. (1988). The origin and affinities of the Ceratohyllaceae. *Taxon* 37**:** 326-345.

Les D. H., D. J. Crawford, R. T. Kimball, M. L. Moody, & E. Landolt (2003). Biogeography of discontinuously distributed hydrophytes: A molecular appraisal of intercontinental disjunctions. *International Journal of Plant Sciences* 164**:** 917-932.

Les D. H., D. K. Garvin, & C. F. Wimpee (1991). Molecular evolutionary history of ancient aquatic angiosperms. *Proceedings of the National Academy of Sciences* 88**:** 10119-10123.

Les D. H., E. L. Schneider, D. J. Padgett, P. S. Soltis, D. E. Soltis, & M. Zanis (1999). Phylogeny, classification and floral evolution of water lilies (Nymphaeaceae; Nymphaeales): A synthesis of non-molecular, *rbcL*, *matK*, and 18S rDNA data. *Systematic Botany* 24**:** 28-46.

Levinson G. & G. Gutman (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4**:** 203-221.

Li H.-L. (1955). Classification and phylogeny of Nymphaeaceae and allied families. *The American Midland Naturalist* 54**:** 33-41.

Liang H.-X. & K. W. Hilu (1996). Application of *matK* sequences to grass systematics. *Canadian Journal of Botany* 74**:** 125-134.

Lidgard S. & P. R. Crane (1990). Angiosperm diversification and Cretaceous floristic trends - a comparison of palynofloras and leaf macrofloras. *Paleobiology* 16**:** 77-93.

Linder H. P., C. R. Hardy, & F. Rutschmann (2005). Taxon sampling effects in molecular clock dating: An example from the African Restionaceae. *Molecular Phylogenetics and Evolution* 35**:** 569-582.

Lippok B., A. A. Gardine, P. S. Williamson, & S. S. Renner (2000). Pollination by flies, bees, and beetles of Nuphar ozarkana and N. advena (Nymphaeaceae). *American Journal of Botany* 87**:** 898-902.

Liu Y.-L., L.-M. Xu, X.-M. Ni, & J.-R. Zhao (2005). Phylogeny of the Nymphaeaceae inferred from ITS sequences. *Acta Phytotaxonomica Sinica* 43**:** 22-30.

Löhne C. & T. Borsch (2005). Molecular evolution and phylogenetic utility of the *petD* group II intron: A case study in basal angiosperms. *Molecular Biology and Evolution* 22**:** 317–332.

Löhne C., T. Borsch, & J. H. Wiersema (subm.). Phylogenetic analysis of Nymphaeales using fast-evolving and non-coding chloroplast markers. *Botanical Journal of the Linnean Society*.

Lutzoni F., P. Wagner, V. Reeb, & S. Zoller (2000). Integrating ambinguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. *Syst. Biol.* 49**:** 628-651.

Macphail M. K., A. D. Partridge, & E. M. Truswell (1999). Fossil pollen records of the problematical primitive angiosperm family Lactoridaceae in Australia. *Plant Systematics & Evolution* 214**:** 199-210.

Magallón S. (2004). Dating lineages: Molecular and paleontological approaches to the temporal framework of clades. *International Journal of Plant Sciences* 165**:** S7-S21.

Magallón S. & M. J. Sanderson (2002). Relationships among seed plants inferred from highly conserved genes: sorting conflicting phylogenetic signals among ancient lineages. *American Journal of Botany* 89**:** 1991-2006.

Magallón S. & M. J. Sanderson (2005). Angiosperm divergence times: The effect of genes, codon positions, and time constraints. *Evolution* 59**:** 1653-1670.

Mai D. H. (1970). Subtropische Elemente im europäischen Tertiär I. Die Gattungen *Gironniera, Sarcococca, Illicium, Evodia, Illex, Mastixia, Alangium, Symplocos* und *Rehderodendron*. *Paläontol. Abh., Abt. B.: Paläobot.* 3**:** 441–503.

Mai D. H. (1988). New Nymphaealean fossils from the Tertiary of central Europe. *Tertiary Res.* 9**:** 87-96.

Mai D. H. (1995). Tertiäre Vegetationsgeschichte Europas. Fischer, Jena, Germany.

Marhold K., J. Lihova, M. Perny, & W. Bleeker (2004). Comparative ITS and AFLP analysis of diploid *Cardamine* (Brassicaceae) taxa from closely related polyploid complexes. *Annals of Botany* 93**:** 507-520.

Martin W., A. Gierl, & H. Saedler (1989). Molecular evidence for pre-Cretaceous angiosperm origins. *Nature* 339**:** 46-48.

Martin W., D. Lydiate, H. Brinkann, G. Forkmann, H. Saedler, & R. Cerff (1993). Molecular phylogenies in angiosperm evolution. *Molecular Biology and Evolution* 10**:** 140-162.

Masters M. T. (1902). The arrangement of the air canals in the stalks of Nymphaeas. *Journal of the Royal Horticultural Society* 26**:** 840-843.

Mathews S. & M. J. Donoghue (2000). Basal angiosperm phylogeny inferred from duplicate phytochromes A and C. *International Journal of Plant Sciences* 161**:** S41-S55.

McGlone M. S. (2005). Goodbye Gondwana. *Journal of Biogeography* 32**:** 739-740.

McLoughlin S. (2001). The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany* 49**:** 271-300.

Meeuse A. D. J. (1987). All about angiosperms. Edburon, Delft.

Michel F. & J.-L. Ferat (1995). Structure and activities of group II introns. *Annual Review of Biochemistry* 64**:** 435-461.

Michel F., K. Umesono, & H. Ozeki (1989). Comparative and functional anatomy of group II catalytic introns - a review. *Gene* 82**:** 5-30.

Monde R.-A., J. C. Greene, & D. B. Stern (2000). Disruption of the *petB-petD* intergenic region on tobacco chloroplasts affects *petD* RNA accumulation and translation. *Molecular and General Genetics* 263**:** 610-618.

Mooers A. O. & D. Schluter (1999). Reconstruction ancestor states with maximum likelihood: support for one- and two-rate models. *Systematic Biology* 48**:** 623-633.

Morgenstern B. (1999). DIALIGN 2: improvement of the segment-to-segment approach to mutliple sequence alignment. *Bioinformatics* 15**:** 211-218.

Morley R. J. (1998). Palynological evidence for Tertiary plant dispersals in the SE Asian region in relation to plate tectonics and climate. *In:* Biogeography and geological evolution of SE Asia (R. Hall, and J. D. Holloway, Eds.), Backhuys, Leiden.

Morley R. J. (2000). Origin and evolution of tropical rain forests. Wiley Chichester.

Morley R. J. (2003). Interplate dispersal routes for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics* 6**:** 5-20.

Morrone J. J. (2005). Cladistic biogeography: identity and place. *Journal of Biogeography* 32**:** 1281-1284.

Moseley M. F. (1958). Morphological studies of the Nymphaeaceae. I. The nature of the stamens. *Phytomorphology* 8**:** 1-29.

Moseley M. F. (1961). Morphological studies of the Nymphaeaceae. II. The flowers of *Nymphaea*. *Botanical Gazette* 122**:** 233-259.

Moseley M. F. (1965). Morphological studies of Nymphaeaceae. III. The floral anatomy of *Nuphar*. *Phytomorphology* 15**:** 54-84.

Moseley M. F., E. L. Schneider, & P. S. Williamson (1993). Phylogenetic interpretations from selected floral vasculature characters in the Nymphaeaceae sensu lato. *Aquatic Botany* 44**:** 325-342.

Muller J. (1970). Palynological evidence on early differentiation of angiosperms. *Biological Reviews of the Cambridge Philosophical Society* 45**:** 415-450.

Müller J. (1970). Description of pollen grains of Ondinea purpurea den Hartog. *Blumea* 18**:** 416-417.

Müller J. & K. Müller (2003). QuickAlign: A new alignment editor. *Plant Molecular Biology Reporter* 21**:** 5.

Müller K. (2004). PRAP - computation of Bremer support for large data sets. *Molecular Phylogenetics and Evolution* 31**:** 780-782.

Müller K. (2005). SeqState: Primer design and sequence statistics for phylogenetic DNA datasets. *Applied Bioinformatics* 4**:** 65-69.

Müller K. (2006). Incorporating information from length-mutational events into phylogenetic analysis. *Molecular Phylogenetics and Evolution* 38**:** 667-676.

Müller K. & T. Borsch (2005). Phylogenetics of Amaranthaceae based on *matK/trnK* sequence data - Evidence from parsimony, likelihood, and Bayesian analyses. *Annals of the Missouri Botanical Garden* 92**:** 66-102.

Müller K., T. Borsch, L. Legendre, S. Porembski, I. Theisen, & W. Barthlott (2004). Evolution of carnivory in Lentibulariaceae and the Lamiales. *Plant Biology* 6**:** 477-490.

Müller K., K. W. Hilu, & T. Borsch (in prep.). Phylogenetic utility of rapidly evolving DNA on high taxonomical levels: comparing three cpDNA datasets of basal angiosperms.

Muñoz J., Á. M. Felicísimo, F. Cabezas, A. R. Burgaz, & I. Martínez (2004). Wind as a long-distance dispersal vehicle in the Southern Hemisphere. *Science* 304**:** 1144-1147.

Mutke J., G. Kier, G. A. Krupnick, & W. Barthlott (2005). Terrestrial Plant diversity. *In:* Plant Conservation: A Natural History Approach (G. J. K. Krupnick, Ed.), pp. 15-25, University of Chicago Press, Chicago.

Nakai T. (1943). *Ordines, familiae, tribi, genera, sectiones, species, varietates, formae et combinationes novae a* Prof. Nakai-Takenosin *adhuc ut novis edita.* Appendix: *Quaestiones characterium naturalium plantarum vel Extractus ex praelectionibus pro aluminis botanicis Universitatis Imperialis Tokyoensis per annos* 1926-1941., Tokyo.

Nandi O. I., M. W. Chase, & P. K. Endress (1998). A combined cladistic analysis of angiosperms using *rbcL* and non-molecular data sets. *Annals of the Missouri Botanical Garden* 85**:** 137-212.

Near J. T. & M. J. Sanderson (2004). Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. One contribution of 16 to a Discussion Meeting Issue 'Plant phylogeny and the origin of major biomes'. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 359**:** 1477 - 1483.

Near T. J. & B. P. Keck (2005). Dispersal, vicariance and timing of diversification in *Nothonotus* darters. *Molecular Ecology* 14**:** 3485-3496.

Neinhuis C., S. Wanke, K. W. Hilu, K. Müller, & T. Borsch (2005). Phylogeny of Aristolochiaceae based on parsimony, likelihood, and Bayesian analyses of *trnL-trnF* sequences. *Plant Systematics & Evolution* 250**:** 7-26.

Nepokroeff M., K. J. Sytsma, W. L. Wagner, & E. A. Zimmer (2003). Reconstructing ancestral patterns of colonization and dispersal in the Hawaiian understorey tree genus Psychotria (Rubiaceae): a comparison of parsimony and likelihood approaches. *Systematic Biology* 52**:** 820-838.

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

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

Nixon K. C. (1999). The Parsimony Ratchet, a new method for rapid parsimony analysis. *Cladistics* 15**:** 407-414.

Nixon K. C. (2002). WinClada ver. 1.00.08. Published by the author, Ithaca, NY.

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

ODSN (1999). Ocean Drilling Stratigraphic Network. http://www.odsn.de/odsn/services/paleomap/paleomap.html (January 10, 2006).

Oh I.-C., T. Denk, & E. M. Friis (2003). Evolution of *Illicium* (Illiciaceae): Mapping morphological characters on the molecular tree. *Plant Systematics & Evolution* 240**:** 175-209.

Okada H. & M. Tamura (1981). Karyomorphological study of the Nymphaeales. *The Journal of Japanese Botany* 56**:** 367-374.

Okada Y. & T. Otaya (1930). Study of *Euryale ferox* Salisb. VI. Cleistogamous versus chasmogamous flowers. *The Botanical Magazin, Tokyo* 44**:** 369-373.

Orgaard M. (1991). The genus *Cabomba* (Cabombaceae) - a taxonomic study. *Nordic Journal of Botany* 11**:** 179-203.

Padgett D. J., D. H. Les, & G. E. Crow (1999). Phylogenetic relationships in *Nuphar* (Nymphaeaceae): Evidence from morphology, chloroplast DNA, and nuclear ribosomal DNA. *American Journal of Botany* 86**:** 1316-1324.

Pagel M., A. Meade, & D. Barker (2004). Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* 53**:** 758-766.

Parenti L. R. & C. J. Humphries (2004). Historical biogeography, the natural science. *Taxon* 53**:** 899-903.

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

Pennington R. T. & C. W. Dick (2004). The role of immigrants in the assembly of the South American rainforest tree flora. One contribution of 16 to a Discussion Meeting Issue 'Plant phylogeny and the origin of major biomes'. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 359**:** 1611 - 1622.

Percy D. M., R. D. M. Page, & Q. C. B. Cronk (2004). Plant-insect interactions: double-dating associated insect and plant lineages reveals asynchronoous radiations. *Systematic Biology* 53**:** 120-127.

Pindall J. L., S. C. Cande, W. C. Pitman, D. B. Rowley, J. F. Dewey, J. LaBrecque, & W. Haxby (1988). A plate-kinematic framework for models of Caribbean evolution. *Tectonophysics* 155**:** 121–138.

Podoplelova Y. & G. Ryzhakov (2005). Phylogenetic analysis of the order Nymphaeales based on the nucleotide sequences of the chloroplast ITS2-4 region. *Plant Science* 169**:** 606-611.

Posada D. & K. A. Crandall (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* 14**:** 817-818.

Prance G. T. (1980). A note on the pollinaton of *Nymphaea amazonum* Mart. & Zucc. (Nymphaeaceae). *Brittonia* 32**:** 505-507.

Prance G. T. & A. B. Anderson (1976). Studies on the floral biology of neotropical Nymphaeaceae. 3. *Acta Amazonica* 6**:** 163-170.

Prance G. T. & G. R. Arias (1975). A study of the floral biology of *Victoria amazonica* (Poepp.) Sowerby (Nymphaeaceae). *Acta Amazonica* 5**:** 109-129.

Prance G. T., H. Beentje, J. Dransfield, & R. Johns (2000). The tropical flora remains undercollected. *Annals of the Missouri Botanical Garden* 87**:** 67-71.

Proctor V. W. (1968). Long-distance dispersal of seeds by retention in digestive tract of birds. *Science* 160**:** 321-322.

Pruchner D., S. Beckert, H. Muhle, & V. Knoop (2002). Divergent intron conservation in the mitochondrial nad2 gene: Signatures for the three bryophyte classes (mosses, liverworts, and hornworts) and the lycophytes. *Journal of Molecular Evolution* 55**:** 265-271.

Qiu Y.-L., Y. Cho, J. C. Cox, & J. D. Palmer (1998). The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* 394**:** 671-674.

Qiu Y.-L., O. Dombrovska, J. Lee, L. Li, B. A. Whitlock, F. Bernasconi-Quadroni, J. S. Rest, C. C. Davis, T. Borsch, K. W. Hilu, S. S. Renner, D. E. Soltis, P. S. Soltis, M. J. Zanis,

J. J. Cannone, R. R. Gutell, M. Powell, V. Savolainen, L. W. Chatrou, & M. W. Chase (2005). Phylogenetic analyses of basal angiosperms based on nine chloroplast, mitochondrial, and nuclear genes. *International Journal of Plant Sciences* 166**:** 815-842.

Qiu Y.-L., J. Lee, B. A. Whitlock, F. Bernasconi-Quadroni, & O. Dombrovska (2001). Was the ANITA rooting of the angiosperm phylogeny affected by long-branch attraction? *Molecular Biology and Evolution* 18**:** 1745-1753.

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

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

Quandt D. (in press). Universal primers for amplification of the trnK/matK-psbA region in land plants. *Anales del Jardín Botánico de Madrid*.

Quandt D., K. Müller, & S. Huttunen (2003). Characterisation of the chloroplast DNA *psbT-H* region and the influence of dyad symmetrical elements on phylogenetic reconstructions. *Plant Biology* 5**:** 400-410.

Quandt D., K. Müller, M. Stech, J.-P. Frahm, W. Frey, K. W. Hilu, & T. Borsch (2004). Molecular evolution of the chloroplast trnL-F region in land plants. *In:* Molecular systematics of bryophytes (B. Goffinet, V. Hollowell, and R. Magill, Eds.), pp. 13-37, Missouri Botanical Garden Press.

Quinn C. J., R. A. Price, & P. A. Gadek (2002). Familial concepts and relationships in the conifers based on *rbcL* and *matK* sequence comparisons. *Kew Bulletin* 57**:** 513-531.

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

Rambaut A. & M. Charleston (2002). TreeEdit: Phylogenetic Tree Editor v1.0 alpha 10, Computer program distributed by the author, available from http://evolve.zoo.ox.ac.uk/software.html?id=TreeEdit.

Ramshaw J. A. M., D. L. Richardson, B. T. Meatyard, R. H. Brown, M. Richardson, E. W. Thompson, & D. Boulter (1972). The time of origin of hte flowering plants determined by using amino acid sequence data of cytochrome c. *New Phytologist* 71**:** 773-779.

Ratcliffe B. C., D. M. Smith, & D. Erwin (2005). *Oryctoantiquus borealis*, new genus and species from the Eocene of Oregon, U.S.A., the world's oldest fossil dynastine and largest fossil scarabaeid (Coleoptera: Scarabaeidae: Dynastinae). *The Coleopterists Bulletin* 59**:** 127-135.

Raven P. H. & D. I. Axelrod (1972). Plate tectonics and Australasian paleobiogeography. *Science* 176**:** 1379-1386.

Ray J. (1682). Methodus plantarum nova. Faithorne and Kersey, London, UK.

Ray J. (1696). De variis plantarum methodis dissertatio brevis. Smith and Walford, London, UK.

Ree R. H., B. R. Moore, C. O. Webb, & M. J. Donoghue (2005). A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59**:** 2299-2311.

Reinink-Smith L. M. & E. B. Leopold (2005). Warm climate in the late Miocene of the south coast of Alaska and the occurrence of Podocarpaceae pollen. *Palynology* 29**:** 205-262.

Reisz R. R. & J. Müller (2004). Molecular timescales and the fossil record: a paleontological perspective. *Trends in Genetics* 20**:** 237-241.

Renner S. S. (2004a). Bayesian analysis of combined chloroplast loci, using multiple calibrations, supports the recent arrival of Melastomataceae in Africa and Madagascar. *American Journal of Botany* 91**:** 1427-1435.

Renner S. S. (2004b). Multiple Miocene Melastomataceae dispersal between Madagascar, Africa and India. One contribution of 16 to a Discussion Meeting Issue 'Plant phylogeny and the origin of major biomes'. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 359**:** 1485 - 1494.

Renner S. S. (2004c). Plant dipersal across the tropical atlantic by wind and sea currents. *International Journal of Plant Sciences* 165**:** S23-S33.

Renner S. S. (2004d). Variation in diversity among Laurales, early Cretaceous to present. *Biologiske skrifter* 55**:** 441-458.

Renner S. S. (2005). Relaxed molecular clocks for dating historical plant dispersal events. *Trends in Plant Science* 10**:** 550-558.

Renner S. S. & A. S. Chanderbali (2000). What is the relationship among Hernadiaceae, Lauraceae, and Monimiaceae, and why is this question so difficult to answer? *International Journal of Plant Sciences* 161**:** S109-S119.

Renner S. S., G. Clausing, & K. Meyer (2001). Historical biogeography of Melastomataceae: the roles of Tertiary migration and long-distance dispersal. *American Journal of Botany* 88**:** 1290-1300.

Renner S. S., D. B. Foreman, & D. Murray (2000). Timing transantarctic disjunctions in the Atherospermataceae (Laurales): Evidence from coding and noncoding chloroplast sequences. *Systematic Biology* 49**:** 579-591.

Retallack G. J. & D. L. Dilcher (1986). Cretaceous angiosperm invasion of North America. *Cretaceous Research 7***:** 227-252.

Richardson J. E., L. W. Chatrou, J. B. Mols, R. H. J. Erkens, & M. D. Pirie (2004). Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae. One contribution of 16 to a Discussion Meeting Issue 'Plant phylogeny and the origin of major biomes'. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 359**:** 1495 - 1508.

Richardson J. E., R. T. Pennington, T. D. Pennington, & P. M. Hollingsworth (2001). Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* 293**:** 2242-2245.

Rock C. D., A. Barkan, & W. C. Taylor (1987). The maize plastid *psbB-psbF-petB-petD* gene cluster: spliced and unspliced *petB* and *petD* RNAs encode alternative products. *Current Genetics* 12**:** 69-77.

Rodriguez-Trelles F., R. Tarrio, & F. J. Ayala (2002). A methodological bias toward overestimation of molecular evolutionary time scales. *Proceedings of the National Academy of Sciences* 99**:** 8112-8115.

Rohde K. (1992). Latitudinal gradients in species diversity: the search for the primary cause. *Oikos* 65**:** 514-527.

Ronquist F. (1994). Ancestral areas and parsimony. *Systematic Biology* 43**:** 267-274.

Ronquist F. (1996). DIVA, Version 1.1. Computer Program and Manual Available by Anonymous FTP from Uppsala University (ftp.uu.se or ftp.systbot.uu.se).

Ronquist F. (1997). Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology* 46**:** 195-203.

Ronquist F. (2004). Bayesian inference of character evolution. *Trends in Ecology and Evolution* 19**:** 475-481.

Ronquist F. & J. P. Huelsenbeck (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* (in press).

Rousset F., M. Pélandakis, & M. Solignac (1991). Evolution of compensatory substitutions through G-U intermediate state in *Drosophila* rRNA. *Proceedings of the National Academy of Sciences* 88**:** 10032-10036.

