Int. J. Plant Sci. 168(5):639–671. 2007.

1058-5893/2007/16805-0012$15.00

PHYLOGENY OF *NYMPHAEA* (NYMPHAEACEAE): EVIDENCE FROM SUBSTITUTIONS AND MICROSTRUCTURAL CHANGES IN THE CHLOROPLAST *trnT-trnF* REGION

Thomas Borsch,[1,*] Khidir W. Hilu,† John H. Wiersema,‡ Cornelia Löhne,* Wilhelm Barthlott,* and Volker Wilde§

*Nees-Institut für Biodiversität der Pflanzen, Universität Bonn, Meckenheimer Allee 170, 53115 Bonn, Germany; †Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia 24061, U.S.A.; ‡USDA, Agricultural Research Service, Building 011A, BARC-West, Beltsville, Maryland 20705, U.S.A.; and §Sektion Paläobotanik, Forschungsinstitut Senckenberg, Senckenberganlage 25, 60325 Frankfurt am Main, Germany

Nymphaea is the most speciose, phenotypically diverse, and geographically widespread (nearly global) genus of Nymphaeales. Phylogenetic relationships among 35 of an estimated 45–50 species of *Nymphaea* are presented based on an analysis of the chloroplast *trnT-trnF* region. Because this is the first phylogenetic analysis of *Nymphaea*, monophyly of the genus had to be tested, and its status in Nymphaeales had to be inferred. Rooting was therefore extended to more distant outgroups (*Amborella*, Austrobaileyales). Monophyly of *Nymphaea* received weak support, with a *Euryale-Victoria* clade appearing as sister. The three major lineages within *Nymphaea* are constituted by the northern temperate subg. *Nymphaea* that is sister to all remaining species, a subgg. *Hydrocallis-Lotos* clade, and a subgg. *Anecphya-Brachyceras* clade. The Australian genus *Ondinea* was nested at species level within *Nymphaea* subg. *Anecphya*. The pantropical subg. *Brachyceras* as currently circumscribed does not appear natural, with *Nymphaea petersiana* belonging to subg. *Lotos*. Microstructural changes are frequent and highly informative, exhibiting lower levels of homoplasy than substitutions. Reconstructing the evolution of microstructural changes shows a strong insertion bias in simple sequence repeats. Complex indels are often explained by mutational events that occurred independently in different parts of the tree rather than being the result of stepwise events at subsequent nodes. AT-rich, satellite-like sequence parts have evolved independently in the P8 stem loop of the *trnL* group I intron in *Nuphar* and in major lineages of *Nymphaea*. They seem to be conserved in sequence within species but are highly variable among species. Moreover, the *trnT-trnF* region provides a signal that allows recognition (bar coding) of most species analyzed so far.

Keywords: chloroplast genome, molecular evolution, *trnL* group I intron, basal angiosperms, *Nymphaea*, *Ondinea*.

Introduction

The water lily genus *Nymphaea* is the most speciose member of the order Nymphaeales. The monophyletic herbaceous Nymphaeales have been proposed as sister to all angiosperms after the New Caledonian endemic shrub *Amborella* based on evidence from multiple genes (Mathews and Donoghue 1999; Qiu et al. 1999; Soltis et al. 1999; Graham and Olmstead 2000; Zanis et al. 2002; Hilu et al. 2003) and rapidly evolving spacers and introns (Borsch et al. 2003, 2005; Löhne and Borsch 2005). An alternative hypothesis suggesting that a clade consisting of *Amborella* and Nymphaeales is sister to all other angiosperms (Barkman et al. 2000; Mathews and Donoghue 2000) has been favored by recent likelihood-based inferences (Leebens-Mack et al. 2005; Qiu et al. 2005). A consensus has been achieved for Austrobaileyales (Austrobaileyaceae, Illiciaceae, Schisandraceae, Trimeniaceae) as the next successive sister to the remaining angiosperms after Nymphaeales and *Amborella* (Zanis et al. 2002; Borsch et al. 2003, 2005; Hilu et al. 2003; Qiu et al. 2005). Among all these extant taxa, the Nymphaeales and *Nymphaea* are the only globally diverse group.