Rydin C. & M. Källersjö (2002). Taxon sampling and seed plant phylogeny. *Cladistics* 18**:** 484-513.

Rydin C., M. Källersjö, & E. M. Friis (2002). Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: Conflicting data, rooting problems, and the monophyly of conifers. *International Journal of Plant Sciences* 163**:** 197-214.

Salisbury R. A. (1806). Description of the natural order of Nymphaeaceae. *Annals of Botany (König & Sims)* 2**:** 68-76.

Samigullin T. K., W. F. Martin, A. V. Troitsky, & A. S. Antonov (1999). Molecular data from the chloroplast *rpoC1* gene suggest a deep and distinct dichotomy of contemporary spermatophytes into two monophyla: Gymnosperms (including Gnetales) and angiosperms. *Journal of Molecular Evolution* 49**:** 310-315.

Sampson F. B. (2000). Pollen diversity in some modern magnoliids. *International Journal of Plant Sciences* 161**:** S193-S210.

Sanders D. R. (1979). The ecology of Cabomba caroliniana. *In:* Weed Control Methods for Public Health Applications (E. O. Gangstad, Ed.), pp. 133-146, CRC Press, Boca Raton, Florida.

Sanderson M. J. (1997). A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* 14**:** 1218-1231.

Sanderson M. J. (1998). Estimating rate and time in molecular phylogenies: Beyond the molecular clock? *In:* Molecular systematics of plants II: DNA sequencing (D. E. Soltis, P. S. Soltis, and J. J. Doyle, Eds.), Kluwer Academic Press, Boston.

Sanderson M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molecular Biology and Evolution* 19**:** 101-109.

Sanderson M. J. & J. A. Doyle (2001). Sources of error and confidence intervals in estimating the age of angiosperms from *rbcL* and 18S rDNA data. *American Journal of Botany* 88**:** 1499-1516.

Sanderson M. J., J. L. Thorne, N. Wikstrom, & K. Bremer (2004). Molecular evidence on plant divergence times. *American Journal of Botany* 91**:** 1656-1665.

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

Sang T., D. J. Crawford, & T. F. Stuessy (1995). Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: Implications for biogeography and concerted evolution. *Proceedings of the National Academy of Science* 92**:** 6813-6817.

Sanmartín I., H. Enghoff, & F. Ronquist (2001). Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society* 73**:** 345-390.

Sanmartín I. & F. Ronquist (2004). Southern hemisphere bigeography inferred by event-based models: Plant versus animal patterns. *Systematic Biology* 53**:** 216-243.

Santamaría L. (2002). Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment. *Acta Oecologica* 23**:** 137-154.

Sauquet H., J. A. Doyle, T. Scharaschkin, T. Borsch, K. W. Hilu, L. W. Chatrou, & A. Le Thomas (2003). Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: implications for character evolution. *Botanical Journal of the Linnean Society* 142**:** 125-186.

Savolainen V. & M. W. Chase (2003). A decade of progress in plant molecular phylogenetics. *Trends in Genetics* 19**:** 717-724.

Savolainen V., M. W. Chase, C. M. Morton, D. E. Soltis, C. Bayer, M. F. Fay, A. De Bruijn, S. Sullivan, & Y.-L. Qiu (2000a). Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences. *Systematic Biology* 49**:** 306-362.

Savolainen V., M. F. Fay, D. C. Albach, A. Backlund, M. van der Bank, K. M. Cameron, S. A. Johnson, M. D. Lledó, J.-C. Pintaud, M. Powell, M. C. Sheahan, D. E. Soltis, P. S. Soltis, P. H. Weston, W. M. Whitten, K. J. Wurdack, & M. W. Chase (2000b). Phylogeny of the eudicots: a nearly complete familial analysis of the *rbcL* gene sequences. *Kew Bulletin* 55**:** 257-309.

Scharaschkin T. & J. A. Doyle (2005). Phylogeny and historical biogeography of *Anaxagorea* (Annonaceae) using morphology and non-coding chloroplast sequence data. *Systematic Botany* 30**:** 712-735.

Schenck H. (1886). Die Biologie der Wassergewächse. Cohen & Sohn (Fr. Cohen), Bonn.

Schmidt M. & H. A. W. Schneider-Poetsch (2002). The evolution of gymnosperms redrawn by Phytochrome genes: The Gnetatae appear at the base of the gymnosperms. *Journal of Molecular Evolution* 54**:** 715-724.

Schmitz-Linneweber C., R. M. Maier, J. P. Alcaraz, A. Cottet, R. G. Herrmann, & R. Mache (2001). The plastid chromosome of spinach (*Spinacia oleracea*): Complete nucleotide sequence and gene organization. *Plant Molecular Biology* 45**:** 307-315.

Schmucker T. (1932). Physiologische und ökologische Untersuchungen an Blüten tropischer *Nymphaea*-Arten. *Planta* 16**:** 376-412.

Schmucker T. (1933). Zur Blütenbiologie tropischer *Nymphaea*-Arten. II. Bor als entscheidender Faktor. *Planta* 18**:** 641-650.

Schneider E. L. (1976). The floral anatomy of *Victoria* Schomb. (Nymphaeaceae). *Botanical Journal of the Linnean Society* 72**:** 115-148.

Schneider E. L. (1982). Observations on the pollination biology of *Nymphaea gigantea* W.J. Hooker (Nymphaeaceae). *The Western Australian Naturalist* 15**:** 71-72.

Schneider E. L. (1983). Gross morphology and floral biology of *Ondinea purpurea* den Hartog. *Australian Journal of Botany* 31**:** 371-382.

Schneider E. L. & S. Carlquist (1995a). Vessel origins in Nymphaeaceae: *Euryale* and *Victoria*. *Botanical Journal of the Linnean Society* 119**:** 185-193.

Schneider E. L. & S. Carlquist (1995b). Vessels in the roots of *Barclaya rotundifolia* (Nymphaeaceae). *American Journal of Botany* 82**:** 1343-1349.

Schneider E. L. & S. Carlquist (1996). Vessels in *Brasenia* (Cabombaceae): new perspectives on vessel origin in primary xylem of angiosperms. *American Journal of Botany* 83**:** 1236-1240.

Schneider E. L., S. Carlquist, K. Beamer, & A. Kohn (1995). Vessels in Nymphaeaceae: *Nuphar*, *Nymphaea*, and *Ondinea*. *International Journal of Plant Sciences* 156**:** 857-862.

Schneider E. L. & E. G. Ford (1978). Morphological studies of the Nymphaeaceae. X. The seed of *Ondinea purpurea* den Hartog. *Bulletin of the Torrey Botanical Club* 195**:** 192-200.

Schneider E. L., M. F. Moseley, & P. S. Williamson (1984). The pollination biology of *Ondinea purpurea* (Nymphaeaceae). *In:* Proceedings of the Vth International Symposium on Pollination, pp. 231-235, INRA Publ. (Les Colloques de I'INRA, n-21).

Schneider E. L., S. C. Tucker, & P. S. Williamson (2003). Floral development in the Nymphaeales. *International Journal of Plant Sciences* 164**:** S279-S292.

Schneider E. L. & P. S. Williamson (1993). Nymphaeaceae. *In:* The families and genera of vascular plants II (K. Kubitzki, J. G. Rohwer, and V. Bittrich, Eds.), pp. 486-493, Springer, Berlin.

Schneider H., E. Schuettpelz, K. M. Pryer, R. Cranfill, S. Magallón, & R. Lupia (2004). Ferns diversified in the shadow of angiosperms. *Nature* 428**:** 553-557.

Schönenberger J., A. A. Anderberg, & K. J. Sytsma (2005). Molecular phylogenetics and patterns of floral evolution in the Ericales. *International Journal of Plant Sciences* 166**:** 265-288.

Scotland R. W. & A. H. Wortley (2003). How many species of seed plants are there? *Taxon* 52**:** 101-104.

Sculthorpe C. D. (1967). The biology of aquatic vascular plants. Edward Arnold, London, U.K.

Shaw A. J., C. J. Cox, B. Goffinet, W. R. Buck, & S. B. Boles (2003). Phylogenetic evidence of a rapid radiation of pleurocarpous mosses (Bryophyta). *Evolution* 57**:** 2226-2241.

Shaw J., E. B. Lickey, J. T. Beck, S. B. Farmer, W. Liu, J. Miller, K. C. Siripun, C. T. Winder, E. E. Schilling, & R. L. Small (2005). The tortoise and the hare II: relative utility of 21

noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92**:** 142-166.
Simmons M. P. & H. Ochoterena (2000). Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49**:** 369-381.
Simmons M. P., H. Ochoterena, & T. G. Carr (2001). Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. *Systematic Biology* 50**:** 454-462.
Simmons M. P., K. M. Pickett, & M. Miya (2004). How meaningful are Bayesian support values? *Molecular Biology and Evolution* 21**:** 188-199.
Simon J. P. (1970). Comparative serology of the order Nymphaeales. I. Preliminary survey on the relationships of Nelumbo. *Aliso 7***:** 243-261.
Sims H. J., P. S. Herendeen, R. Lupia, R. A. Christopher, & P. R. Crane (1999). Fossil flowers with Normapolles pollen from the Upper Cretaceous of southeastern North America. *Review of Palaeobotany and Palynology* 106**:** 131-151.
Small R. L., R. C. Cronn, & J. F. Wendel (2004). L.A.S. Johnson Review No. 2. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17**:** 145-170.
Smith A. G., D. G. Smith, & B. M. Funnell (1994). Atlas of Mesozoic and Cenozoic coastlines. Cambridge University Press, Cambridge / UK.
Smits A. J. M., R. Van Ruremonde, & G. Van Der Velde (1989). Seed dispersal of three nymphaeid macrophytes. *Aquatic Botany* 35**:** 167-180.
Smyth D. R. (2005). Morphogenesis of flowers - Our evolving view. *The Plant Cell* 17**:** 330-341.
Soltis D. E. & P. S. Soltis (1998). Choosing an approach and an appropriate gene for phylogenetic analysis. *In:* Molecular systematics of plants II: DNA sequencing (D. E. Soltis, P. S. Soltis, and J. J. Doyle, Eds.), pp. 1-42, Kluwer Academic Publishers, Boston.
Soltis D. E. & P. S. Soltis (2004a). *Amborella* not a "basal angiosperm"? Not so fast. *American Journal of Botany* 91**:** 997.
Soltis D. E., P. S. Soltis, M. D. Bennett, & I. J. Leitch (2003). Evolution of genome size in the angiosperms. *American Journal of Botany* 90**:** 1596-1603.
Soltis D. E., P. S. Soltis, M. W. Chase, M. E. Mort, D. C. Albach, M. Zanis, V. Savolainen, W. H. Hahn, S. B. Hoot, M. F. Fay, M. Axtell, S. M. Swensen, L. M. Prince, J. W. Kress, K. C. Nixon, & J. S. Farris (2000). Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133**:** 381-461.
Soltis D. E., P. S. Soltis, D. L. Nickrent, L. A. Johnson, W. J. Hahn, S. B. Hoot, J. A. Sweere, R. K. Kuzoff, K. A. Kron, M. W. Chase, S. M. Swensen, E. A. Zimmer, S.-M. Chaw, L. J. Gillespie, J. W. Kress, & K. J. Sytsma (1997). Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Annals of the Missouri Botanical Garden* 84**:** 1-49.
Soltis D. E., P. S. Soltis, & M. Zanis (2002a). Phylogeny of seed plants based on evidence from eight genes. *American Journal of Botany* 89**:** 1670-1681.
Soltis P. S. & D. E. Soltis (2004b). The origin and diversification of angiosperms. *American Journal of Botany* 91**:** 1614-1626.
Soltis P. S., D. E. Soltis, & M. W. Chase (1999a). Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402**:** 402-404.
Soltis P. S., D. E. Soltis, V. Savolainen, P. R. Crane, & T. G. Barraclough (2002b). Rate heterogeneity among lineages of tracheophytes: Integration of molecular and fossil data and evidence for molecular living fossils. *Proceedings of the National Academy of Sciences* 99**:** 4430-4435.
Soltis P. S., D. E. Soltis, P. G. Wolf, D. L. Nickrent, S.-M. Chaw, & R. L. Chapman (1999b). The phylogeny of land plants inferred from 18S rDNA sequences: Pushing the limits of rDNA signal? *Molecular Biology and Evolution* 16**:** 1774-1784.
Sorenson M. D. (1999). TreeRot, version 2., Boston University, Boston, MA.
Springer M. S. (1995). Molecular clocks and the incompleteness of the fossil record. *Journal of Molecular Evolution* 41**:** 531-538.

Stebbins L. G. (1974). Flowering Plants: Evolution above the Species Level. Belknap Press of Harvard University, Cambridge.

Steele K. P. & R. Vilgalys (1994). Phylogenetic analyses of Polemoniaceae using nucleotide-sequences of the plastid gene *matK*. *Systematic Botany* 19**:** 126-142.

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

Stefanovic S., D. W. Rice, & J. D. Palmer (2004). Long branch attraction, taxon sampling, and the earliest angiosperms: *Amborella* or monocots? *BMC Evolutionary Biology* 4**:** 35.

Suh Y., L. B. Thien, H. E. Reeve, & E. A. Zimmer (1993). Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. *American Journal of Botany* 80**:** 1042-1055.

Sun G., D. L. Dilcher, S. Zheng, & Z. Zhou (1998). In Search of the First Flower: A Jurassic Angiosperm, Archaefructus, from Northeast China. *Science* 282**:** 1692-1695.

Sun G., Q. Ji, D. L. Dilcher, S. Zheng, K. C. Nixon, & X. Wang (2002). Archaefructaceae, a New Basal Angiosperm Family. *Science* 296**:** 899-904.

Swisher C. C., Y.-Q. Wang, X.-L. Wang, X.-X. Wang, & Y. Wang (1999). Cretaceous age for the feathered dinosaurs of Liaoning, China. *Nature* 400**:** 58-61.

Swofford D. L. (2002). PAUP* Phylogenetic analysis using parsimony (*and other methods) v. 4.0b10, Sinauer Associates, Sunderland, Massachusetts.

Sytsma K. J., A. Litt, M. L. Zjhra, C. Pires, M. Nepokroeff, E. Conti, J. Walker, & P. G. Wilson (2004). Clades, clocks, and continents: Historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the southern hemisphere. *International Journal of Plant Sciences* 165**:** S85-S105.

Taberlet P., L. Gielly, G. Pautou, & J. Bouvet (1991). Universal primers for amplification of 3 noncoding regions of chloroplast DNA. *Plant Molecular Biology* 17**:** 1105-1109.

Takhtajan A. L. (1980). Outline of the classification of flowering plants (Magnoliophyta). *The Botanical Review* 46**:** 225-359.

Takhtajan A. L. (1997). Diversity and classification of flowering plants. Columbia University Press, New York, NY.

Tamura M. (1982). Relationship of *Barclaya* and classification of Nymphaeales. *Acta Phytotaxonomica et Geobotanica* 33**:** 336-345.

Tanaka M., J. Obokata, J. Chunwongse, J. Shinozaki, & M. Suguira (1987). Rapid splicing and stepwise processing of a transcript from the *psbB* operon in tobacco chloroplasts: Determination of the intron sites in *petB* and *petD*. *Molecular and General Genetics* 209**:** 427-431.

Tang W. & G. Eisenbrand (1992). Chinese drugs of plant origin: chemistry, pharmacology, and use in traditional and modern medicine Springer, Berlin.

Tarver D. P. & D. R. Sanders (1977). Selected life cycle features of fanwort. *Journal of Aquatic Plant Management* 15**:** 18-22.

Taylor D. W. & L. J. Hickey (1990). An Aptian plant with attached leaves and flowers: implications for angiosperm origin. *Science* 247**:** 702-704.

Taylor D. W. & L. J. Hickey (1992). Phylogenetic evidence for the herbaceous origin of angiosperms. *Plant Systematics and Evolution* 180**:** 137-156.

Taylor D. W. & L. J. Hickey (1996). Evidence for and implications of an herbaceous origin for angiosperms. *In:* Flowering plant origin, evolution and phylogeny (D. W. Taylor, and L. J. Hickey, Eds.), Chapman & Hall, New York.

Thien L. B., H. Azuma, & S. Kawano (2000). New perspectives on the pollination of basal angiosperms. *International Journal of Plant Sciences* 161**:** S225-S235.

Thorne J. L. & H. Kishino (1992). Freeing phylogenies form artifacts of alignment. *Molecular Biology and Evolution* 9**:** 1148-1162.

Thorne J. L. & H. Kishino (2002). Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* 51**:** 689-702.

Thorne J. L., H. Kishino, & I. S. Painter (1998). Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution* 15**:** 1647-1657.
Thorne R. (1976). A phylogenetic classification of the Angiospermae. *Evol. Biol.* 9**:** 35-106.
Thorne R. F. (1992). Classification and Geography of Flowering Plants. *The Botanical Review* 58**:** 225-348.
Thorne R. F. (2000). The classification and geography of the flowering plants: Dicotyledons of the class Angiospermae. *The Botanical Review* 66**:** 441-647.
Thorne R. F. (2002). How many species of seed plants are there? *Taxon* 51**:** 511-512.
Tiffney B. H. (1985). The Eocene North Atlantic land bridge: Its importance in Tertiary and modern phytogeography of the northern hemisphere. *Journal of the Arnold Arboretum* 66**:** 243-273.
Tiffney B. H. & S. R. Manchester (2001). The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the northern hemisphere Tertiary. *International Journal of Plant Sciences* 162**:** S3-S17.
Tréca B. (1981). Régime alimentaire de la Sarcelle d'été (Anas querquedula L.) dans le delta du Sénégal. *L'Oiseau et la Revue Francaise d'Ornithologie* 51**:** 33-58.
Tsudzuki J., K. Nakashima, T. Tsudzuki, J. Hirasuka, M. Shibata, T. Wakasugi, & M. Sugiura (1992). Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ*, *trnK*, *psbA*, *trnI* and *trnH* and the absence of *rps16*. *Molecular and General Genetics* 232**:** 206-214.
Valla J. J. & D. R. Cirino (1972). Biología floral del irupé, *Victoria cruziana* D'Orb. (Nymphaeaceae). *Darwiniana* 17**:** 447-500.
Van Vierssen W. (1982). On the identity and autecology of *Zannichellia peltata* Bertol. in western Europe. *Aquatic Botany* 13**:** 367-383.
Vangerow S., T. Teerkorn, & V. Knoop (1999). Phylogenetic information in the mitochondrial nad5 gene of pteridophytes: RNA editing and intron sequences. *Plant Biology* 1**:** 235-243.
Vawter L. & W. M. Brown (1993). Rates and patterns of base change in the small subunit ribosomal RNA gene. *Genetics* 134**:** 597-608.
Velichkevich F. (1994). The microevolution aspect of biostratigraphy of the Belarussian Pleistocene deposits. *Lithosphere* 1**:** 74-80.
Vickulin S. V. (1999). Palaeogene leaf compressions of myrtaceous affinity from Pasekovo, Middle Russian Upland, southern European Russia. *Botanical Journal of the Linnean Society* 131**:** 65-98.
Vogel V. J. (1970). American Indian Medicine. University of Oklahoma Press, Norman, OK.
Vriesendorp B. & F. T. Bakker (2005). Reconstructing patterns of reticulate evolution in angiosperms: what can we do? *Taxon* 54**:** 593-604.
Wang X. Q., D. C. Tank, & T. Sang (2000). Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Molecular Biology and Evolution* 17**:** 773-781.
Webb S. D. (1991). Ecogeography and the great american interchange. *Paleobiology* 17**:** 266-280.
Weberbauer A. (1894). Beiträge zur Samenanatomie der Nymphaeaceen. *Engler´s Jahrbücher* 18**:** 213-258.
Wehr W. & S. R. Manchester (1996). Paleobotanical significance of Eocene flowers, fruits and seeds from Republic, Washington. *Washington Geology* 24**:** 25-27.
Weidlich W. H. (1976a). The organization of the vascular system in the stems of the Nymphaeaceae. I. *Nymphaea* subgenera *Castalia* and *Hydrocallis*. *American Journal of Botany* 63**:** 499-509.
Weidlich W. H. (1976b). The organization of the vascular system in the stems of the Nymphaeaceae. II. *Nymphaea* subgenera *Anecphya*, *Lotos*, and *Brachyceras*. *American Journal of Botany* 63**:** 1365-1379.
Weidlich W. H. (1980). The organization of the vascular system in the stems of the Nymphaeaceae. III. *Victoria* and *Euryale*. *American Journal of Botany* 67**:** 790-803.

Welch J. J. & L. Bromham (2005). Molecular dating when rates vary. *Trends in Ecology and Evolution* 20**:** 320-327.

Westhoff P. & R. G. Herrmann (1988). Complex RNA maturation in the chloroplast: The *psbB* operon from spinach. *European Journal of Biochemistry* 171**:** 551-564.

Wettstein R. v. (1924). Handbuch der Systematischen Botanik. 3. Aufl. Franz Deuticke, Leipzig und Wien.

White T. J., T. Bruns, S. Lee, & J. Taylor (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In:* PCR protocols: A guide to methods and applications (D. G. M. Innis, J. Sninsky, and T. White, Eds.), pp. 315-322, Academic Press, San Diego, California.

Wiersema J. H. (1987). A monograph of *Nymphaea* subgenus *Hydrocallis* (Nymphaeaceae). *Systematic Botany Monographs* 16**:** 1-112.

Wiersema J. H. (1988). Reproductive biology of *Nymphaea* (Nymphaeaceae). *Annals of the Missouri Botanical Garden* 75**:** 795-804.

Wiffin T. (2002). Plant biogeography of the Asian-Australian region. *In:* Advances in Geoecology (P. Kershaw, D. Bruno, N. Tapper, D. Penny, and J. Brown, Eds.), pp. 60-82, Catena, Reiskirchen.

Wikström N., V. Savolainen, & M. W. Chase (2001). Evolution of the angiosperms: Calibrating the family tree. *Proceedings of the Royal Society of London Series B* 268**:** 2211-2220.

Williamson P. S. & M. F. Moseley (1989). Morphological studies of the Nymphaeaceae sensu lato. XVII. Floral anatomy of *Ondinea purpurea* subspecies *purpurea* (Nymphaeaceae). *American Journal of Botany* 76**:** 1779-1794.

Williamson P. S. & E. L. Schneider (1993). Cabombaceae. *In:* The families and genera of vascular plants II (K. Kubitzki, J. G. Rohwer, and V. Bittrich, Eds.), pp. 157-161, Springer, Berlin.

Williamson P. S. & E. L. Schneider (1994). Floral aspects of *Barclaya* (Nymphaeaceae): Pollination, ontogeny and structure. *Plant Systematics & Evolution [Suppl.]* 8**:** 159-173.

Williamson P. S., E. L. Schneider, & L. Malins (1989). Tuber and leaf structure of *Ondinea purpurea* den Hartog (Nymphaeaceae). *The Western Australian Naturalist* 18**:** 52-61.

Wing S. L. & L. D. Boucher (1998). Ecological aspects of the Cretaceous flowering plant radiation. *Annual Review of Earth and Planetary Sciences* 26**:** 379-421.

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

Wolfe J. A. (1975). Some aspects of plant geography of the Northern Hemisphere during Late Cretaceous and Tertiary. *Annals of the Missouri Botanical Garden* 62**:** 264-279.

Wolfe J. A. (1992). Climatic, floristic, and vegetational changes near the Eocene / Oligocene boundary in North America. *In:* Eocene-Oligocene climatic and biotic evolution (D. A. Prothero, and W. A. Berggren, Eds.), pp. 421-436, Princeton University Press, Princeton, NJ.

Wolfe K. H., M. Guoy, Y.-W. Yang, P. M. Sharp, & W.-S. Li (1989). Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proceedings of the National Academy of Sciences* 86**:** 6201-6205.

Wolfe K. H., C. W. Morden, & J. D. Palmer (1992). Function and evolution of a minimal plastid genome from a nonphotosynthetic parasite plant. *Proceedings of the National Academy of Sciences* 89**:** 10648-10652.

Woods K., K. W. Hilu, T. Borsch, & J. H. Wiersema (2005a). Pattern of variation and systematics of *Nymphaea odorata*: II. Sequence information from ITS and *trnL-trnF*. *Systematic Botany* 30**:** 481-493.

Woods K., K. W. Hilu, J. H. Wiersema, & T. Borsch (2005b). Pattern of variation and systematics of *Nymphaea odorata*: I. Evidence from morphology and Inter-Simple Sequence Repeats (ISSRs). *Systematic Botany* 30**:** 471-480.

Xiang Q.-Y. & D. E. Soltis (2001). Dispersal-vicariance analyses of intercontinental disjunctions: Historical biogeography implications for angiosperms in the northern hemisphere. *International Journal of Plant Sciences* 162**:** S29-S39.

Yoder A. D. & Z. Yang (2000). Estimation of primate speciation dates using local molecular clocks. *Molecular Biology and Evolution* 17**:** 1081-1090.

Yoo M.-J., C. D. Bell, P. S. Soltis, & D. E. Soltis (2005). Divergence times and historical biogeography of Nymphaeales. *Systematic Botany* 30**:** 693-704.

Zachos J., M. Pagani, L. Sloan, E. Thomas, & K. Billups (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292**:** 686-693.

Zander R. H. (2004). Minimal values for reliability of bootstrap and jackknife proportions, decay index, and Bayesian posterior probability. *Phyloinformatics* 1**:** 1-13.

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

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

Zavada M. S. & J. M. Benson (1987). Fossil evidence for the primitive angiosperm family Lactoridaceae. *American Journal of Botany* 74**:** 1590-1594.

Zhang D. & T. Sang (1999). Physical mapping of ribosomal RNA genes in peonies (*Paeonia*, Paeoniaceae) by fluorescent in situ hybridization: implications for phylogeny and concerted evolution. *American Journal of Botany* 86**:** 735-740.

Zhang L.-B. & S. S. Renner (2003). The deepest splits in Chloranthaceae as resolved by chloroplast sequences. *International Journal of Plant Sciences* 164**:** S383-S392.

Zhang W. (2000). Phylogeny of the grass family (Poaceae) from *rpl16* intron sequence data. *Molecular Phylogenetics and Evolution* 15**:** 135-146.

Zimmermann W. (1930). Die Phylogenie der Pflanzen: Ein Überblick über Tatsachen und Probleme. Fischer, Jena.

Zonneveld B. J. M., I. J. Leitch, & M. D. Bennett (2005). First nuclear DNA amounts in more than 300 angiosperms. *Annals of Botany* 96**:** 229-244.

Zuckerkandl E. & L. Pauling (1962). Molecular disease, evolution, and genetic heterogeneity. *In:* Horizons in biochemistry (M. Kasha, and B. Pullman, Eds.), pp. 189-225, Academic Press, New York.

List of Abbreviations

(in alphabetical order)

BI	=	Bayesian Inference
bp	=	base pairs
BrS	=	Bremer Support
BS	=	Bootstrap Support
CI	=	Consistency Index
cp	=	chloroplast
JK	=	Jackknife Support
Ma	=	million years
MP	=	Maximum Parsimony
mt	=	mitochondrial
nt	=	nucleotide
PBS	=	Partitioned Bremer Support
PP	=	Bayesian Posterior Probability
RC	=	Rescaled Consistency Index
rDNA	=	ribosomal DNA
SD	=	Standard Deviation
SE	=	Standard Error
SSR	=	simple sequence repeat

List of Figures

List of Tables

Appendix

Appendix 1 — List of length mutations that were coded as binary characters in the *petD* indel matrix (Chapter 2, page 23 ff.). Note that the three microsatellite regions are also listed, but since they were not coded in the indel matrix they do not have numbers. "Position" refers to the position of indels in the complete alignment (1803 characters including hotspots).

Indel No.	Position	Kind of length mutation	Length	Motif	Taxa
1	1-68	deletion	25	-	
s.n.	1-4	microsatellite		multiple A	Saururaceae, *Piper* sp.
2	7-13	SSR	7	TATAGAG	*Platanus*
3	14-48	deletion	10	-	*Impatiens*
4	16-31	insertion (may be 2x SSR)	16	TTGAATTTATAGAGAA	*Amborella*
5	32-40	SSR	9	TATAGCGAW	*Chloranthus off.*
6	45-49	deletion	5	-	*Aristolochia* and *Chloranthus brach.*
7	47-49	deletion	3	-	Gymnosperms
8	63	deletion	1	-	all angiosperms
10	70-79	insertion (may be SSR)	10	AGTATCAATG	*Araucaria*
11	80-96	SSR	17	AWGACAGATTATMGATA	*Chloranthus off.*
12	80-102	SSR	6	TTTGTA	*Chloranthus* (both species)
13	110-114	insertion	5	ATATA	*Nymphaea*
14	118	deletion	1	-	*Piper*
15	120-124	deletion	1	-	*Amborella*
16	120-125	deletion	2	-	*Ginkgo*
17	121-124	insertion	4	GATA	*Lactoris*
18	127-131	insertion	5	KAGAA	*Ceratophyllum*
19	132-136	insertion	5	TTGAA	*Tofieldia*
20	137-139	insertion	3	TCG	*Pinus*
21	142-143	deletion	2	-	*Victoria*
22	143	deletion	1	-	Nymphaeales
23	145-149	SSR	5	TTTYG	*Aristolochia*
24	150-153	insertion	4	TTTT	*Ginkgo*
25	150-154	insertion	1	T	*Ginkgo* and *Araucaria*
26	160-162	SSR	3	TAA	*Cabomba*
27	160-179	deletion	4	-	*Araucaria*
28	165-170	SSR	6	GTAGGA	*Nymphaea odorata*
29	173-179	SSR	7	GAGGAAC	*Trochodendron*
30	181-184	SSR	4	AACA	*Amborella*
31	185-189	SSR	5	AAACA	*Amborella*
s.n.	190-195	microsatellite		multiple A	Nymphaeales

Appendix 1 (continued)

Indel No.	Position	Kind of length mutation	Length	Motif	Taxa
32	197-208	deletion	1	-	*Ceratophyllum*
33	198-201	SSR	4	AAAT	*Brasenia*
34	202-208	insertion	7	AATAGAC	*Araucaria*
35	216-219	SSR	6	GTATTT	*Nymphaea nouchali* and *N. micrantha*
36	216-222	SSR	4	ATTT	*Drimys* and *Pseudowintera*
37	216-225	SSR	10	YATAGTATTT	*Cabomba*
38	246-265	SSR	20	CATTGCTACAAATATGGATT	*Nypa*
39	268-272	SSR	5	TRAAA	*Dicentra*
40	273-279	insertion	7	TCAAAAA	*Pinus*
41	282-285	inversion	4	AAAA -> TTTT	*Cabomba*
42	286-291	SSR	6	YGAAAW	*Amborella*
s.n.	292-302	microsatellite		multiple A	*Trochodendron, Dicentra, Aristolochia*
43	303-308	SSR	6	GAAAAA	*Illicium*
44	307-308	SSR	2	GAAA	*Brasenia*
45	309-311	SSR	3	TAA	*Dicentra*
46	314-321	SSR	8	TGGAAATA	*Araucaria*
47	322-326	insertion	5	TCCAT	*Pinus*
48	336-345	SSR	11	GYTCCTCGGA	*Araucaria*
49	347	deletion	1	-	eudicots
50	349	deletion	1	-	*Umbellularia*
51	358	SSR	1	T	*Piper*
52	367-378	insertion	12	multiple T	*Gingko*
53	384-387	SSR	4	AACT	eumagnoliids
54	387-389	deletion	3	-	*Saruma*
55	387-457	inversion	33	ATTCATGTTAAAAAATG ACTAGATAGAGTATGG	*Impatiens*
56	396	insertion	1	T	*Impatiens*
57	397	insertion	1	T	*Dicentra*
58	397-398	deletion	1	-	*Nelumbo*
59	397-402	insertion	5	TTGTA	Chloranthaceae, Canellales, Laurales, Magnoliales, eudicots, monocots
60	397-405	deletion	8	-	*Ceratophyllum*
61	409-413	SSR	5	AYATT	*Amborella*
62	414	SSR	1	A	*Aristolochia*
63	419	deletion	1	-	*Lactoris*

Appendix 1 (continued)

Indel No.	Position	Kind of length mutation	Length	Motif	Taxa
64	424-431	SSR	8	TTTGATAY	*Tofieldia* and *Orontium*
65	435-438	SSR	4	CAAA	*Pinus*
66	439-444	SSR	6	GAAYAR	*Cabomba* and *Brasenia*
67	439-450	SSR	12	ATTTAATACGAA	*Austrobaileya*
68	456	SSR	1	T	*Acorus*
69	462-467	SSR	6	TGAAGT	*Saruma*
70	478-482	SSR	5	CTCAG	*Araucaria*
71	484	deletion	1	-	all angiosperms
72	489-493	deletion	1	-	Austrobaileyales and *Aristolochia*
73	490-493	SSR	4	ARAT	*Piper*
74	534-538	SSR	5	ATYGG	*Ginkgo*
75	551-555	SSR	5	AGATA	*Ceratophyllum*
76	558-561	deletion	4	-	*Dicentra*
77	558-562	deletion	5	-	Nymphaeales
78	560-569	deletion	5	-	*Piper*
79	565-569	SSR	5	ATTYT	*Tofieldia*
80	573-576	SSR	4	TATC	*Aextoxicon*
81	577-580	SSR	4	TGSC	*Impatiens*
82	585-597	deletion	8	-	*Araucaria*
83	593-597	SSR	5	TTGKA	*Piper*
84	611	SSR	1	A	*Saruma*
85	616	deletion	1	-	*Myristica*
86	623	insertion	1	A	monocots
87	624	deletion	1	-	*Amborella*
88	686	insertion	1	T	*Ginkgo*
89	708	deletion	1	-	*Dicentra*
90	727-731	SSR	5	ATTAT	*Orontium*
91	735	deletion	1	-	Laurales
92	741-745	SSR	5	GAATC	*Orontium*
93	746	SSR	1	C	*Tofieldia* and *Orontium*
94	750-771	insertion	10	TCRTGCATGA	all angiosperms
95	752-753	SSR	2	TC	*Illicium* and *Schisandra*
96	757-761	SSR	5	TTRTG	*Umbellularia*
97	762-766	SSR	5	TCRTG	*Chloranthus brach.*
98	775-779	SSR	5	CAACA	*Dicentra*
99	796-799	insertion	4	ACGG	*Annonaceae*

Appendix 1 (continued)

Indel No.	Position	Kind of length mutation	Length	Motif	Taxa
100	801-818	deletion	8	-	Araucaria
101	801	deletion	1	-	Pinus
102	805-809	SSR	5	GTASA	Nelumbo
103	810-813	insertion	4	multiple A	Ceratophyllum
104	816	insertion	1	G	Ginkgo
105	819-822	SSR	4	ATAA	Impatiens
106	823-825	SSR	3	TAA	Nuphar
107	826-836	insertion (may be SSR)	11	AATAAAAKTAA	Umbellularia
108	831-836	SSR	5	AATAR	Nypa
109	841-843	SSR	3	TAA	Ginkgo
110	844-849	SSR	6	WAGTGA	Schisandra and Amborella
111	963-982	insertion (probably SSR)	20	AATACTATATACTATGGAAT	Piper
112	990	SSR	6	TAGTAT	Chloranthus brach.
113	990-995	SSR	5	AGTAW	Myristica and Chloranthus brachystachys
114	990-996	deletion	1	-	Magnoliales
115	996	deletion	1	-	Magnoliales
116	1000	deletion	1	-	Nymphaeales
117	1023-1027	deletion	4	-	Araucaria
118	1027	insertion	1	T	Amborella
119	1030-1032	SSR	3	CCG	Acorus
120	1033-1036	deletion	4	-	Nypa
122	1036	deletion	1	-	Pinus
123	1040-1044	insertion (may be SSR)	5	GTCTA	Impatiens
124	1058	deletion	1	-	Impatiens
125	1078	deletion	1	-	Nelumbo
126	1082	insertion	1	T	Chloranthus
127	1083	deletion	1	-	Aristolochia
128	1093-1101	insertion	5	CTNKA	all angiosperms
129	1095-1098	SSR	4	GACT	Myristica
130	1095-1099	deletion	1	-	Piper and Saururaceae
131	1107-1113	SSR	7	TATTAGT	Annonaceae
132	1125	deletion	1	-	Tofieldia
133	1130-1133	SSR	4	CTTT	Buxus
134	1136-1139	SSR	4	TTGTA	Austrobaileya

Appendix 1 (continued)

Indel No.	Position	Kind of length mutation	Length	Motif	Taxa
135	1142	deletion	1	-	Laurales and Magnoliales
136	1143-1169	deletion	8	-	Araucaria
137	1144-1146	insertion	3	GTASA	Impatiens
138	1147-1152	insertion (may be SSR)	6	AAGTGT	Buxus
139	1153-1156	SSR	4	WATG	Aristolochia
140	1153-1157	SSR	1	T	Houttuynia and Aristolochia
141	1165-1169	SSR	5	GAARA	Amborella
142	1170	deletion	1	-	Tofieldia
143	1173-1190	deletion	5	-	Lactoris
144	1176-1179	deletion	4	-	all angiosperms
145	1175-1194	deletion	5	-	Chloranthus
146	1180-1186	insertion	6	AAAACM	Nymphaeales
147	1186	insertion	1	T	Cabomba
148	1187-1189	SSR	3	ATA	Ceratophyllum
149	1193	SSR	1	A	Buxus
150	1196-1200	SSR	5	RTATT	Annonaceae and Saururus
151	1201	deletion	1	-	all angiosperms
152	1204-1209	deletion	1	-	Asimina
153	1205-1209	SSR	5	MTGTG	Nuphar
154	1210-1217	deletion	8	-	Nymphaeales (except Cabomba)
155	1211-1217	deletion	7	-	Cabomba
156	1220-1224	SSR	5	AAATA	Orontium
157	1235-1236	SSR	2	AY	Calycanthus
158	1235-1238	SSR	2	AA	Calycanthus and Myristica
159	1241	deletion	1	-	Tofieldia
160	1243-1254	SSR	12	TCACARGGCCTG	Annonaceae
161	1274-1278	SSR	5	AAGCA	Schisandra
162	1280	deletion	1	-	all angiosperms
163	1287-1292	deletion	6	-	Impatiens
164	1295-1297	deletion	3	-	Nymphaeales
165	1299-1304	SSR	6	AAAATT	Amborella
166	1307-1316	SSR (2x)	9	TTTGTTTTG	Dicentra
167	1346-1355	SSR	10	VCATTTTAYYT	Piper and Saururaceae
168	1359-1383	deletion	10	-	Ceratophyllum
169	1360-1365	SSR	6	TTGTAA	Nypa
170	1374-1375	insertion	2	multiple A	Araucaria

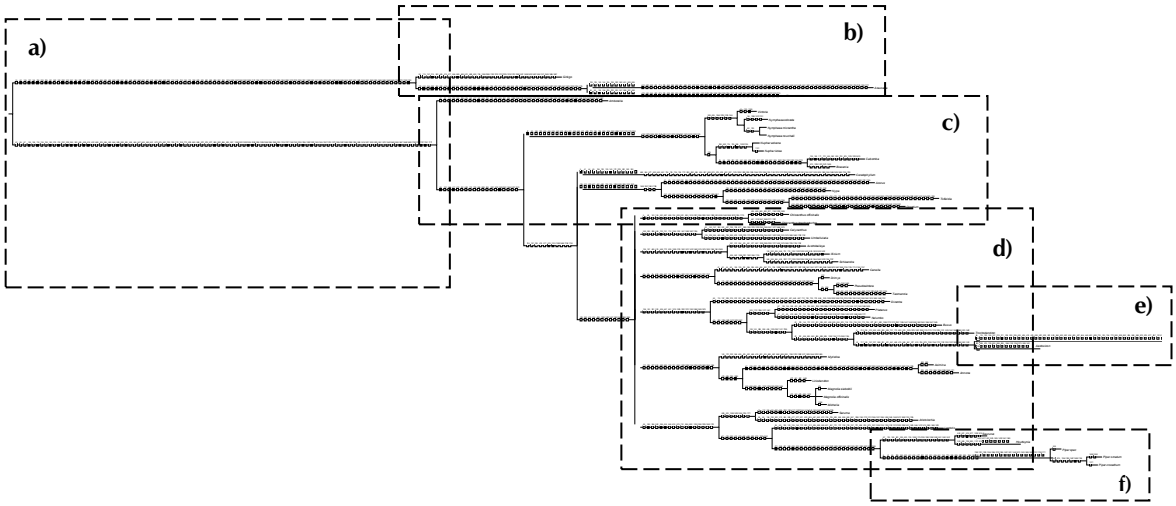
Appendix 1 (continued)

Indel No.	Position	Kind of length mutation	Length	Motif	Taxa
171	1376-1382	insertion	7	TTTAATT	Impatiens
172	1386-1406	deletion	12	-	Dicentra
173	1390-1405	deletion	7	-	Ceratophyllum
174	1390-1568	deletion	76	-	Nymphaeales
175	1390	deletion	1	-	Araucaria and Pinus
176	1393-1396	insertion (may be 2x SSR)	4	CAAA	Tofieldia
177	1393-1397	SSR	1	A	Tofieldia and Ginkgo
178	1399-1409	deletion	7	-	Houttuynia
179	1402-1404	SSR	3	TAA	Orontium
180	1402-1405	SSR	1	A	Orontium and Annonaceae
181	1411-1415	SSR	5	TTCTT	Ginkgo
182	1478-1499	deletion	6	-	Ceratophyllum
183	1478-1568	deletion		-	Araucaria
184	1479	deletion	1	-	Aristolochia
185	1483	SSR	1	A	Piper
186	1484-487	SSR	4	CTAT	Impatiens
187	1488-1492	SSR	5	YCTTA	Nelumbo and Chloranthus off.
188	1493-1498	SSR	6	ACCTTA	Chloranthus
189	1501-1502	SSR	2	CA	Buxus
190	1505-1508	SSR	4	CATA	Saruma
191	1512-1517	insertion	5	YCTWT	Impatiens and Aextoxicon
192	1517-1522	insertion	6	TGTCCC	Ginkgo
193	1523-1524	insertion (may be SSR)	2	TH	Piper and Saururaceae
194	1525-1526	deletion	2	-	Ceratophyllum
195	1527-1528	deletion	2	-	Saururus
196	1529-1530	SSR	2	RT	Piper
197	1529-1558	deletion		-	Ginkgo
198	1531-1533	insertion (may be SSR)	2	ATT	Winteraceae
199	1534-1536	insertion (may be SSR)	3	TRT	monocots
200	1541-1542	SSR	2	AT	Amborella
201	1543-1546	SSR	4	ATAC	Tofieldia
202	1547-1550	SSR	4	ATAC	Tofieldia
203	1551-1553	SSR	3	SAT	Impatiens
204	1551-1557	SSR	4	ATRT	Buxus and Impatiens