[1] Author for correspondence; telephone 49-228-732681; fax 49-228-733120; e-mail borsch@uni-bonn.de.

Manuscript received May 2006; revised manuscript received October 2006.

The Nymphaeales are now generally divided into two families (Schneider and Williamson 1993; Williamson and Schneider 1993), the Cabombaceae (*Brasenia* and *Cabomba*) and the Nymphaeaceae (*Barclaya*, *Euryale*, *Nuphar*, *Nymphaea*, *Ondinea*, and *Victoria*), with a total of ca. 70 species. Ito's (1987) phylogenetic analysis of Nymphaeales included *Nelumbo* and *Ceratophyllum*, based on morphological characters. In his tree, *Nymphaea* is sister to a *Euryale* plus *Victoria* clade. Using floral vasculature characters, Moseley et al. (1993) found support for the monophyly of the clade including *Nymphaea*, *Euryale*, and *Victoria*. The earliest molecular phylogenetic analyses (Les et al. 1991, 1999) found a well-supported topology of *Nuphar* (*Barclaya* (*Ondinea* (*Nymphaea* (*Euryale* + *Victoria*)))). All of these studies included only a single species of *Nymphaea* (*Nymphaea odorata*), thus leaving monophyly and relationships of the genus *Nymphaea* as open questions. Löhne and Borsch (2005) further added *N. micrantha* and *N. heudelotii* in an analysis of the *petD* data set for Nymphaeales, but these species appeared in an unresolved position with *Victoria*.

Nymphaea occurs almost worldwide, comprising 45–50 species in five subgenera (*Anecphya* [seven to 10 species], *Brachyceras* [14–16 species], *Hydrocallis* [14 species], *Lotos* [two to three species], and *Nymphaea* [eight species]). Each subgenus displays a characteristic distribution. *Nymphaea* subg. *Nymphaea* ranges throughout the Northern Hemisphere in temperate

regions, subg. *Lotos* is paleotropical, and subg. *Hydrocallis* is restricted to the Neotropics. *Nymphaea* subg. *Brachyceras* has a pantropical range, and subg. *Anecphya* is restricted to Australia and New Guinea. This widely accepted classification into five subgenera traces to Conard's (1905) monograph of *Nymphaea*. Conard grouped subgg. *Anecphya* and *Brachyceras* as Apocarpiae, on the basis of carpel walls that are only partially fused. He included subgg. *Hydrocallis*, *Lotos*, and *Nymphaea* in the Syncarpiae, with more complete carpel wall fusion. Conard's Apocarpiae and Syncarpiae reflect sections *Leptopleura* and *Symphytopleura*, previously published by Caspary (1865, 1888).

As indicated above, phenotypic characters have been used to infer relationships among genera of Nymphaeales (Ito 1987; Moseley et al. 1993; Les et al. 1999). However, the considerable differences in phenotypic characters present among subgenera within *Nymphaea* have never been analyzed in order to test monophyly of the genus. Differences among species of *Nymphaea* have been pointed out in floral morphology and anatomy (Caspary 1865, 1888; Conard 1905), capillary appendages (Wiersema 1988; Schneider and Williamson 1993), pollen morphology (Gabarayeva and El-Ghazaly 1997; Borsch 2000; Hesse and Zetter 2005), seed morphology (Weberbauer 1894; Collinson 1980; Wiersema 1987), and vegetative anatomy (Weidlich 1976*a*, 1976*b*). Species of subgg. *Anecphya*, *Brachyceras*, and *Nymphaea*, as well as *Ondinea* (Schneider 1983), are diurnally flowering, whereas subgg. *Hydrocallis*, *Lotos*, and *Victoria* are nocturnally flowering (Prance and Arias 1975; Wiersema 1988). Nocturnal flowering is associated with beetle pollination (Ervik and Knudsen 2003; Hirthe and Porembski 2003) and diurnal flowering with a variety of different pollinators, including hymenopterans, dipterans, and coleopterans (Wiersema 1988 and references therein). Flowers in *Euryale* have been reported as being predominantly cleistogamous in some populations (Kadono and Schneider 1987), which may be a derived feature. It remains to be tested whether the two predominant pollination syndromes (diurnal vs. nocturnal flowering) mark two evolutionary lineages, in line with a paraphyletic genus *Nymphaea*, or whether nocturnal flowering arose independently in subgg. *Hydrocallis*, *Lotus*, and *Victoria*.