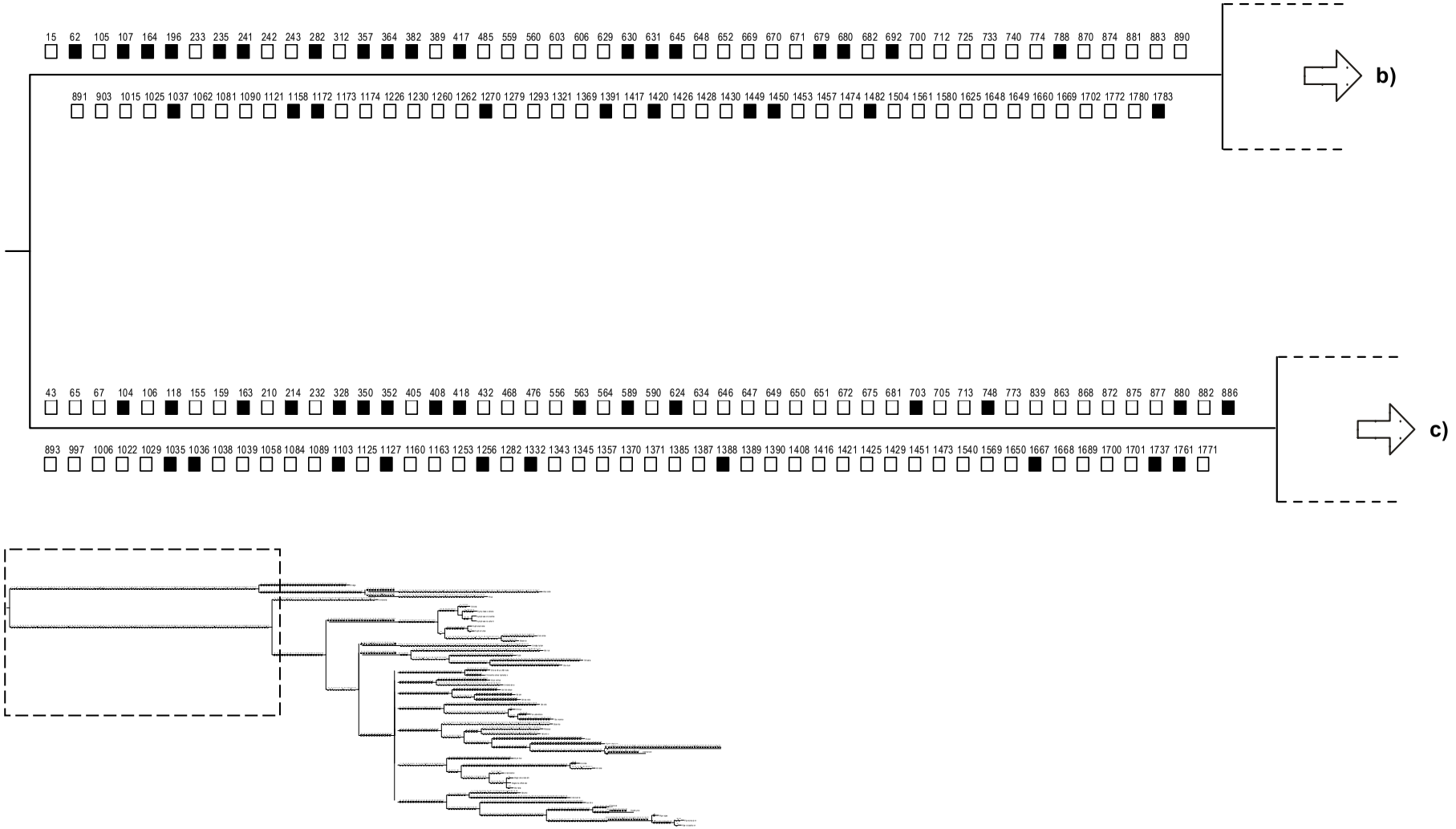
Appendix 1 (continued)

Indel No.	Position	Kind of length mutation	Length	Motif	Taxa
205	1541-1568	deletion	11	-	Pinus
206	1558-1571	deletion	14	-	Ceratophyllum
207	1563-1568	deletion	6	-	eumagnoliids
208	1565-1568	deletion	4	-	Ginkgo
209	1572	deletion	1	-	Ginkgo
210	1575-1576	deletion	2	-	Amborella
211	1575	deletion	1	-	Araucaria
212	1581-1606	deletion	2	-	Annonaceae
213	1582-1585	insertion	4	CATT	Ceratophyllum
214	1582-1609	deletion	5	-	Ginkgo
215	1582-1625	deletion	7	-	Araucaria
216	1583-1669	deletion	19	-	Pinus
217	1586-1593	insertion	8	TCTTATC	Tofieldia
218	1594-1603	insertion	10	TTTTTTATAG	Impatiens
219	1594-1604	SSR	1	A	Impatiens and Lactoris
220	1610-1614	SSR	5	TTTVA	Pseudowintera, Piper sp.
221	1615-1623	insertion	9	TYAAATTTA	Piper
222	1626-1627	insertion	2	GG	Acorus
223	1626-1629	insertion	2	GA	Acorus and Tofieldia
224	1627-1629	SSR	1	G	Ceratophyllum
225	1630-1633	SSR	4	TAWA	Orontium
226	1634-1643	insertion	10	TTATATGGTA	Saruma
227	1644-1645	SSR	2	TA	Annonaceae
228	1649-1656	deletion	2	-	Amborella
229	1650-1660	deletion	4	-	Ginkgo
230	1651-1656	SSR	6	ATGGRT	Annonaceae
231	1659	SSR	1	C	Austrobaileyales
232	1660	deletion	1	-	Houttuynia
233	1671-1674	SSR	4	KTAT	Canella
234	1675-1678	SSR	4	KTCW	Nypa
235	1679-1683	SSR	5	TTTSA	Amborella
236	1686-1688	deletion	3	-	Houttuynia
237	1690-1699	SSR	9	TTGGTTCTT	Impatiens
238	1702-1705	deletion	4	-	Amborella
239	1708-1711	SSR	4	TCTT	Annonaceae
240	1733	deletion	1	-	Pinus and Araucaria
241	1774-1777	SSR	4	AGAA	Umbellularia

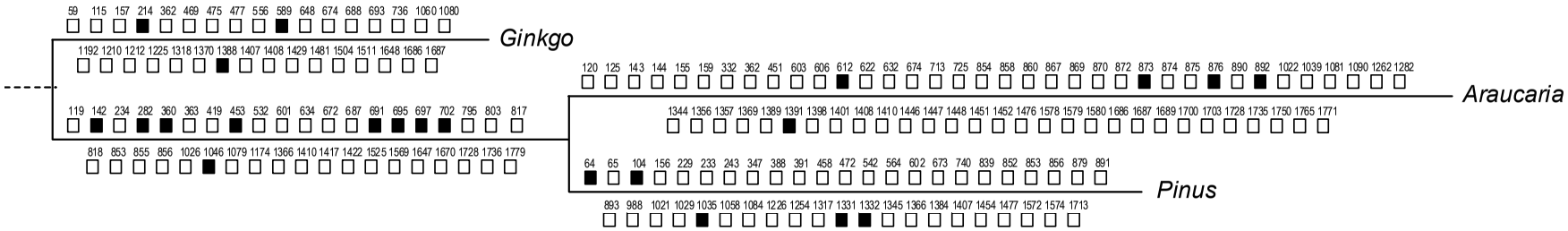
Appendix 2 — Strict consensus of 113 most parsimonious trees obtained from the intron + spacer matrix (i.e., substitutions only) of the *petD* data set (Chapter 2, page 23 ff.). Characters are optimized on the tree using Winclada 1.00.08 (Nixon 2002) assuming accelerated transformation (ACCTRAN). The tree is shown in subdivisions on the subsequent pages according to the following scheme:

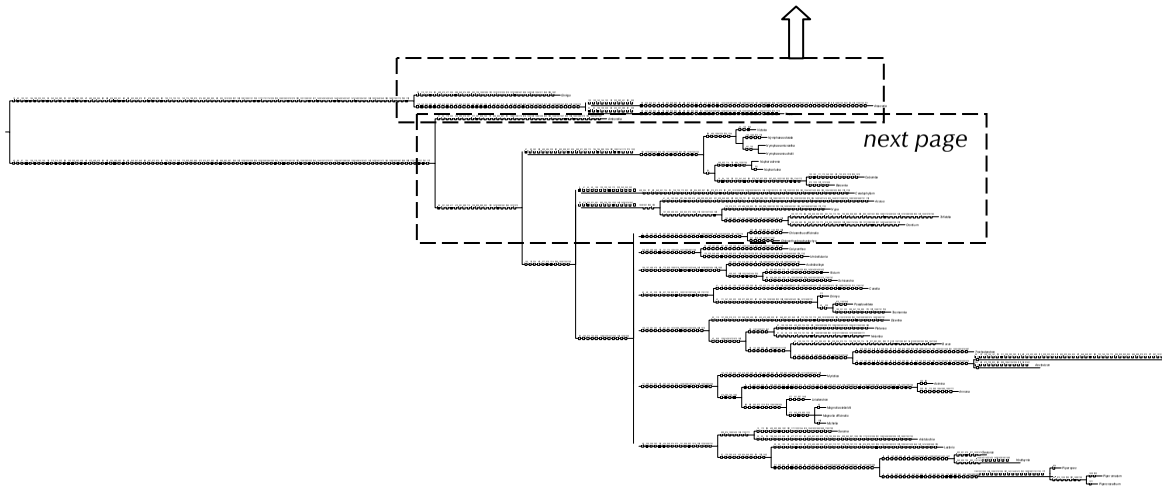


Appendix 2 — Section (a) of the strict consensus from the *petD* intron + spacer matrix (Chapter 2). Numbers refer to the position in the alignment. Parsimony informative characters are mapped onto the tree. Black boxes show states synapomorphic for the respective nodes, whereas white boxes indicate homoplastic characters.

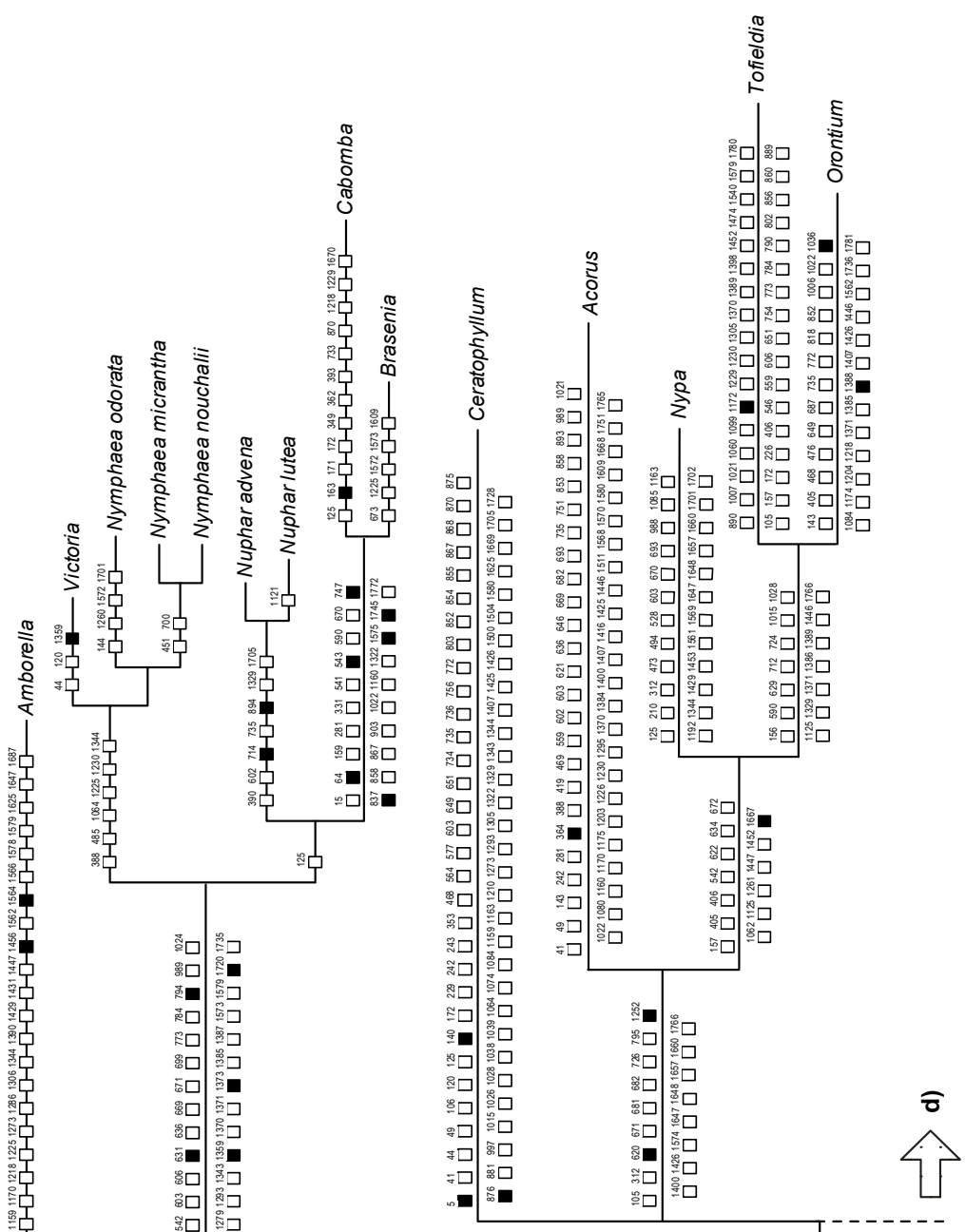


Appendix 2 — Section (b) of the strict consensus from the *petD* intron+spacer matrix. Numbers refer to the position in the alignment. Parsimony informative characters are mapped onto the tree. Black boxes show states synapomorphic for the respective nodes, whereas white boxes indicate homoplastic characters.

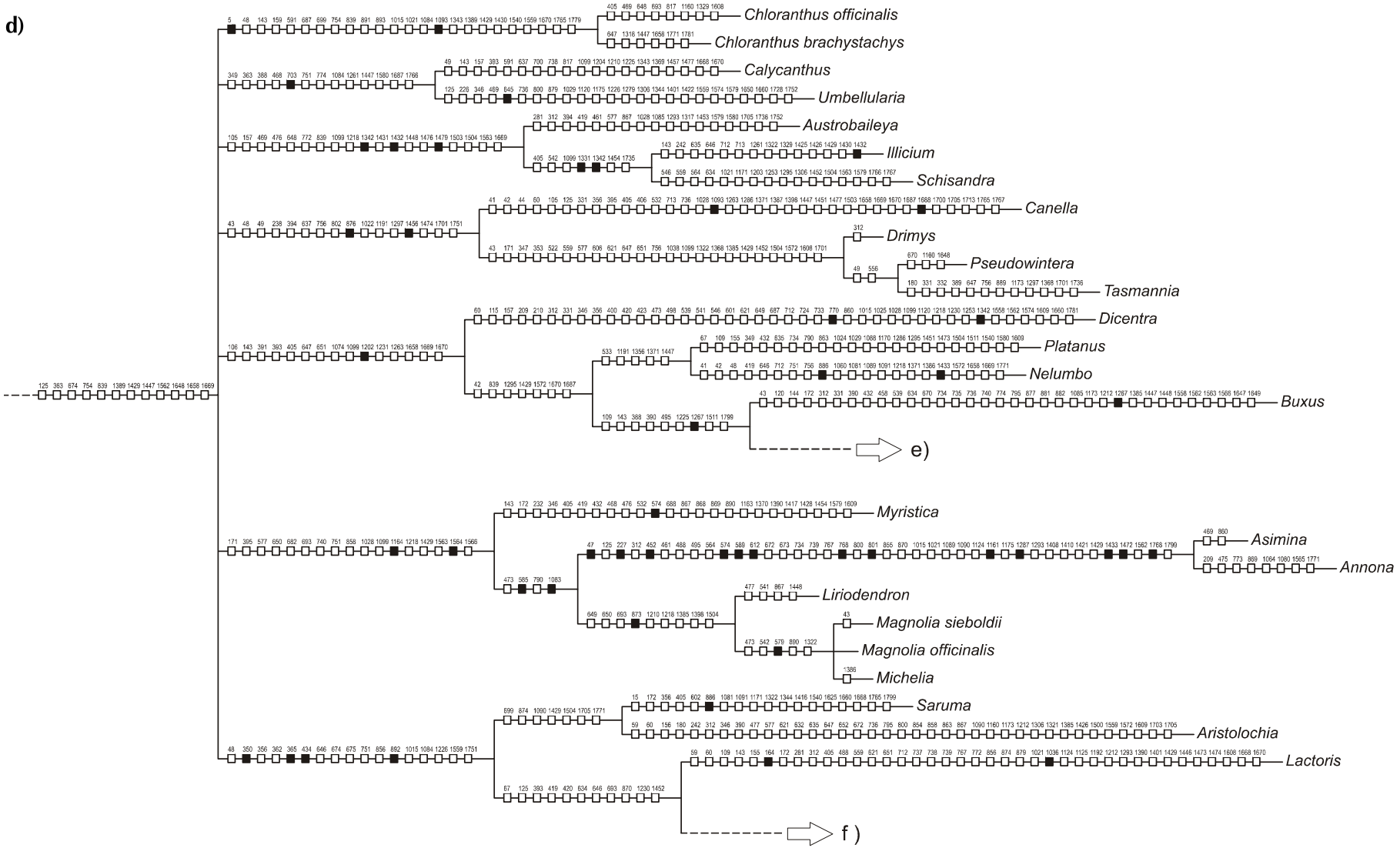




Appendix 2 (continued)

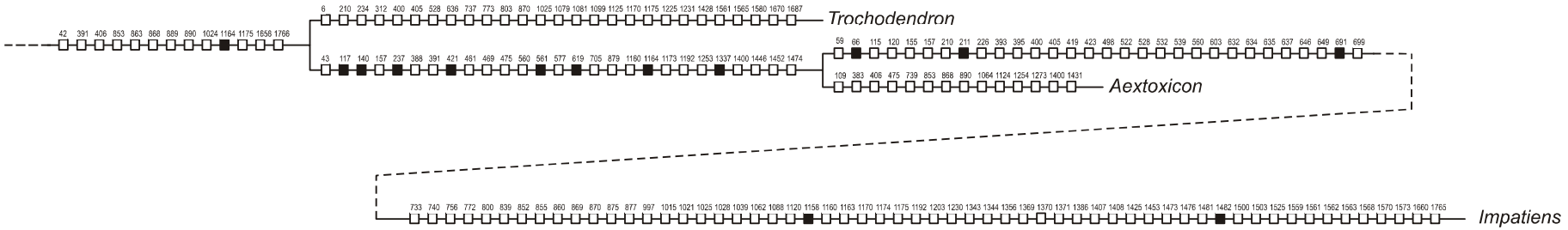


d)

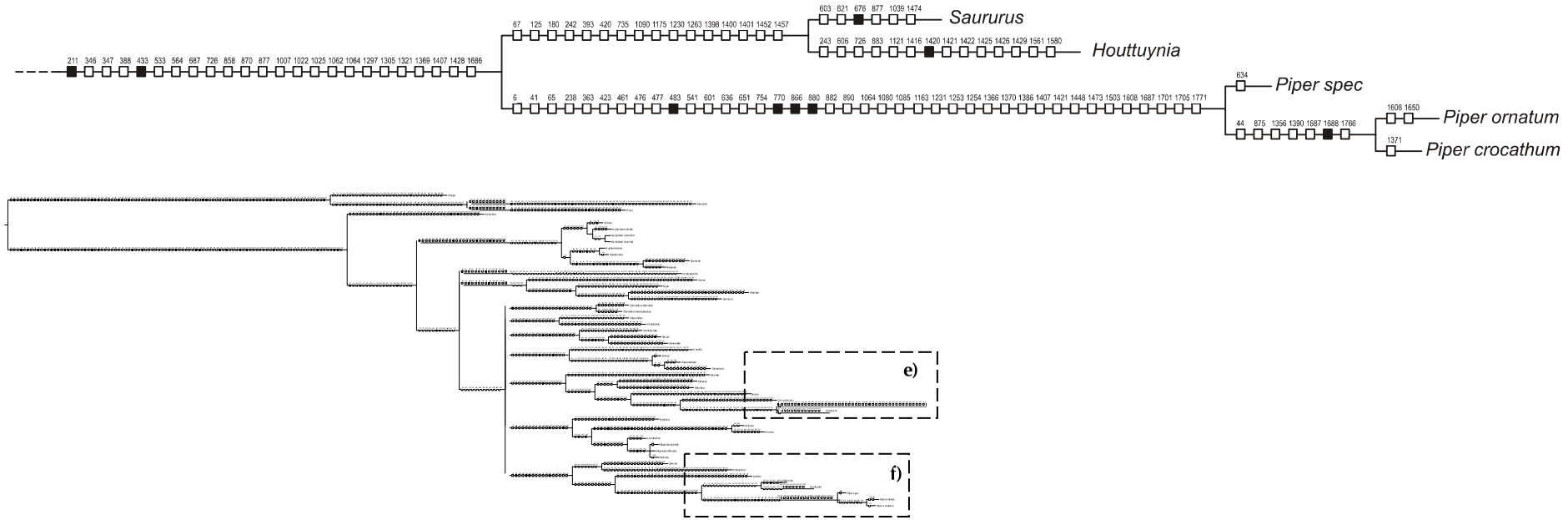

Chloranthus officinalis
Chloranthus brachystachys
Calycanthus
Umbellularia
Austrobaileya
Illicium
Schisandra
Canella
Drimys
Pseudowintera
Tasmannia
Dicentra
Platanus
Nelumbo
Buxus
e)
Myristica
Asimina
Annona
Liriodendron
Magnolia sieboldii
Magnolia officinalis
Michelia
Saruma
Aristolochia
Lactoris
f)

Appendix 2 (continued)

e)



f)

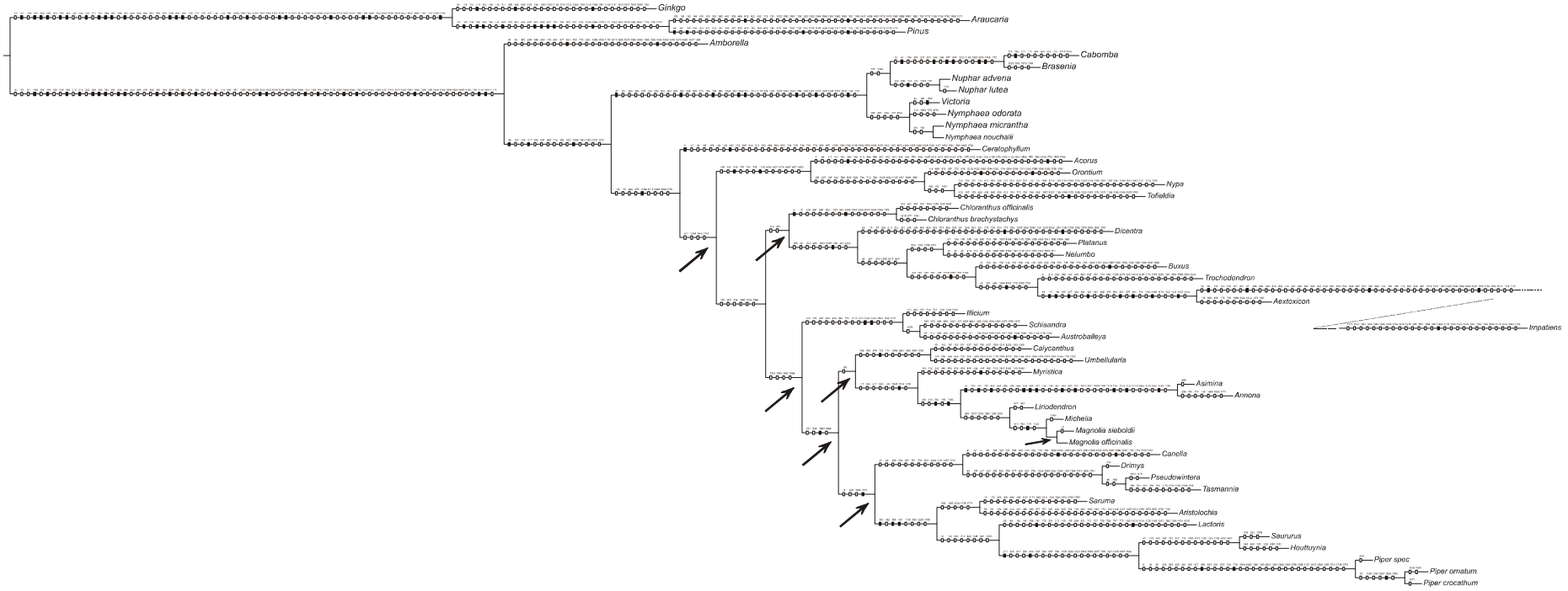


Appendix 3 — Distribution of parsimony informative characters on one of the 113 most parsimonious trees obtained from intron+spacer matrix of the *petD* data set (i.e., indels not included; see Chapter 2). Characters are optimised on the tree using Winclada 1.00.08 (Nixon 2002) assuming accelerated transformation (ACCTRAN). This figure (and all other figures) is available for download under http://mbe.oxfordjournals.org/cgi/content/abstract/22/2/317 (file number 5 of the supplementary material).

a) Overview of the tree with arrows indicating nodes, which collapse in the strict consensus.

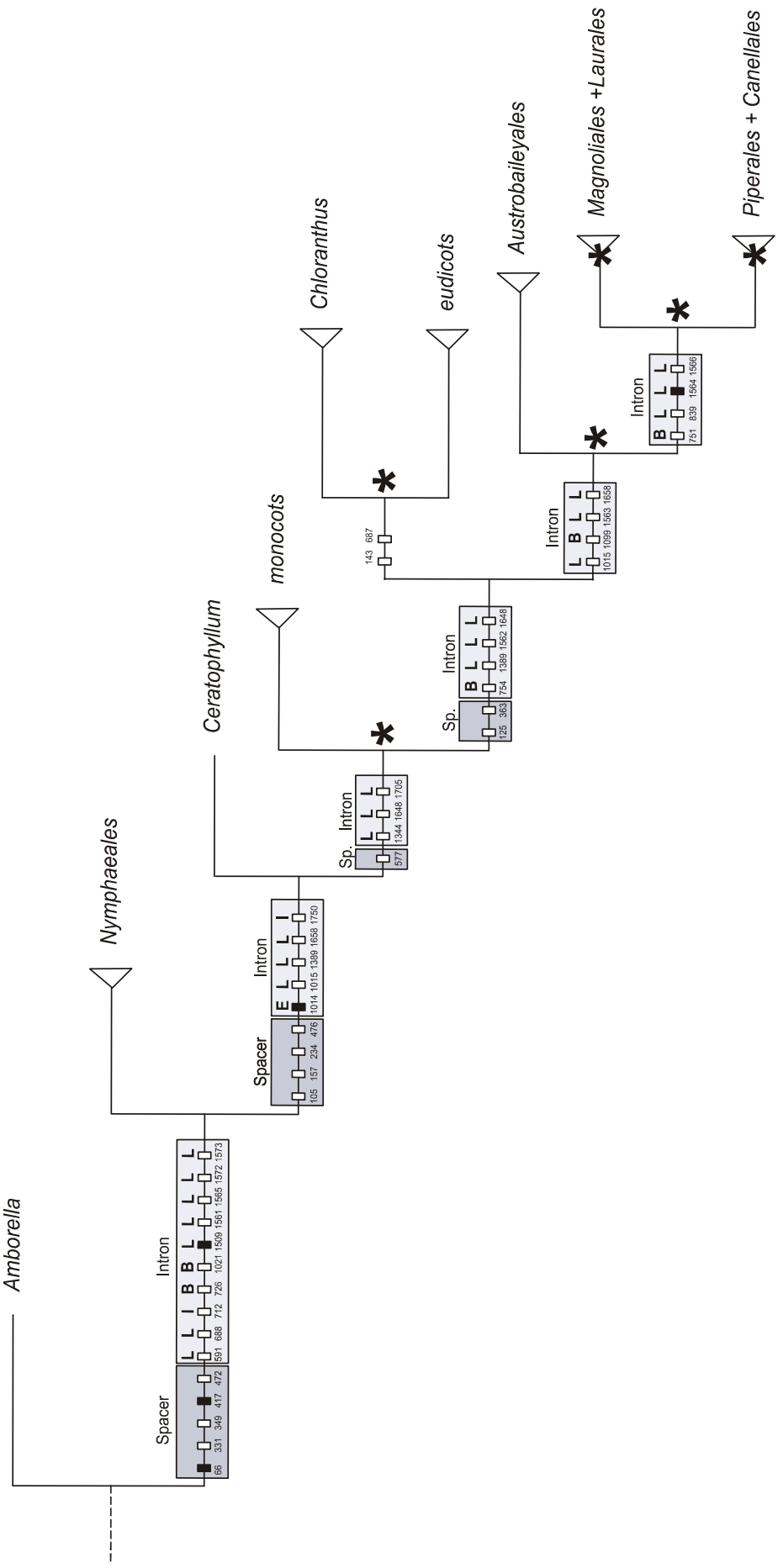
b, next page) An enlarged section of the total tree, showing the basalmost nodes within angiosperms. Structural elements that have been assigned to the respective characters are indicated by letters (L = loop, B = bulge, I = interhelical, E = exon binding site 1 - a conserved stretch within a loop). Asterisk mark those nodes which are collapsing in the strict consensus.

a)

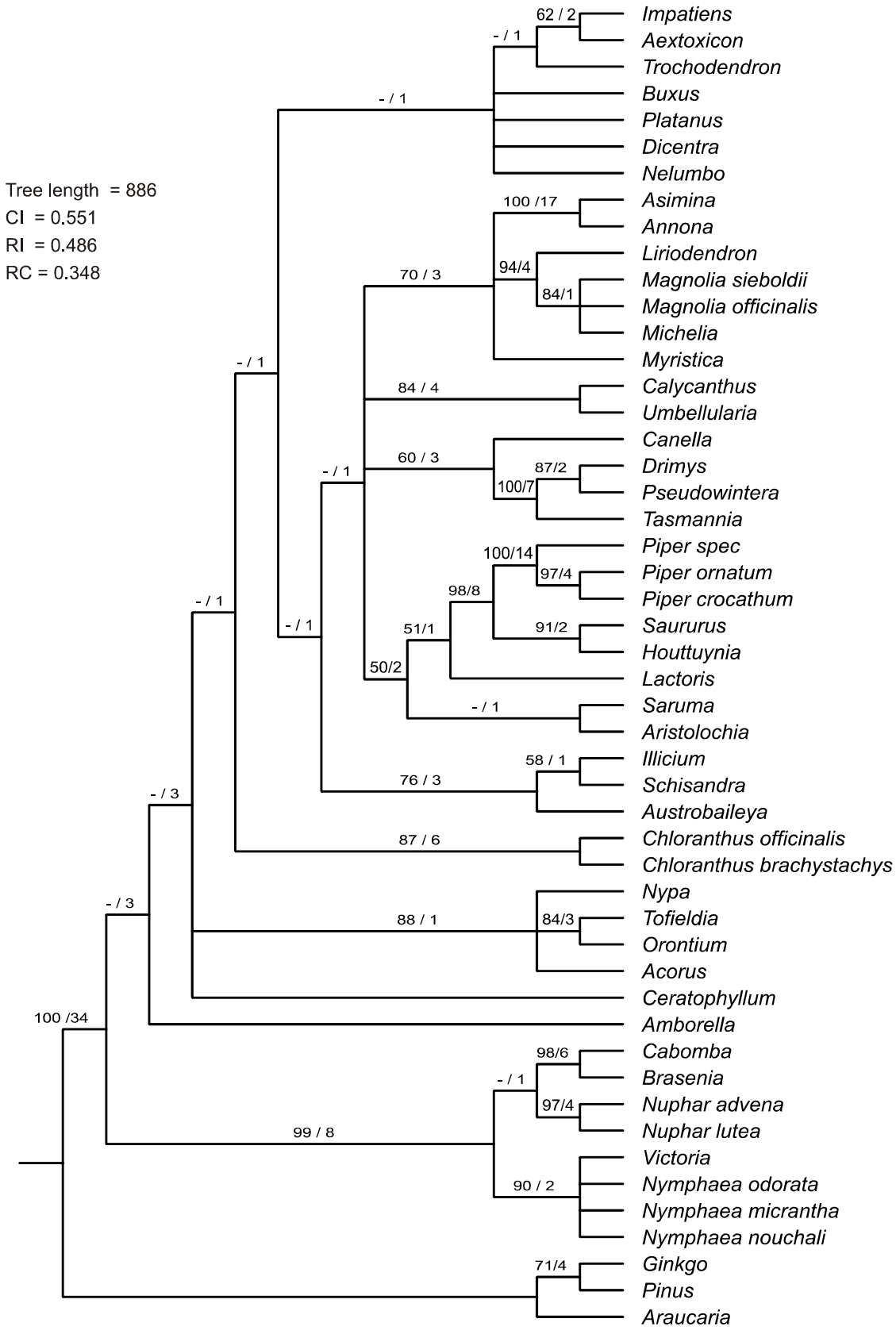


Appendix 3 (continued)

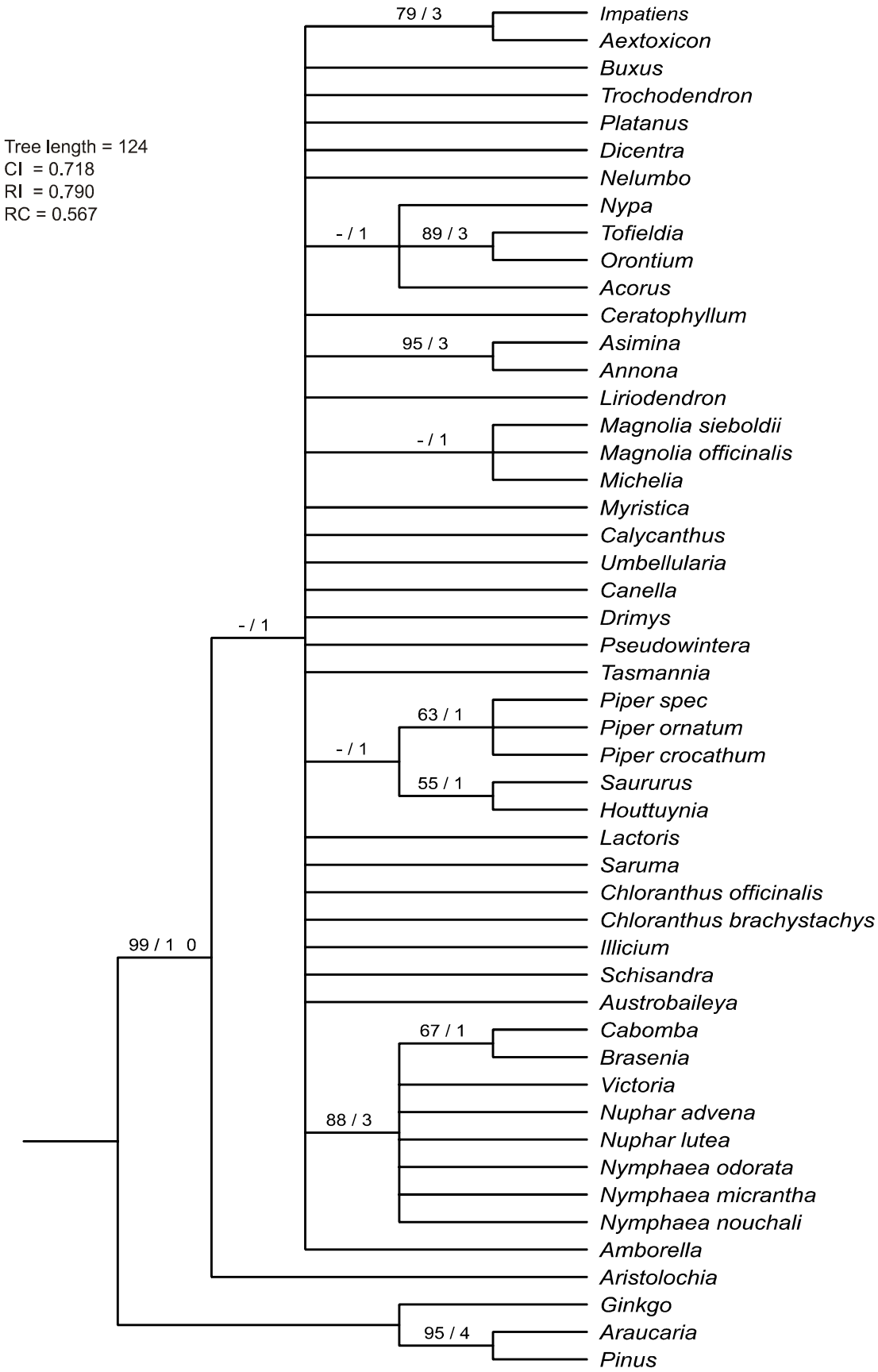
b)



Appendix 4 — a) Strict consensus of 113 trees obtained from the partition of non-pairing structural elements of the *petD* intron (loops, bulges, interhelical stretches = 800 characters of the intron matrix, see Chapter 2). Numbers above branches indicate bootstrap and Bremer support.



Appendix 4 — b) Strict consensus of 10,000 trees obtained from the partition of stem elements of the *petD* intron (192 characters of the intron matrix; see Chapter 2). Numbers above branches indicate bootstrap and Bremer support.



Appendix 5 — Overview on PCR and Cycle–Sequencing conditions used in this study**.** "Program" refers to (inofficial) names and numbers of the PCR–programs used at the thermocycler in the molecular systematics lab of the Nees Institute / University of Bonn.

***petD* region (see Chapter 2 and 3)**			
Primer set	**Reaction mixture**	**PCR conditions**	**Program**
PipetB1411F / PipetD738R (whole fragment)	**(50µl)**: 26.7 µl H2O, 5 µl Buffer, 2 µl Primer (each), 10 µl dNTP's, 0.3 µl Taq Polymerase, 4 µl DNA (1:10)	**1) initial denaturation** (1.5 min at 96° C), **2) 35 cycles** of denaturation (30 sec at 95°C), annealing (1 min at 50°C) and extension (1.5 min at 72°C), **3) final extension** (20 min at 72°C)	Kai1 (No. 4)

***rpl16* region (see Chapter 3)**			
Primer set	**Reaction mixture**	**PCR conditions**	**Program**
rpl16F / rpl16R (whole fragment)	**(50µl)**: 23.7 µl H2O, 5 µl Buffer, 3 µl $MgCl_2$, 2 µl Primer (each), 10 µl dNTP's, 0.3 µl Taq Polymerase, 4 µl DNA (1:10)	**1) 35 cycles** of denaturation (1 min at 94°C), annealing (1 min at 48°C) and extension (2 min at 72°C), **2) final extension** (15 min at 72°C)	trnTF–AN (No. 29)
NYrps3F / rpl16R (larger fragment, because NYrps3F is located further upstream from the position of rpl16F)	**(50µl)**: 23.7 µl H2O, 5 µl Buffer, 3 µl $MgCl_2$, 2 µl Primer (each), 10 µl dNTP's, 0.3 µl Taq Polymerase, 4 µl DNA (1:10)	**1) initial denaturation** (1.5 min at 96° C), **2) 35 cycles** of denaturation (30 sec at 95°C), annealing (1 min at 50°C) and extension (1.5 min at 72°C), **3) final extension** (20 min at 72°C)	Kai1 (No. 4)

***trnK* / *matK* region (see Chapter 3)**			
Primer set	**Reaction mixture**	**PCR conditions**	**Program**
trnKf-bry /psbAR (whole fragment)	**(50µl)**: 25.8 µl H2O, 5 µl Buffer, 2 µl Primer (each), 10 µl dNTP's, 0.3 µl Taq Polymerase, 5 µl DNA (1:10)	**1) initial denaturation** (3 min at 96° C), annealing (3 min at 50° C), extension (3 min 72° C) **2) 40 cycles** of denaturation (30 s at 94° C), annealing (1.5 min at 48°C), and extension (3 min at 72° C) **3) final extension** (20min at 72°C)	matK+
trnKf–bry / ARmatK660R (5'-fragment)	**(25µl)**: 3µl H_2O, 2.5 µl Buffer, 0.5 µl Primer (each), 3.3 µl dNTP's, 0.2 µl Taq Polymerase, 15 µl DNA (1:100)	**1) initial denaturation** (1.5 min at 96° C), **2) 35 cycles** of denaturation (30 sec at 95°C), annealing (1 min at 50°C) and extension (1.5 min at 72°C), **3) final extension** (20 min at 72°C)	Kai1 (No. 2)
NYmatK480F / psbAR (3'-fragment)			

Appendix 5 (continued)

***trnT–trnF* region (see Chapter 3 and 4)**			
Primer set	**Reaction mixture**	**PCR conditions**	**Program**
rps4–5R / trnL–110R	**(50µl)**: 25.7 µl H2O, 5 µl Buffer, 3 µl $MgCl_2$, 2 µl Primer (each), 10 µl dNTP's, 0.3 µl Taq Polymerase, 2 µl DNA (1:10)	**1) 36 cycles** of denaturation (1 min at 94°), annealing (1 min at 52°C) and extension (2 min at 72°C), **2) final extension** (15 min at 72°C)	trnTF (No. 2)
C – F	**(50µl)**: 25.7 µl H2O, 5 µl Buffer, 3 µl $MgCl_2$, 2 µl Primer (each), 10 µl dNTP's, 0.3 µl Taq Polymerase, 2 µl DNA (1:10)	**1) 36 cycles** of denaturation (1 min at 94°), annealing (1 min at 52°C) and extension (2 min at 72°C), **2) final extension** (15 min at 72°C)	trnTF (No. 2)

ITS region (see Chapter 4)			
Primer set	**Reaction mixture**	**PCR conditions**	**Program**
ITS5 / ITS4 (whole fragment)	**(25µl)**: 1.5 µl H2O, 2.5 µl Buffer, 0.4 µl $MgCl_2$, 0.75 µl Primer (each), 3.9 µl dNTP's, 0.2 µl Taq Polymerase, 15 µl DNA (1:100)	**1) 35 cycles** of denaturation (1 min at 97°C), annealing (1 min at 48°C) and extension (45 s at 72°C), **3) final extension** (7 min at 72°C)	ITS-TB1 (No. 9)

Chemicals used:

- Primer working concentration 20mmol / µl (MWG Biotech, Ebersberg / Germany)
- Taq Polymerase: SAWADY Taq-DNA-Polymerase (Peqlab Biotechnologie GmbH, Erlangen / Germany; catalogue number 01-1020)
- Buffer (comes with Peqlab Taq–Polymerase): 10x Reaktionspuffer S für hohe Spezifitäten (100 mM Tris-HCl pH 8.8, 500 mM KCl, 0.1 % Tween 20, 15 mM $MgCl_2$)
- dNTPs: peqGOLD dNTP-Set 4x25 µmol (Peqlab Biotechnologie GmbH, Erlangen / Germany; catalogue number 20-2010)

Reaction kits used:

- QIAquick Gel Extraction Kit (QIAGEN GmbH, Hilden / Germany; catalogue number 28706)
- Cycle Sequencing (CEQ): DTCS Quick Start Kit (Beckman Coulter GmbH, Krefeld / Germany; catalogue number 608120)
- Cycle Sequencing (ABI): BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems Applera Deutschland GmbH, Darmstadt / Germany; catalogue number 4337450), and BigDye® Terminator v3.1 (catalogue number 4337455)

Cycle–Sequencing reaction mixtures (total volume 10µl):

- for ABI sequencer: 0–6 µl H2O, 0.5–7 µl template, 0.5 µl Primer (20 pmol/µl), 2.5 µl BigDye Terminator Ready Reaction Kit version 1.1
- for CEQ sequencer: 0–6 µl H2O, 0.5–6.5 µl template, 0.5 µl Primer (20 pmol/µl), 3 µl CEQ DTCS Quick Start Kit

Cycle–Sequencing Conditions:

- ABI sequencer: 30 cycles of denaturation (96°C, 5s), annealing (50°C, 15s) and elongation (50°C, 4 min)
- CEQ sequencer: 30 cycles of denaturation (96°C, 20 s), annealing (50°C, 20 s) and elongation (60°, 4 min)

Appendix 6 — Currently recognized species of Nymphaeaceae and Cabombaceae (Nymphaeales). Distribution data was compiled for all currently recognized species of Nymphaeales, and was used for the production of genera distribution maps (Figure 5.3–5.5) and the diversity map (Figure 5.6) presented in Chapter 5.