Polyploidy plays an important role in *Nymphaea* evolution (Gupta 1980), although the state of knowledge is rather scarce and is based mostly on earlier studies (Langlet and Söderberg 1929; Gupta 1978, 1980; Okada and Tamura 1981). Chromosome counts indicate a base number of x=14 for the genus, with polyploidy evident in all subgenera, especially subgg. *Anecphya* (2n=224), which lacks counts for most species; *Brachyceras* (2n=28, 56, 84), with most species still uncounted; *Nymphaea* (2n=56, 84, 112), with counts for most species; and *Lotos* (2n=28, 56, 84), with all species counted. While diploids occur rarely in other subgenera, they are common in subg. *Hydrocallis* (2n=18, 20, 28, 42, 84), where most species are diploid (Wiersema 1987). Somatic counts for several species of *Nuphar* (2n=34), two species of *Barclaya* (2n=36), *Euryale* (2n=58), and both species of *Victoria* (2n=20, 24) indicate a wide range of base numbers in these other genera.

The fossil record of water lilies includes predominantly pollen, seeds, and leaves, many of which have been described in a number of form genera. Fossil remains that can be clearly assigned to one of the lineages of *Nymphaea* appear to be comparatively young, such as seeds in Upper Eocene/Lower Oligocene strata from England (Collinson 1980), whereas seeds from the Middle Eocene of Canada have been compared with *Victoria* (Cevalloz-Ferriz and Stockey 1989) and seeds of *Nuphar* unambiguously date back to the Early Eocene of China (Chen et al. 2004). The discussion on the fossil history of Nymphaeales has been stimulated by the discovery of small flowers from the Lower Cretaceous of Portugal, considered to be the earliest water lilies (Friis et al. 2001). Other small flowers from the Turonian (Upper Cretaceous) of New Jersey have been regarded as Nymphaeaceae close to *Victoria* and *Euryale* (*Microvictoria*; Gandolfo et al. 2004).

Analyses of basal angiosperms have consistently found a long branch leading to the crown group of Nymphaeales (Qiu et al. 1999, 2005; Zanis et al. 2002; Borsch et al. 2003), supported by a large number of synapomorphic indels (Löhne and Borsch 2005). Resolving relationships within this crown group presented a rather difficult setting because of the genetic distance of potential outgroups such as *Amborella* and the possibility of long-branch attraction through the outgroup. This situation may have motivated the use of Cabombaceae to root Nymphaeaceae in earlier analyses (Les et al. 1999) of relationships within the clade. Yoo et al. (2005) provided molecular clock evidence for a diversification of the Nymphaeales crown group during the Eocene (44.6 ± 7.9 Ma), much more recent than the divergence of the stem lineage (estimated to 125–115 Ma).

Borsch et al. (2003) applied sequence data of the two spacers and the group I intron in the chloroplast *trnT-trnF* region to phylogenetic analyses of basal angiosperms. It was shown that extreme length-mutational dynamics resulting in difficult-to-align sequence parts at greater genetic distances is confined to certain hotspots. In the *trnL* intron, these mutational hotspots are located within the P6 and P8 stem loops, which are structurally least confined (Borsch et al. 2003; Quandt et al. 2004). The resulting *trnT-trnF* phylogeny was well resolved and highly supported. The *trnT-trnF* region was therefore a promising molecular marker that, on the one hand, is alignable with *Amborella* and Austrobaileyales and, on the other, is variable enough to provide resolution among species of *Nymphaea*.