Cabombaceae

***Cabomba* Aubl.**

C. aquatica Aubl.
C. caroliniana A. Gray
C. furcata Schult & Schult f.
C. haynesii Wiersema
C. palaeformis Fassett

***Brasenia* Schreb.**

Brasenia schreberi J. F. Gmel.

Nymphaeaceae

***Nuphar* Sm.**

sect. *Astylus*

N. advena (Aiton) W. T. Aiton
N. polysepala Engelm.
N. sagittifolia (Walter) Pursh
N. variegata Engelm. ex Durand

sect. *Nuphar*

N. japonica DC.
N. lutea (L.) Sm.
N. microphylla (Pers.) Fernald
N. pumila (Timm) DC.

Nymphaea* L. subg. *Brachyceras

N. ampla (Salisb.) DC.
N. caerulea Savigny
N. capensis Thunb.
N. colorata Peter
N. divaricata Hutch.
N. elegans Hook.
N. gracilis Zucc.
N. guineensis Schumach. & Thonn.
N. heudelotii Planch.
N. micrantha Guill. & Perr.
N. nouchali Burm. f.
N. ovalifolia Conard
N. pulchella DC.
N. stuhlmannii (Engl.) Schweinf. & Gilg
N. sulphurea Gilg
N. thermarum Eb. Fisch.

***Barclaya* Wall.**

B. kunstleri (King) Ridl.
B. longifolia Wall.
B. motleyi Hook. f.
B. rotundifolia Hotta

***Victoria* Lindl.**

V. amazonica (Poepp.) J. C. Sowerby
V. cruziana A. D. Orb

***Euryale* Salisb.**

E. ferox Salisb.

Nymphaea* L. subg. *Hydrocallis

N. amazonum Mart. & Zucc.
N. belophylla Trickett
N. conardii Wiersema
N. gardneriana Planch.
N. *glandulifera* Rodschied
N. jamesoniana Planch.
N. lasiophylla Mart. & Zucc.
N. lingulata Wiersema
N. novogranatensis Wiersema
N. oxypetala Planch.
N. potamophila Wiersema
N. prolifera Wiersema
N. rudgeana G. Mey.
N. tenerinervia Casp.

***Ondinea* Hartog**

O. purpurea Hartog [a]

Nymphaea* L. subg. *Anecphya

N. atrans S. W. L. Jacobs
N. elleniae S. W. L. Jacobs
N. gigantea Hook.
N. hastifolia Domin
N. immutabilis S. W. L. Jacobs
N. macrosperma Merr. & L. M. Perry
N. violacea Lehm.

Nymphaea* L. subg. *Lotos

N. lotos L.
N. petersiana Klotzsch [b]
N. pubescens Willd.
(*N. rubra* Roxb. ex Andrews) [c]

Nymphaea* L. subg. *Nymphaea

sect. *Nymphaea*
N. alba L.
N. candida C. Presl.
N. odorata Aiton

sect. *Chamaenymphaea*
N. leibergii Morong
N. pygmaea (Salisb.) W. T. Aiton
N. tetragona Georgi

sect. *Xanthantha*
N. mexicana Zucc.

Notes. — [a] distribution of *Ondinea* was recorded, but *Ondinea* was not treated as a distinct taxon in DIVA–analysis (see explanation in the text).

[b] *Nymphaea petersiana*, currently classified under subg. *Brachyceras*, is here subsumed under subg. *Lotos* because recent molecular studies show close affinity of *N. petersiana* to *N. lotus* and *N. pubescens* (Borsch 2000, Borsch *et al.* in prep., Löhne *et al.* in prep., see also Chapter 3)

[c] *Nymphaea rubra* (subg. *Lotos*) from East India and Bangladesh was not sampled here because the identity of this taxon is considered uncertain (sterile populations occurring, might be of hybrid origin)

Appendix 7 — List of floras used for compilation of distribution data for Nymphaeales (Chapter 5, page 109 ff.). The references listed here were checked in the course of this study. More information on species distribution was obtained from the online database of the Germplasm Resources Information Network (USDA, ARS; Cabombaceae: http://www.ars-grin.gov/cgi-bin/npgs/html/gnlist.pl?185, Nymphaeaceae: http://www.ars-grin.gov/cgi-bin/npgs/html/gnlist.pl?779). Date of data accession: May–Decemb er 2005. See this online–ressource for more references.

North America

- Rzedowski, J. & G. Rzedowski, eds. 1979–. Flora fanerogámica del Valle de México.
- Scoggan, H. J. 1978–1979. The flora of Canada, 4 vol.
- Wiersema, J. H. 1997. Cabombaceae, Nymphaeaceae. In: Flora of North America Vol. 3 (FNA Editorial Committee, Ed.)
  (online ressource: http://www.efloras.org/volume_page.aspx?volume_id=1003&flora_id=1)

Central America

- Adams C. D. (1972). Flowering plants of Jamaica. The University Press, Glasgow / UK.
- Correll D. S. & H. B. Correll (1982) Flora of the Bahama archipelago. J. Cramer, Vaduz.
- Howard R. (1974–1989). Flora of the lesser Antilles. Harvard University, Massachusetts.
- Leon (J. S. Sauget) & Alain (E. E. Liogier) (1946–1953). Flora de Cuba.. Vol. 1

South America

- Burkart, A. E., ed. 1969. Flora ilustrada de Entre Rios (Argentina). Buenos Aires.
- Cornejo, X. & C. Bonifaz (2003). Nymphaeaceae. In: Flora of Ecuador (Harling, G. & L. Andersson, Eds.).
- Dubs, B. 1998. Prodromus florae matogrossensis. Botany of Mato Grosso, series B, no. 3.
- Jørgensen, P. M. & S. León-Yánez, eds. 1999. Catalogue of the vascular plants of Ecuador. Monogr. Syst. Bot. Missouri Bot. Gard. 75.
- Wiersema, J. H. 1987. A monograph of *Nymphaea* subgenus *Hydrocallis* (Nymphaeaceae). *Syst. Bot. Monogr.* 16: 80 pp.
- Wiersema, J. H. 2003. Nymphaeaceae. In: Steyermark, J. A. et al., eds., Flora of the Venezuelan Guayana.

Europe & temperate Asia

- Davis P. H., Ed. (1965–1988). Flora of Turkey and the east Aegean islands.
- Fu D.-Z. & J. H. Wiersema. (2001). Nymphaeaceae. In: Flora of China Vol. 6 (Wu Zheng-yi & P. H. Raven et al., eds.)
  (online resource: http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=10618)
- Jalas J. & J. Suominen, Eds. (1989). Atlas Florae Europaeae: Distribution of vascular plants in Europe. Helsinki.
- Kadono Y. (1994). Aquatic plants of Japan. Bun-ichi Sogo Shuppan Co., Ltd., Tokyo.
- Khan Z. *et al*. (1997). Aquatic plants of Iran and Pakistan: 8. Nymphaeaceae. *Biologia* (Lahore) 43:10.
- Komarov V. L. *et al*., Eds. (1934–1964). Flora SSSR.
- Li H. L. et al., Eds. (1975–1979). Flora of Taiwan. 2:546.
- Ohwi J. (1965). Flora of Japan (Engl. ed.).
- Sharma B. D. *et al*., Eds. (1993–). Flora of India.
- Zohary M. & N. Feinbrun-Dothan. (1966–). Flora palaestina.

Appendix 7 (continued)

Africa

- Arnold T. H. & B. C. De Wet, Eds. (1993). Plants of southern Africa: names and distribution. Mem. Bot. Surv. S. Africa no. 62.
- Berhaut J. (1979). Flore illustrée du Sénégal. Governement du Sénégal, Dakar.
- Brunel J. F., P. Hiepko & H. Scholz, Eds. (1984). Flore analytique du Togo. Phanerogames. Berlin
- Cufodontis G. (1953–1972). Enumeratio plantarum aethiopiae: Spermatophyta.
- Exell A. W. et al., Eds. (1960–). Flora Zambesiaca. (online ressource: http://www.rbgkew.org.uk/efloras/search.do)
- Hauman L. (1948). Nymphaeaceae. In: Flore du Congo Belge et du Ruanda–Urundi (Comité exécutif de la Flore du Congo belge et le Jardin Botanique de l'État, Eds.).
- Humbert J.-H., Ed. (1936–). Flore de Madagascar et des Comores.
- Hutchinson J., J. M. Dalziel & R. W. J. Keay (1954). Flora of west tropical Africa, 2nd ed.
- Lisowski S., F. Malaisse & J. J. Symoens (1970). Cabombaceae. In: Flore du Congo du Rwanda et du Burundi (Jardin botanique national de Belgique, Ed.).
- Thulin M., Ed. (1993–). Flora of Somalia.
- Troupin G. (1978–). Flore du Rwanda.
- Verdcourt B. (1989). Nymphaeaceae. In: Flora of tropical East Africa (R. M. Polhill, Ed.).

Tropical Asia & Australia

- Dassanayake M. D. & F. R. Fosberg, Eds. (1980–). A revised handbook to the flora of Ceylon.
- Jacobs S. W. L. & C. L. Porter (in press). Nymphaeaceae. In: Flora of Australia Vol. 2 (A. Wilson, Ed.).
- Matthew K. M. (1983). The flora of the Tamilnadu Carnatic.
- Saldanha C. J. (1985–). Flora of Karnataka.
- Sharma B. D. *et al.*, Eds. (1993–). Flora of India.
- van Royen, P. (1962). Sertulum papuanum 5. Nymphaeaceae. Nova Guinea, Bot. 8:107.
- Womersley, J. S., Ed. (1978–). Handbooks of the flora of Papua New Guinea.

Appendix 8 — Dispersal–vicariance analyses of the Nymphaeales (Chapter 5). Shown are the original data files used for analyses (topology A and B, only extant or extant plus fossil distribution coded) as well as the output files for each analysis.

a) data file 1 (topology A, only extant distribution coded), written in nexus–format

```
#NEXUS
BEGIN data;
       DIMENSIONS NTAX=16 nchar=9;
       FORMAT DATATYPE=STANDARD MISSING=? GAP=-;
       MATRIX
[      unit area      ABCDEFGHI]
       Amborella      000000001
       Austrobaileya  000000010
       Illicium       110001100
       Kadsura        000001100
       Schisandra     100001000
       Cabomba        111000000
       Brasenia       111011110
       Nuphar         100101000
       Barclaya       000000100
       Victoria       001000000
       Euryale        000001100
       Ny_Anecphya    000000010
       Ny_Brachyceras 111011110
       Ny_Hydrocallis 011000000
       Ny_Lotos       000111100
       Ny_Nymphaea    110101000;
END;

BEGIN TREES;
TREE 1 =[&R] (Amborella ,(((Nuphar ,(Barclaya ,(Ny_Nymphaea ,((Ny_Anecphya
,Ny_Brachyceras),((Ny_Hydrocallis ,Ny_Lotos),(Victoria ,Euryale ))))))),(Cabomba ,Brasenia
)),(Austrobaileya ,(Illicium ,(Kadsura ,Schisandra ))))));
END;

output NY9A_recent2.txt;
echo status;
optimize maxareas=2;
output --;

output NY9A_recent5.txt;
echo status;
optimize maxareas=5;
output --;

output NY9A_recent9.txt;
echo status;
optimize maxareas=9;
output --;
```

results – data file 1, maxareas = 2

```
ny8a_fossil2.txt opened as output file
echo status on
optimization successful - exact solution
       settings: maxareas=2, bound=250, hold=1000, weight=1.000, age=1.000
       optimal reconstruction requires 35 dispersals

optimal distributions at each node:
node 17 (anc. of terminals ny_anecphya-ny_brachyceras): H  FH
node 18 (anc. of terminals ny_hydrocallis-ny_lotos): BF  CF
node 19 (anc. of terminals victoria-euryale): CF
node 20 (anc. of terminals ny_hydrocallis-euryale): F
node 21 (anc. of terminals ny_anecphya-euryale): F  FH
node 22 (anc. of terminals ny_nymphaea-euryale): F
node 23 (anc. of terminals barclaya-euryale): FG
node 24 (anc. of terminals nuphar-euryale): F
node 25 (anc. of terminals cabomba-brasenia): A  B  C  AF  BF  CF
node 26 (anc. of terminals nuphar-brasenia): F  AF  BF  CF
node 27 (anc. of terminals kadsura-schisandra): F
node 28 (anc. of terminals illicium-schisandra): F
node 29 (anc. of terminals austrobaileya-schisandra): FH
node 30 (anc. of terminals nuphar-schisandra): F
node 31 (anc. of terminals amborella-schisandra): FH
```

results data file 1 – maxareas = 5

```
ny8a_fossil5.txt opened as output file
echo status on
optimization successful - exact solution
       settings: maxareas=5, bound=250, hold=1000, weight=1.000, age=1.000
       optimal reconstruction requires 34 dispersals

optimal distributions at each node:
node 17 (anc. of terminals ny_anecphya-ny_brachyceras): H  AH  ABH  CH  ACH  BCH  ABCH
     AEH  ABEH  CEH  ACEH  BCEH  ABCEH  FH  AFH  BFH  ABFH  CFH  ACFH  BCFH  ABCFH  EFH
     AEFH  BEFH  ABEFH  CEFH  ACEFH  BCEFH  GH  AGH  BGH  ABGH  CGH  ACGH  BCGH  ABCGH  EGH
     AEGH  BEGH  ABEGH  CEGH  ACEGH  BCEGH  FGH  AFGH  BFGH  ABFGH  CFGH  ACFGH  BCFGH
     EFGH  AEFGH  BEFGH  CEFGH
node 18 (anc. of terminals ny_hydrocallis-ny_lotos): CD  BCD  CE  BCE  CDE  BCDE  BF  CF
     BCF  BDF  CDF  BCDF  BEF  CEF  BCEF  BDEF  CDEF  BCDEF  BG  CG  BCG  BDG  CDG  BCDG
     BEG  CEG  BCEG  BDEG  CDEG  BCDEG  BFG  CFG  BCFG  BDFG  CDFG  BCDFG  BEFG  CEFG
     BCEFG  BDEFG  CDEFG
node 19 (anc. of terminals victoria-euryale): CF  CG  CFG
node 20 (anc. of terminals ny_hydrocallis-euryale): C  F  G
node 21 (anc. of terminals ny_anecphya-euryale): C  AC  ABC  ACE  ABCE  F  AF  ABF  ACF
     ABCF  AEF  ABEF  ACEF  ABCEF  G  AG  ABG  ACG  ABCG  AEG  ABEG  ACEG  ABCEG  AFG  ABFG
     ACFG  ABCFG  AEFG  ABEFG  ACEFG  CH  ACH  ABCH  ACEH  ABCEH  FH  AFH  ABFH  ACFH
     ABCFH  AEFH  ABEFH  ACEFH  GH  AGH  ABGH  ACGH  ABCGH  AEGH  ABEGH  ACEGH  AFGH  ABFGH
     ACFGH  AEFGH
node 22 (anc. of terminals ny_nymphaea-euryale): A  AC  ABC  ACD  ABCD  F  AF  ABF  ACF
     ABCF  ADF  ABDF  ACDF  ABCDF  AG  BG  ABG  DG  ADG  BDG  ABDG  FG  AFG  BFG  ABFG  DFG
     ADFG  BDFG  ABDFG  AH  ABH  ACH  ABCH  ADH  ABDH  ACDH  ABCDH  AFH  ABFH  ACFH  ABCFH
     ADFH  ABDFH  ACDFH  AGH  BGH  ABGH  DGH  ADGH  BDGH  ABDGH  FGH  AFGH  BFGH  ABFGH
     DFGH  ADFGH  BDFGH
node 23 (anc. of terminals barclaya-euryale): G  AG  ABG  ACG  ABCG  ADG  ABDG  ACDG
     ABCDG  FG  AFG  ABFG  ACFG  ABCFG  ADFG  ABDFG  ACDFG  AGH  ABGH  ACGH  ABCGH  ADGH
     ABDGH  ACDGH  AFGH  ABFGH  ACFGH  ADFGH
node 24 (anc. of terminals nuphar-euryale): A  F  AF  ADF  AG  DG  ADG  AFG  DFG  ADFG
node 25 (anc. of terminals cabomba-brasenia): A  B  C  ABC  ABE  ACE  BCE  ABCE  AF  BF
     ABF  CF  ACF  BCF  ABCF  AEF  BEF  ABEF  CEF  ACEF  BCEF  ABCEF  ABG  ACG  BCG  ABCG
     AEG  BEG  ABEG  CEG  ACEG  BCEG  ABCEG  AFG  BFG  ABFG  CFG  ACFG  BCFG  ABCFG  AEFG
     BEFG  ABEFG  CEFG  ACEFG  BCEFG  ABCH  ABEH  ACEH  BCEH  ABCEH  AFH  BFH  ABFH  CFH
     ACFH  BCFH  ABCFH  AEFH  BEFH  ABEFH  CEFH  ACEFH  BCEFH  ABGH  ACGH  BCGH  ABCGH
     AEGH  BEGH  ABEGH  CEGH  ACEGH  BCEGH  AFGH  BFGH  ABFGH  CFGH  ACFGH  BCFGH  AEFGH
     BEFGH  CEFGH
node 26 (anc. of terminals nuphar-brasenia): A  ABD  ACD  F  AF  BF  CF  ABCF  ABDF  ACDF
     ABEF  ACEF  BCEF  ABCEF  ABG  ACG  ADG  BDG  ABDG  CDG  ACDG  ABFG  ACFG  BCFG  ABCFG
     ADFG  BDFG  ABDFG  CDFG  ACDFG  AEFG  BEFG  ABEFG  CEFG  ACEFG  BCEFG  ABCFH  ABEFH
     ACEFH  BCEFH  ABFGH  ACFGH  BCFGH  AEFGH  BEFGH  CEFGH
node 27 (anc. of terminals kadsura-schisandra): F
node 28 (anc. of terminals illicium-schisandra): F  ABF  AFG  BFG  ABFG
node 29 (anc. of terminals austrobaileya-schisandra): ABH  FH  ABFH  AGH  BGH  ABGH  AFGH
     BFGH  ABFGH
node 30 (anc. of terminals nuphar-schisandra): ABCF  ABDF  ACDF  ABEF  ACEF  BCEF  ABCEF
     ABCG  ABFG  ACFG  BCFG  ABCFG  ADFG  BDFG  ABDFG  CDFG  ACDFG  AEFG  BEFG  ABEFG  CEFG
     ACEFG  BCEFG  ABCEH  ABCFH  ABDFH  ACDFH  ABEFH  ACEFH  BCEFH  ABCGH  ABDGH  ACDGH
     ABEGH  ACEGH  BCEGH  ABFGH  ACFGH  BCFGH  ADFGH  BDFGH  CDFGH  AEFGH  BEFGH  CEFGH
node 31 (anc. of terminals amborella-schisandra): ABCEH  ABCFH  ABDFH  ACDFH  ABEFH  ACEFH
     BCEFH  ABCGH  ABDGH  ACDGH  ABEGH  ACEGH  BCEGH  ABFGH  ACFGH  BCFGH  ADFGH  BDFGH
     CDFGH  AEFGH  BEFGH  CEFGH
```

results data file 1 – maxareas = 9

```
ny9a_recent9.txt opened as output file
echo status on
optimization successful - exact solution
       settings: maxareas=9, bound=250, hold=1000, weight=1.000, age=1.000
       optimal reconstruction requires 31 dispersals

optimal distributions at each node:
node 17 (anc. of terminals ny_anecphya-ny_brachyceras): H  CH  ACH  BCH  ABCH  CEH  ACEH
     BCEH  ABCEH  FH  AFH  BFH  ABFH  CFH  ACFH  BCFH  ABCFH  EFH  AEFH  BEFH  ABEFH  CEFH
     ACEFH  BCEFH  ABCEFH  GH  AGH  BGH  ABGH  CGH  ACGH  BCGH  ABCGH  EGH  AEGH  BEGH
     ABEGH  CEGH  ACEGH  BCEGH  ABCEGH  FGH  AFGH  BFGH  ABFGH  CFGH  ACFGH  BCFGH  ABCFGH
     EFGH  AEFGH  BEFGH  ABEFGH  CEFGH  ACEFGH  BCEFGH  ABCEFGH
node 18 (anc. of terminals ny_hydrocallis-ny_lotos): BD  CD  BCD  CE  BCE  BDE  CDE  BCDE
     BF  CF  BCF  BDF  CDF  BCDF  BEF  CEF  BCEF  BDEF  CDEF  BCDEF  BG  CG  BCG  BDG  CDG
     BCDG  BEG  CEG  BCEG  BDEG  CDEG  BCDEG  BFG  CFG  BCFG  BDFG  CDFG  BCDFG  BEFG  CEFG
     BCEFG  BDEFG  CDEFG  BCDEFG
node 19 (anc. of terminals victoria-euryale): CF  CG  CFG
```

```
node 20 (anc. of terminals ny_hydrocallis-euryale): C  CD  BCD  CDE  BCDE  F  DF  BDF  CDF
     BCDF  DEF  BDEF  CDEF  BCDEF  G  DG  BDG  CDG  BCDG  DEG  BDEG  CDEG  BCDEG  DFG  BDFG
     CDFG  BCDFG  DEFG  BDEFG  CDEFG  BCDEFG
node 21 (anc. of terminals ny_anecphya-euryale): C  F  G  CH  DH  BDH  CDH  BCDH  DEH
     BDEH  CDEH  BCDEH  FH  DFH  BDFH  CDFH  BCDFH  DEFH  BDEFH  CDEFH  BCDEFH  GH  DGH
     BDGH  CDGH  BCDGH  DEGH  BDEGH  CDEGH  BCDEGH  DFGH  BDFGH  CDFGH  BCDFGH  DEFGH
     BDEFGH  CDEFGH  BCDEFGH
node 22 (anc. of terminals ny_nymphaea-euryale): D  CD  ACD  BCD  ABCD  F  DF  ADF  BDF
     ABDF  CDF  ACDF  BCDF  ABCDF  AG  BG  ABG  DG  ADG  BDG  ABDG  FG  AFG  BFG  ABFG  DFG
     ADFG  BDFG  ABDFG  DH  ADH  BDH  ABDH  CDH  ACDH  BCDH  ABCDH  DFH  ADFH  BDFH  ABDFH
     CDFH  ACDFH  BCDFH  ABCDFH  AGH  BGH  ABGH  DGH  ADGH  BDGH  ABDGH  FGH  AFGH  BFGH
     ABFGH  DFGH  ADFGH  BDFGH  ABDFGH
node 23 (anc. of terminals barclaya-euryale): G  DG  ADG  BDG  ABDG  CDG  ACDG  BCDG
     ABCDG  FG  DFG  ADFG  BDFG  ABDFG  CDFG  ACDFG  BCDFG  ABCDFG  DGH  ADGH  BDGH  ABDGH
     CDGH  ACDGH  BCDGH  ABCDGH  DFGH  ADFGH  BDFGH  ABDFGH  CDFGH  ACDFGH  BCDFGH  ABCDFGH
node 24 (anc. of terminals nuphar-euryale): D  F  DF  ADF  DG  ADG  DFG  ADFG
node 25 (anc. of terminals cabomba-brasenia): BCE  ABCE  BCEF  ABCEF  BCEG  ABCEG  BCEFG
     ABCEFG  BCEH  ABCEH  BCEFH  ABCEFH  BCEGH  ABCEGH  BCEFGH  ABCEFGH
node 26 (anc. of terminals nuphar-brasenia): ABCDEG  ABCEFG  ABCDEFG  ABCDEGH  ABCEFGH
     ABCDEFGH
node 27 (anc. of terminals kadsura-schisandra): F
node 28 (anc. of terminals illicium-schisandra): F
node 29 (anc. of terminals austrobaileya-schisandra): FH
node 30 (anc. of terminals nuphar-schisandra): ABCEFGH  ABCDEFGH
node 31 (anc. of terminals amborella-schisandra): ABCEFGHI  ABCDEFGHI
```

Appendix 8 (continued)

b) data file 2 (topology A, extant and fossil distribution coded), written in nexus–format

```
#NEXUS
BEGIN data;
        DIMENSIONS NTAX=16 nchar=9;
        FORMAT DATATYPE=STANDARD MISSING=? GAP=-;
        MATRIX
[       unit area       ABCDEFGHI]
        Amborella       000000001
        Austrobaileya   000000010
        Illicium        110101100
        Kadsura         000101100
        Schisandra      100101000
        Cabomba         111000000
        Brasenia        111111110
        Nuphar          100101000
        Barclaya        000000100
        Victoria        001000000
        Euryale         000101100
        Ny_Anecphya     000000010
        Ny_Brachyceras  111011110
        Ny_Hydrocallis  011000000
        Ny_Lotos        000111100
        Ny_Nymphaea     110101000;
END;

BEGIN TREES;
TREE tree =[&R] (Amborella ,(((Nuphar ,(Barclaya ,(Ny_Nymphaea ,((Ny_Anecphya
,Ny_Brachyceras),((Ny_Hydrocallis ,Ny_Lotos),(Victoria ,Euryale ))))))),(Cabomba ,Brasenia
)),(Austrobaileya ,(Illicium ,(Kadsura ,Schisandra )))));
END;

output NY9A_fossil2.txt;
echo status;
optimize maxareas=2;
output --;

output NY9A_fossil5.txt;
echo status;
optimize maxareas=5;
output --;

output NY9A_fossil9.txt;
echo status;
optimize maxareas=9;
output --;
```

results data file 2 – maxareas = 2

```
ny9a_fossil2.txt opened as output file
echo status on
optimization successful - exact solution
        settings: maxareas=2, bound=250, hold=1000, weight=1.000, age=1.000
        optimal reconstruction requires 40 dispersals

optimal distributions at each node:
node 17 (anc. of terminals ny_anecphya-ny_brachyceras): H  FH
node 18 (anc. of terminals ny_hydrocallis-ny_lotos): BD  CD  BF  CF
node 19 (anc. of terminals victoria-euryale): CD  CF
node 20 (anc. of terminals ny_hydrocallis-euryale): D  F
node 21 (anc. of terminals ny_anecphya-euryale): F  DH  FH
node 22 (anc. of terminals ny_nymphaea-euryale): D  F
node 23 (anc. of terminals barclaya-euryale): DG  FG
node 24 (anc. of terminals nuphar-euryale): D  F
node 25 (anc. of terminals cabomba-brasenia): A  B  C  AD  BD  CD  AF  BF  CF
node 26 (anc. of terminals nuphar-brasenia): D  AD  BD  CD  F  AF  BF  CF
node 27 (anc. of terminals kadsura-schisandra): D  F
node 28 (anc. of terminals illicium-schisandra): D  F
node 29 (anc. of terminals austrobaileya-schisandra): DH  FH
node 30 (anc. of terminals nuphar-schisandra): D  F
node 31 (anc. of terminals amborella-schisandra): DI  FI
```

results data file 2 – maxareas = 5

```
ny9a_fossil5.txt opened as output file
echo status on
optimization successful - exact solution
       settings: maxareas=5, bound=250, hold=1000, weight=1.000, age=1.000
       optimal reconstruction requires 38 dispersals

optimal distributions at each node:
node 17 (anc. of terminals ny_anecphya-ny_brachyceras): H  FH  AFH  BFH  ABFH  CFH  ACFH
    BCFH  ABCFH  EFH  AEFH  BEFH  ABEFH  CEFH  ACEFH  BCEFH  FGH  AFGH  BFGH  ABFGH  CFGH
    ACFGH  BCFGH  EFGH  AEFGH  BEFGH  CEFGH
node 18 (anc. of terminals ny_hydrocallis-ny_lotos): BD  CD  BCD  BDE  CDE  BCDE  BF  CF
    BCF  BDF  CDF  BCDF  BEF  CEF  BCEF  BDEF  CDEF  BCDEF  BDG  CDG  BCDG  BDEG  CDEG
    BCDEG  BFG  CFG  BCFG  BDFG  CDFG  BCDFG  BEFG  CEFG  BCEFG  BDEFG  CDEFG
node 19 (anc. of terminals victoria-euryale): CD  CF  CDF  CDG  CFG  CDFG
node 20 (anc. of terminals ny_hydrocallis-euryale): D  F
node 21 (anc. of terminals ny_anecphya-euryale): F  DH  FH
node 22 (anc. of terminals ny_nymphaea-euryale): D  F
node 23 (anc. of terminals barclaya-euryale): DG  FG
node 24 (anc. of terminals nuphar-euryale): D  F
node 25 (anc. of terminals cabomba-brasenia): A  B  C
node 26 (anc. of terminals nuphar-brasenia): AD  BD  CD  AF  BF  CF
node 27 (anc. of terminals kadsura-schisandra): D  F
node 28 (anc. of terminals illicium-schisandra): D  F
node 29 (anc. of terminals austrobaileya-schisandra): DH  FH
node 30 (anc. of terminals nuphar-schisandra): ADFH  BDFH  CDFH
node 31 (anc. of terminals amborella-schisandra): ADFHI  BDFHI  CDFHI
```

results data file 2 – maxareas = 9

```
ny9a_fossil9.txt opened as output file
echo status on
optimization successful - exact solution
       settings: maxareas=9, bound=250, hold=1000, weight=1.000, age=1.000
       optimal reconstruction requires 35 dispersals

optimal distributions at each node:
node 17 (anc. of terminals ny_anecphya-ny_brachyceras): H  FH  AFH  BFH  ABFH  CFH  ACFH
    BCFH  ABCFH  EFH  AEFH  BEFH  ABEFH  CEFH  ACEFH  BCEFH  ABCEFH  FGH  AFGH  BFGH
    ABFGH  CFGH  ACFGH  BCFGH  ABCFGH  EFGH  AEFGH  BEFGH  ABEFGH  CEFGH  ACEFGH  BCEFGH
    ABCEFGH
node 18 (anc. of terminals ny_hydrocallis-ny_lotos): BD  CD  BCD  BDE  CDE  BCDE  BF  CF
    BCF  BDF  CDF  BCDF  BEF  CEF  BCEF  BDEF  CDEF  BCDEF  BDG  CDG  BCDG  BDEG  CDEG
    BCDEG  BFG  CFG  BCFG  BDFG  CDFG  BCDFG  BEFG  CEFG  BCEFG  BDEFG  CDEFG  BCDEFG
node 19 (anc. of terminals victoria-euryale): CD  CF  CDF  CDG  CFG  CDFG
node 20 (anc. of terminals ny_hydrocallis-euryale): D  F
node 21 (anc. of terminals ny_anecphya-euryale): F  DH  FH
node 22 (anc. of terminals ny_nymphaea-euryale): D  F
node 23 (anc. of terminals barclaya-euryale): DG  FG
node 24 (anc. of terminals nuphar-euryale): D  F
node 25 (anc. of terminals cabomba-brasenia): ABCEG  ABCDEG  ABCEFG  ABCDEFG  ABCEGH
    ABCDEGH  ABCEFGH  ABCDEFGH
node 26 (anc. of terminals nuphar-brasenia): ABCDEG  ABCEFG  ABCDEFG  ABCDEGH  ABCEFGH
    ABCDEFGH
node 27 (anc. of terminals kadsura-schisandra): D  F
node 28 (anc. of terminals illicium-schisandra): D  F
node 29 (anc. of terminals austrobaileya-schisandra): DH  FH
node 30 (anc. of terminals nuphar-schisandra): ABCDEFGH
node 31 (anc. of terminals amborella-schisandra): ABCDEFGHI
```

Appendix 8 (continued)

c) data file 3 (topology B, only extant distribution coded), written in DIVA–format

```
tree 1 = [&R]
(Amborella,(((Illicium,Schisandra),Austrobaileya),((Cabomba,Brasenia),(((Nuphar_japonica,N
uphar_lutea),(Nuphar_advenaN080,Nuphar_specN108)),(Barclaya,((Victoria,Euryale),(((Nymphae
a_alba_NY056,Nymphaea_candida_NY063),Nymphaea_odorata_N012),(((Nymphaea_amazonum_NY428,Nym
phaea_gardneriana_NY026),(((Nymphaea_lotus_therm_NY003,Nymphaea_lotus_lotus_NY078),Nymphae
a_pubescens_NY406),Nymphaea_petersiana_NY058)),((Nymphaea_gracilis_NY025,(Nymphaea_elegans
_NY006,Nymphaea_pulchella_NY100)),((Nymphaea_cf_nouchalii_NY066,Nymphaea_thermarum_NY065),
(((Nymphaea_hastifolia_NY134,(Nymphaea_violacea_NY110,Nymphaea_elleniae_NY137)),Nymphaea_v
iolacea_NY135),(Nymphaea_macrosperma_NY127,Nymphaea_gigantea_NY067)))))))))));

distribution
1       i
2       abfg
3       h
4       ah
5       abc
6       abcefgh
7       f
8       df
9       a
10      a
11      g
12      c
13      fg
14      df
15      df
16      ab
17      bc
18      c
19      e
20      e
21      g
22      e
23      a
24      a
25      abc
26      fg
27      e
28      h
29      h
30      h
31      h
32      h
33      h;

output 33B9_rezent2.txt;
echo status;
optimize maxareas=2;
output --;

output 33B9_rezent5.txt;
echo status;
optimize maxareas=5;
output --;

output 33B9_rezent9.txt;
echo status;
optimize maxareas=9;
output --;
```

results data file 3 – maxareas = 2

```
33b9_rezent2.txt opened as output file
echo status on
optimization successful - exact solution
        settings: maxareas=2, bound=250, hold=1000, weight=1.000, age=1.000
        optimal reconstruction requires 33 dispersals

optimal distributions at each node:
node 34 (anc. of terminals illicium-schisandra): AH
node 35 (anc. of terminals illicium-austrobaileya): A
node 36 (anc. of terminals cabomba-brasenia): A
```

```
node 37 (anc. of terminals nuphar_japonica-nuphar_lutea): F
node 38 (anc. of terminals nuphar_advenan08-nuphar_specn108): A
node 39 (anc. of terminals nuphar_japonica-nuphar_specn108): AF
node 40 (anc. of terminals victoria-euryale): CF  CG
node 41 (anc. of terminals nymphaea_alba_ny-nymphaea_candida): D  F
node 42 (anc. of terminals nymphaea_alba_ny-nymphaea_odorata): AD  BD  AF  BF
node 43 (anc. of terminals nymphaea_amazonu-nymphaea_gardner): C
node 44 (anc. of terminals nymphaea_lotus_t-nymphaea_lotus_l): E
node 45 (anc. of terminals nymphaea_lotus_t-nymphaea_pubesce): EG
node 46 (anc. of terminals nymphaea_lotus_t-nymphaea_petersi): E  EG
node 47 (anc. of terminals nymphaea_amazonu-nymphaea_petersi): CE  CG
node 48 (anc. of terminals nymphaea_elegans-nymphaea_pulchel): A
node 49 (anc. of terminals nymphaea_gracili-nymphaea_pulchel): A
node 50 (anc. of terminals nymphaea_cf_nouc-nymphaea_thermar): EF  EG
node 51 (anc. of terminals nymphaea_violace-nymphaea_ellenia): H
node 52 (anc. of terminals nymphaea_hastifo-nymphaea_ellenia): H
node 53 (anc. of terminals nymphaea_hastifo-nymphaea_violace): H
node 54 (anc. of terminals nymphaea_macrosp-nymphaea_gigante): H
node 55 (anc. of terminals nymphaea_hastifo-nymphaea_gigante): H
node 56 (anc. of terminals nymphaea_cf_nouc-nymphaea_gigante): EH  FH  GH
node 57 (anc. of terminals nymphaea_gracili-nymphaea_gigante): AE  AF  AG  AH
node 58 (anc. of terminals nymphaea_amazonu-nymphaea_gigante): AC  E  AE  CF  EF
     G  CG  EG
node 59 (anc. of terminals nymphaea_alba_ny-nymphaea_gigante): A  AE  BE  DE  F
     EF  AG  BG  DG  FG
node 60 (anc. of terminals victoria-nymphaea_gigante): AC  F  AF  G  AG  BG  DG
     EG  FG
node 61 (anc. of terminals barclaya-nymphaea_gigante): G  AG  FG
node 62 (anc. of terminals nuphar_japonica-nymphaea_gigante): A  F  AF  AG
node 63 (anc. of terminals cabomba-nymphaea_gigante): A  AF
node 64 (anc. of terminals illicium-nymphaea_gigante): A
node 65 (anc. of terminals amborella-nymphaea_gigante): AI
```

results data file 3 – maxareas = 5

```
33b9_rezent5.txt opened as output file
echo status on
optimization successful - exact solution
        settings: maxareas=5, bound=250, hold=1000, weight=1.000, age=1.000
        optimal reconstruction requires 32 dispersals

optimal distributions at each node:
node 34 (anc. of terminals illicium-schisandra): AH  BH  ABH  FH  AFH  BFH  ABFH  GH  AGH
     BGH  ABGH  FGH  AFGH  BFGH  ABFGH
node 35 (anc. of terminals illicium-austrobaileya): A  AB  ABF  AG  ABG  AFG  ABFG  H  AH
     BH  ABH  AFH  BFH  ABFH  GH  AGH  BGH  ABGH  FGH  AFGH  BFGH  ABFGH
node 36 (anc. of terminals cabomba-brasenia): A  B  AB  C  AC  BC  ABC  AE  BE  ABE  CE
     ACE  BCE  ABCE  AF  BF  ABF  CF  ACF  BCF  ABCF  AEF  BEF  ABEF  CEF  ACEF  BCEF
     ABCEF  AG  BG  ABG  CG  ACG  BCG  ABCG  AEG  BEG  ABEG  CEG  ACEG  BCEG  ABCEG  AFG
     BFG  ABFG  CFG  ACFG  BCFG  ABCFG  AEFG  BEFG  ABEFG  CEFG  ACEFG  BCEFG  BH  ABH  CH
     ACH  BCH  ABCH  AEH  BEH  ABEH  CEH  ACEH  BCEH  ABCEH  AFH  BFH  ABFH  CFH  ACFH
     BCFH  ABCFH  AEFH  BEFH  ABEFH  CEFH  ACEFH  BCEFH  AGH  BGH  ABGH  CGH  ACGH  BCGH
     ABCGH  AEGH  BEGH  ABEGH  CEGH  ACEGH  BCEGH  AFGH  BFGH  ABFGH  CFGH  ACFGH  BCFGH
     AEFGH  BEFGH  CEFGH
node 37 (anc. of terminals nuphar_japonica-nuphar_lutea): F
node 38 (anc. of terminals nuphar_advenan08-nuphar_specn108): A
node 39 (anc. of terminals nuphar_japonica-nuphar_specn108): AF
node 40 (anc. of terminals victoria-euryale): CF  CG  CFG
node 41 (anc. of terminals nymphaea_alba_ny-nymphaea_candida): D  F
node 42 (anc. of terminals nymphaea_alba_ny-nymphaea_odorata): AD  BD  ABD  AF  BF  ABF
node 43 (anc. of terminals nymphaea_amazonu-nymphaea_gardner): C
node 44 (anc. of terminals nymphaea_lotus_t-nymphaea_lotus_l): E
node 45 (anc. of terminals nymphaea_lotus_t-nymphaea_pubesce): EG
node 46 (anc. of terminals nymphaea_lotus_t-nymphaea_petersi): E  EG
node 47 (anc. of terminals nymphaea_amazonu-nymphaea_petersi): CE  CG  CEG
node 48 (anc. of terminals nymphaea_elegans-nymphaea_pulchel): A
node 49 (anc. of terminals nymphaea_gracili-nymphaea_pulchel): A
node 50 (anc. of terminals nymphaea_cf_nouc-nymphaea_thermar): EF  EG  EFG
node 51 (anc. of terminals nymphaea_violace-nymphaea_ellenia): H
node 52 (anc. of terminals nymphaea_hastifo-nymphaea_ellenia): H
node 53 (anc. of terminals nymphaea_hastifo-nymphaea_violace): H
node 54 (anc. of terminals nymphaea_macrosp-nymphaea_gigante): H
node 55 (anc. of terminals nymphaea_hastifo-nymphaea_gigante): H
node 56 (anc. of terminals nymphaea_cf_nouc-nymphaea_gigante): EH  FH  EFH  GH  EGH  FGH
     EFGH
node 57 (anc. of terminals nymphaea_gracili-nymphaea_gigante): AE  AF  AEF  AG  AEG  AFG
     AEFG  AH  AEH  AFH  AEFH  AGH  AEGH  AFGH  AEFGH
```

```
node 58 (anc. of terminals nymphaea_amazonu-nymphaea_gigante): AC  E  AE  ACE  CF  ACF  EF
     AEF  CEF  ACEF  G  CG  ACG  EG  AEG  CEG  ACEG  CFG  ACFG  EFG  AEFG  CEFG  ACEFG  ACH
     AEH  ACEH  CFH  ACFH  EFH  AEFH  CEFH  ACEFH  CGH  ACGH  EGH  AEGH  CEGH  ACEGH  CFGH
     ACFGH  EFGH  AEFGH  CEFGH
node 59 (anc. of terminals nymphaea_alba_ny-nymphaea_gigante): A  AE  BE  ABE  DE  ADE
     BDE  ABDE  F  EF  AEF  BEF  ABEF  AG  BG  ABG  ACG  BCG  ABCG  DG  ADG  BDG  ABDG  CDG
     ACDG  BCDG  ABCDG  AEG  BEG  ABEG  ACEG  BCEG  ABCEG  DEG  ADEG  BDEG  ABDEG  CDEG
     ACDEG  BCDEG  FG  AFG  BFG  ABFG  CFG  ACFG  BCFG  ABCFG  DFG  ADFG  BDFG  ABDFG  CDFG
     ACDFG  BCDFG  EFG  AEFG  BEFG  ABEFG  CEFG  ACEFG  BCEFG  DEFG  ADEFG  BDEFG  CDEFG
     AGH  BGH  ABGH  ACGH  BCGH  ABCGH  DGH  ADGH  BDGH  ABDGH  CDGH  ACDGH  BCDGH  AEGH
     BEGH  ABEGH  ACEGH  BCEGH  DEGH  ADEGH  BDEGH  CDEGH  FGH  AFGH  BFGH  ABFGH  CFGH
     ACFGH  BCFGH  DFGH  ADFGH  BDFGH  CDFGH  EFGH  AEFGH  BEFGH  CEFGH  DEFGH
node 60 (anc. of terminals victoria-nymphaea_gigante): AC  ABC  ACD  ABCD  ACE  ABCE  ACDE
     ABCDE  F  AF  ABF  ACF  ABCF  ADF  ABDF  ACDF  ABCDF  AEF  ABEF  ACEF  ABCEF  ADEF
     ABDEF  ACDEF  G  AG  BG  ABG  ACG  BCG  ABCG  DG  ADG  BDG  ABDG  CDG  ACDG  BCDG
     ABCDG  EG  AEG  BEG  ABEG  CEG  ACEG  BCEG  ABCEG  DEG  ADEG  BDEG  ABDEG  CDEG  ACDEG
     BCDEG  FG  AFG  BFG  ABFG  CFG  ACFG  BCFG  ABCFG  DFG  ADFG  BDFG  ABDFG  CDFG  ACDFG
     BCDFG  EFG  AEFG    BEFG  ABEFG  CEFG  ACEFG  BCEFG  DEFG  ADEFG  BDEFG  CDEFG
node 61 (anc. of terminals barclaya-nymphaea_gigante): G  AG  ABG  ACG  ABCG  ADG  ABDG
     ACDG  ABCDG  AEG  ABEG  ACEG  ABCEG  ADEG  ABDEG  ACDEG  FG  AFG  ABFG  ACFG  ABCFG
     ADFG  ABDFG  ACDFG  AEFG  ABEFG  ACEFG  ADEFG
node 62 (anc. of terminals nuphar_japonica-nymphaea_gigante): A  F  AF  AG  FG  AFG
node 63 (anc. of terminals cabomba-nymphaea_gigante): A  F  AF  BF  ABF  CF  ACF  BCF
     ABCF  AEF  BEF  ABEF  CEF  ACEF  BCEF  ABCEF  ABG  ACG  AFG  BFG  ABFG  CFG  ACFG
     BCFG  ABCFG  EFG  AEFG  BEFG  ABEFG  CEFG  ACEFG  BCEFG  BFH  ABFH  CFH  ACFH  BCFH
     ABCFH  EFH  AEFH  BEFH  ABEFH  CEFH  ACEFH  BCEFH  FGH  AFGH  BFGH  ABFGH  CFGH  ACFGH
     BCFGH  EFGH  AEFGH  BEFGH  CEFGH
node 64 (anc. of terminals illicium-nymphaea_gigante): ABCE  ABF  ACF  ABCF  ABEF  ACEF
     ABCEF  ABCG  ABEG  ACEG  ABCEG  ABFG  ACFG  BCFG  ABCFG  AEFG  ABEFG  ACEFG  ABCH
     ABEH  ACEH  BCEH  ABCEH  AFH  BFH  ABFH  CFH  ACFH  BCFH  ABCFH  AEFH  BEFH  ABEFH
     CEFH  ACEFH  BCEFH  ABGH  ACGH  BCGH  ABCGH  AEGH  BEGH  ABEGH  CEGH  ACEGH  BCEGH
     AFGH  BFGH  ABFGH  CFGH  ACFGH  BCFGH  EFGH  AEFGH  BEFGH  CEFGH
node 65 (anc. of terminals amborella-nymphaea_gigante): ABCEI  ABFI  ACFI  ABCFI  ABEFI
     ACEFI  BCEFI  ABCGI  ABEGI  ACEGI  BCEGI  ABFGI  ACFGI  BCFGI  AEFGI  BEFGI  CEFGI
     ABCHI  ABEHI  ACEHI  BCEHI  AFHI  BFHI  ABFHI  CFHI  ACFHI  BCFHI  AEFHI  BEFHI  CEFHI
     ABGHI  ACGHI  BCGHI  AEGHI  BEGHI  CEGHI  AFGHI  BFGHI  CFGHI  EFGHI
```

results data file 3 – maxareas = 9

```
33b9_rezent9.txt opened as output file
echo status on
optimization successful - exact solution
     settings: maxareas=9, bound=250, hold=1000, weight=1.000, age=1.000
     optimal reconstruction requires 29 dispersals

optimal distributions at each node:
node 34 (anc. of terminals illicium-schisandra): AH  BH  ABH  FH  AFH  BFH  ABFH  GH  AGH
     BGH  ABGH  FGH  AFGH  BFGH  ABFGH
node 35 (anc. of terminals illicium-austrobaileya): A  H
node 36 (anc. of terminals cabomba-brasenia): A  B  C  ABCEG  ABCEFG  BCEGH  ABCEGH
     BCEFGH  ABCEFGH
node 37 (anc. of terminals nuphar_japonica-nuphar_lutea): F
node 38 (anc. of terminals nuphar_advenan08-nuphar_specn108): A
node 39 (anc. of terminals nuphar_japonica-nuphar_specn108): AF
node 40 (anc. of terminals victoria-euryale): CF  CG  CFG
node 41 (anc. of terminals nymphaea_alba_ny-nymphaea_candida): D  F
node 42 (anc. of terminals nymphaea_alba_ny-nymphaea_odorata): AD  BD  ABD  AF  BF  ABF
node 43 (anc. of terminals nymphaea_amazonu-nymphaea_gardner): C
node 44 (anc. of terminals nymphaea_lotus_t-nymphaea_lotus_l): E
node 45 (anc. of terminals nymphaea_lotus_t-nymphaea_pubesce): EG
node 46 (anc. of terminals nymphaea_lotus_t-nymphaea_petersi): E
node 47 (anc. of terminals nymphaea_amazonu-nymphaea_petersi): CE
node 48 (anc. of terminals nymphaea_elegans-nymphaea_pulchel): A
node 49 (anc. of terminals nymphaea_gracili-nymphaea_pulchel): A
node 50 (anc. of terminals nymphaea_cf_nouc-nymphaea_thermar): EF  EG  EFG
node 51 (anc. of terminals nymphaea_violace-nymphaea_ellenia): H
node 52 (anc. of terminals nymphaea_hastifo-nymphaea_ellenia): H
node 53 (anc. of terminals nymphaea_hastifo-nymphaea_violace): H
node 54 (anc. of terminals nymphaea_macrosp-nymphaea_gigante): H
node 55 (anc. of terminals nymphaea_hastifo-nymphaea_gigante): H
node 56 (anc. of terminals nymphaea_cf_nouc-nymphaea_gigante): EH  FH  EFH  EGH  FGH  EFGH
node 57 (anc. of terminals nymphaea_gracili-nymphaea_gigante): AE  AF  AEF  AEG  AFG  AEFG
     AEH  AFH  AEFH  AEGH  AFGH  AEFGH
node 58 (anc. of terminals nymphaea_amazonu-nymphaea_gigante): E  CF  ACF  EF  AEF  CEF
     ACEF  CFG  ACFG  EFG  AEFG  CEFG  ACEFG  CFH  ACFH  EFH  AEFH  CEFH  ACEFH  CFGH
     ACFGH  EFGH  AEFGH  CEFGH  ACEFGH
```

```
node 59 (anc. of terminals nymphaea_alba_ny-nymphaea_gigante): DE  ADE  BDE  ABDE  F  EF
     AEF  BEF  ABEF
node 60 (anc. of terminals victoria-nymphaea_gigante): CDE  ACDE  BCDE  ABCDE  F  BDEF
     ABDEF  CDEF  ACDEF  BCDEF  ABCDEF  BDEG  ABDEG  CDEG  ACDEG  BCDEG  ABCDEG  BDEFG
     ABDEFG  CDEFG  ACDEFG  BCDEFG  ABCDEFG
node 61 (anc. of terminals barclaya-nymphaea_gigante): BDEG  ABDEG  CDEG  ACDEG  BCDEG
     ABCDEG  FG  BDEFG  ABDEFG  CDEFG  ACDEFG  BCDEFG  ABCDEFG
node 62 (anc. of terminals nuphar_japonica-nymphaea_gigante): F  ABDEFG  ACDEFG  BCDEFG
     ABCDEFG
node 63 (anc. of terminals cabomba-nymphaea_gigante): ABCEFG  ABCDEFG  BCEFGH  ABCEFGH
node 64 (anc. of terminals illicium-nymphaea_gigante): ABCEFGH  ABCDEFGH
node 65 (anc. of terminals amborella-nymphaea_gigante): ABCEFGHI  ABCDEFGHI
```

Appendix 8 (continued)

d) data file 4 (topology B, extant and fossil distribution coded), written in DIVA format

```
tree 1 = [&R]
     (Amborella,(((Illicium,Schisandra),Austrobaileya),((Cabomba,Brasenia),(((Nuphar_japoni
     ca,Nuphar_lutea),(Nuphar_advenaN080,Nuphar_specN108)),(Barclaya,((Victoria,Euryale),((
     (Nymphaea_alba_NY056,Nymphaea_candida_NY063),Nymphaea_odorata_N012),(((Nymphaea_amazon
     um_NY428,Nymphaea_gardneriana_NY026),(((Nymphaea_lotus_therm_NY003,Nymphaea_lotus_lotu
     s_NY078),Nymphaea_pubescens_NY406),Nymphaea_petersiana_NY058)),((Nymphaea_gracilis_NY0
     25,(Nymphaea_elegans_NY006,Nymphaea_pulchella_NY100)),((Nymphaea_cf_nouchalii_NY066,Ny
     mphaea_thermarum_NY065),(((Nymphaea_hastifolia_NY134,(Nymphaea_violacea_NY110,Nymphaea
     _elleniae_NY137)),Nymphaea_violacea_NY135),(Nymphaea_macrosperma_NY127,Nymphaea_gigant
     ea_NY067))))))))))));

distribution
1    i
2    abdfg
3    adf
4    h
5    abc
6    abcdefgh
7    f
8    df
9    a
10   a
11   g
12   c
13   dfg
14   df
15   df
16   ab
17   bc
18   c
19   d
20   e
21   g
22   e
23   a
24   a
25   abc
26   fg
27   e
28   h
29   h
30   h
31   h
32   h
33   h;

output 33B9_fossil2.txt;
echo status;
optimize maxareas=2;
output --;

output 33B9_fossil5.txt;
echo status;
optimize maxareas=5;
output --;

output 33B9_fossil9.txt;
echo status;
optimize maxareas=9;
output --;
```

results data file 4 – maxareas = 2

```
33b9_fossil2.txt opened as output file
echo status on
optimization successful - exact solution
     settings: maxareas=2, bound=250, hold=1000, weight=1.000, age=1.000
     optimal reconstruction requires 38 dispersals

optimal distributions at each node:
node 34 (anc. of terminals illicium-schisandra): A  F
node 35 (anc. of terminals illicium-austrobaileya): AH  FH
```

```
node 36 (anc. of terminals cabomba-brasenia): A  B  C  AF  BF  CF
node 37 (anc. of terminals nuphar_japonica-nuphar_lutea): F
node 38 (anc. of terminals nuphar_advenan08-nuphar_specn108): A
node 39 (anc. of terminals nuphar_japonica-nuphar_specn108): AF
node 40 (anc. of terminals victoria-euryale): CD  CF  CG
node 41 (anc. of terminals nymphaea_alba_ny-nymphaea_candida): D  F
node 42 (anc. of terminals nymphaea_alba_ny-nymphaea_odorata): AD  BD  AF  BF
node 43 (anc. of terminals nymphaea_amazonu-nymphaea_gardner): C
node 44 (anc. of terminals nymphaea_lotus_t-nymphaea_lotus_l): DE
node 45 (anc. of terminals nymphaea_lotus_t-nymphaea_pubesce): DG  EG
node 46 (anc. of terminals nymphaea_lotus_t-nymphaea_petersi): E  EG
node 47 (anc. of terminals nymphaea_amazonu-nymphaea_petersi): CE  CG
node 48 (anc. of terminals nymphaea_elegans-nymphaea_pulchel): A
node 49 (anc. of terminals nymphaea_gracili-nymphaea_pulchel): A
node 50 (anc. of terminals nymphaea_cf_nouc-nymphaea_thermar): EF  EG
node 51 (anc. of terminals nymphaea_violace-nymphaea_ellenia): H
node 52 (anc. of terminals nymphaea_hastifo-nymphaea_ellenia): H
node 53 (anc. of terminals nymphaea_hastifo-nymphaea_violace): H
node 54 (anc. of terminals nymphaea_macrosp-nymphaea_gigante): H
node 55 (anc. of terminals nymphaea_hastifo-nymphaea_gigante): H
node 56 (anc. of terminals nymphaea_cf_nouc-nymphaea_gigante): EH  FH  GH
node 57 (anc. of terminals nymphaea_gracili-nymphaea_gigante): AE  AF  AG  AH
node 58 (anc. of terminals nymphaea_amazonu-nymphaea_gigante): AC  E  AE  CF  EF G  CG  EG
node 59 (anc. of terminals nymphaea_alba_ny-nymphaea_gigante): A  AE  BE  DE  F  EF  AG
    BG  DG  FG
node 60 (anc. of terminals victoria-nymphaea_gigante): AC  D  AD  F AF G AG BG  DG  EG  FG
node 61 (anc. of terminals barclaya-nymphaea_gigante): G  AG  DG  FG
node 62 (anc. of terminals nuphar_japonica-nymphaea_gigante): A  AD  F  AF  AG
node 63 (anc. of terminals cabomba-nymphaea_gigante): A  F  AF  BF  CF
node 64 (anc. of terminals illicium-nymphaea_gigante): A  F
node 65 (anc. of terminals amborella-nymphaea_gigante): AI  FI
```

results data file 4 – maxareas = 5

```
33b9_fossil5.txt opened as output file
echo status on
optimization successful - exact solution
    settings: maxareas=5, bound=250, hold=1000, weight=1.000, age=1.000
    optimal reconstruction requires 36 dispersals

optimal distributions at each node:
node 34 (anc. of terminals illicium-schisandra): A  D
node 35 (anc. of terminals illicium-austrobaileya): AH  DH
node 36 (anc. of terminals cabomba-brasenia): A  B  C
node 37 (anc. of terminals nuphar_japonica-nuphar_lutea): F
node 38 (anc. of terminals nuphar_advenan08-nuphar_specn108): A
node 39 (anc. of terminals nuphar_japonica-nuphar_specn108): AF
node 40 (anc. of terminals victoria-euryale): CF  CDF  CFG  CDFG
node 41 (anc. of terminals nymphaea_alba_ny-nymphaea_candida): F
node 42 (anc. of terminals nymphaea_alba_ny-nymphaea_odorata): AF  BF  ABF
node 43 (anc. of terminals nymphaea_amazonu-nymphaea_gardner): C
node 44 (anc. of terminals nymphaea_lotus_t-nymphaea_lotus_l): DE
node 45 (anc. of terminals nymphaea_lotus_t-nymphaea_pubesce): EG  DEG
node 46 (anc. of terminals nymphaea_lotus_t-nymphaea_petersi): E
node 47 (anc. of terminals nymphaea_amazonu-nymphaea_petersi): CE
node 48 (anc. of terminals nymphaea_elegans-nymphaea_pulchel): A
node 49 (anc. of terminals nymphaea_gracili-nymphaea_pulchel): A
node 50 (anc. of terminals nymphaea_cf_nouc-nymphaea_thermar): EF  EG  EFG
node 51 (anc. of terminals nymphaea_violace-nymphaea_ellenia): H
node 52 (anc. of terminals nymphaea_hastifo-nymphaea_ellenia): H
node 53 (anc. of terminals nymphaea_hastifo-nymphaea_violace): H
node 54 (anc. of terminals nymphaea_macrosp-nymphaea_gigante): H
node 55 (anc. of terminals nymphaea_hastifo-nymphaea_gigante): H
node 56 (anc. of terminals nymphaea_cf_nouc-nymphaea_gigante): EH  FH  EFH  EGH  FGH  EFGH
node 57 (anc. of terminals nymphaea_gracili-nymphaea_gigante): AE  AF  AEF  AEG  AFG  AEFG
    AEH  AFH  AEFH  AEGH  AFGH  AEFGH
node 58 (anc. of terminals nymphaea_amazonu-nymphaea_gigante): E  CF  ACF  EF  AEF  CEF
    ACEF  CFG  ACFG  EFG  AEFG  CEFG  ACEFG  CFH  ACFH  EFH  AEFH  CEFH  ACEFH  CFGH
    ACFGH  EFGH  AEFGH  CEFGH
node 59 (anc. of terminals nymphaea_alba_ny-nymphaea_gigante): F  EF  AEF  BEF  ABEF
node 60 (anc. of terminals victoria-nymphaea_gigante): F
node 61 (anc. of terminals barclaya-nymphaea_gigante): FG
node 62 (anc. of terminals nuphar_japonica-nymphaea_gigante): F
node 63 (anc. of terminals cabomba-nymphaea_gigante): AF  BF  CF
node 64 (anc. of terminals illicium-nymphaea_gigante): ABFH  ACFH  ADFH  BDFH  CDFH
node 65 (anc. of terminals amborella-nymphaea_gigante): ABFHI  ACFHI  ADFHI  BDFHI  CDFHI
```

results data file 4 – maxareas = 9

```
echo status on
optimization successful - exact solution
     settings: maxareas=9, bound=250, hold=1000, weight=1.000, age=1.000
     optimal reconstruction requires 33 dispersals

optimal distributions at each node:
node 34 (anc. of terminals illicium-schisandra): A  D  F
node 35 (anc. of terminals illicium-austrobaileya): AH  DH  FH
node 36 (anc. of terminals cabomba-brasenia): ABCEG  BCDEG  ABCDEG  ABCEFG
     BCDEFG  ABCDEFG  ABCEGH  BCDEGH  ABCDEGH  ABCEFGH  BCDEFGH  ABCDEFGH
node 37 (anc. of terminals nuphar_japonica-nuphar_lutea): F
node 38 (anc. of terminals nuphar_advenan08-nuphar_specn108): A
node 39 (anc. of terminals nuphar_japonica-nuphar_specn108): AF
node 40 (anc. of terminals victoria-euryale): CF  CDF  CFG  CDFG
node 41 (anc. of terminals nymphaea_alba_ny-nymphaea_candida): F
node 42 (anc. of terminals nymphaea_alba_ny-nymphaea_odorata): AF  BF  ABF
node 43 (anc. of terminals nymphaea_amazonu-nymphaea_gardner): C
node 44 (anc. of terminals nymphaea_lotus_t-nymphaea_lotus_l): DE
node 45 (anc. of terminals nymphaea_lotus_t-nymphaea_pubesce): EG  DEG
node 46 (anc. of terminals nymphaea_lotus_t-nymphaea_petersi): E
node 47 (anc. of terminals nymphaea_amazonu-nymphaea_petersi): CE
node 48 (anc. of terminals nymphaea_elegans-nymphaea_pulchel): A
node 49 (anc. of terminals nymphaea_gracili-nymphaea_pulchel): A
node 50 (anc. of terminals nymphaea_cf_nouc-nymphaea_thermar): EF  EG  EFG
node 51 (anc. of terminals nymphaea_violace-nymphaea_ellenia): H
node 52 (anc. of terminals nymphaea_hastifo-nymphaea_ellenia): H
node 53 (anc. of terminals nymphaea_hastifo-nymphaea_violace): H
node 54 (anc. of terminals nymphaea_macrosp-nymphaea_gigante): H
node 55 (anc. of terminals nymphaea_hastifo-nymphaea_gigante): H
node 56 (anc. of terminals nymphaea_cf_nouc-nymphaea_gigante): EH  FH  EFH  EGH  FGH  EFGH
node 57 (anc. of terminals nymphaea_gracili-nymphaea_gigante): AE  AF  AEF  AEG  AFG  AEFG
     AEH  AFH  AEFH  AEGH  AFGH  AEFGH
node 58 (anc. of terminals nymphaea_amazonu-nymphaea_gigante): E  CF  ACF  EF  AEF  CEF
     ACEF  CFG  ACFG  EFG  AEFG  CEFG  ACEFG  CFH  ACFH  EFH  AEFH  CEFH  ACEFH  CFGH
     ACFGH  EFGH  AEFGH  CEFGH  ACEFGH
node 59 (anc. of terminals nymphaea_alba_ny-nymphaea_gigante): F  EF  AEF  BEF  ABEF
node 60 (anc. of terminals victoria-nymphaea_gigante): F
node 61 (anc. of terminals barclaya-nymphaea_gigante): FG
node 62 (anc. of terminals nuphar_japonica-nymphaea_gigante): F
node 63 (anc. of terminals cabomba-nymphaea_gigante): ABCEFG  BCDEFG  ABCDEFG  ABCEFGH
     BCDEFGH  ABCDEFGH
node 64 (anc. of terminals illicium-nymphaea_gigante): ABCDEFGH
node 65 (anc. of terminals amborella-nymphaea_gigante): ABCDEFGHI
```

Hiermit versichere ich, dass ich die vorliegende Dissertation selbständig angefertigt und die benutzten Quellen und Hilfsmittel vollständig angegeben habe. Diese Arbeit wurde an keiner anderen Hochschule als Dissertation eingereicht und wurde, abgesehen von der im Folgenden angegebenen Teilpublikation, noch nicht veröffentlicht.

Kapitel 2 wurde wie folgt veröffentlicht:

> Löhne C. & T. Borsch (2005). Molecular evolution and phylogenetic utility of the *petD* group II intron: A case study in basal angiosperms. *Molecular Biology and Evolution* 22: 317–332.

Folgende Veröffentlichungen sind in Vorbereitung:

> Löhne C., J. H. Wiersema, & T. Borsch (subm.). Phylogenetic analysis of Nymphaeales using fast–evolving and non–coding chloroplast markers. S*ubmitted to: Botanical Journal of the Linnean Society*.
>
> Löhne C., T. Borsch, S. W. L. Jacobs, C. B. Hellquist, & J. H. Wiersema (in prep.). Nuclear and plastid DNA sequences reveal complex evolutionary patterns in Australian water–lilies (*Nymphaea* subgenus *Anecphya*). *To be submitted to: Australian Systematic Botany*
>
> Löhne C., M.–J. Yoo, T. Borsch, C. D. Bell, J. H. Wiersema, V. Wilde, P. S. Soltis, D. E. Soltis & W. Barthlott (accepted). Biogeography of Nymphaeales: Extant patterns and historical events. In: Nymphaeales — the first globally diverse clade of angiosperms. (Borsch T. & P. S. Soltis, Eds.) *Taxon Special Volume*.

Bonn, 23. März 2006 _____