The *trnT-trnF* region has become one of the most widely used regions in plants since the availability of universal primers annealing to the t-RNA genes (Taberlet et al. 1991). Initially, *trnT-trnF* sequences were used for analyzing relationships among species (Mes and 't Hart 1994; Gielly et al. 1996) and genera (van Ham et al. 1994; Bayer and Starr 1998). More recently, the region appeared informative for inferring relationships among families (Renner 1999; Sauquet et al. 2003) and major lineages of angiosperms and land plants (Borsch et al. 2003; Quandt et al. 2004). However, in most cases, only the *trnL* intron and the *trnL-trnF* spacer had been sequenced, whereas analyses involving the whole *trnT-trnF* region are comparatively few (e.g., Böhle et al. 1994; Won and Renner 2005). The *trnL-trnF* spacer and the *trnL* intron appear as cotranscribed in land plants (Kanno and Hirai 1993; Quandt et al. 2004; Won and Renner 2005). The *trnT-trnL* spacer is not transcribed and seems to evolve slightly differently. Based on a character resampling approach, the *trnT-trnF* region was recently shown to comprise more phylogenetic structure per informative character than *matK* (Müller et al. 2006).

Aims of this study are twofold. First, we intend to reconstruct phylogenetic relationships of the genus *Nymphaea* using a dense taxon sampling. In this context, it is of particular importance to

test the monophyly of *Nymphaea* as currently circumscribed and to evaluate its position within Nymphaeales. Close and distant outgroups shall be examined for possible effects. Second, we aim to explore fully the information content of the *trnT-trnF* region, including information from microstructural changes, and to assess the relative performance of individual spacers, the *trnL* intron, and the satellite-like region within the *trnL* intron.

Material and Methods

Taxon and Character Sampling

Most specimens were collected from wild populations (table C1). If available, widespread species were represented by individuals from several geographically distant populations in order to test the utility of satellite-like regions in the P8 stem loop of the *trnL* intron for inferring species relationships and for species identification. Before DNA isolation, young leaves were washed in distilled water to remove algae and other periphyton and were subsequently dried in silica gel. About three-fourths of the species of *Nymphaea* (35 out of 45–50), covering all five subgenera (fig. 1), and representatives of all other Nymphaeales genera were sampled.

Both distant and close outgroups were selected to test relationships within *Nymphaea* as well as within the Nymphaeales as a whole. A series of phylogenetic analyses was conducted using three taxon sets: set A comprised all Nymphaeales (61 taxa), with Austrobaileyales (three taxa) plus *Amborella* as outgroup; set B comprised all Nymphaeales and was rooted with *Amborella* alone; and set C comprised only the core of Nymphaeaceae (*Barclaya*, *Euryale*, *Nymphaea*, *Ondinea*, *Victoria*) and used *Brasenia*, *Cabomba*, and four species of *Nuphar* as outgroup. For each taxon set, character sets were successively added to the substitution-based matrix (matrix 1) to test their influence on the inferred trees; matrix 2 was composed of substitutions and indels (excluding P8), matrix 3 of P8 data (substitutions and indels), and matrix 4 of all characters.

DNA Isolation, Amplification, and Sequencing

Total genomic DNA was isolated from dried leaf tissue using a modified CTAB method for optimal yield of DNA, as described by Borsch et al. (2003). The *trnT-trnF* region was polymerase chain reaction (PCR) amplified in two overlapping halves with universal primers. Amplification of the upstream half used primers rps4-5F, annealing to the *rps4* gene upstream of *trnT* (Sauquet et al. 2003), and trnL110R (Borsch et al. 2003), annealing at a site 16 nucleotides (nt) upstream of the P element in the *trnL* intron that is length conserved in angiosperms. The downstream half was amplified using primers c and f (Taberlet et al. 1991). These amplification primers were also used for sequencing, and when reads were not long enough, primers a, d, and e, designed by Taberlet et al. (1991), also were employed. This strategy enabled the generation of complete and reliable sequences of both spacers and the *trnL* gene, including its intron. Amplification conditions were as follows: 34 cycles of 94°C (1 min) denaturation, 52°C (1 min) annealing, 72°C (2 min) extension, and 72°C (15 min) final extension. PCR products were purified using a QiaQuick gel extraction kit (Qiagen, Valencia, CA) and directly sequenced with ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kits, version 1.1 or 3.1 (Applied Biosystems, Foster City, CA), on ABI 310 and 377 automated sequencers. Alternatively, direct sequencing of PCR products was performed with the CEQ DTCS Quick start Kit (Beckman Coulter, Fullerton, CA), and extension products were electrophoresed on the CEQ 8000 automated sequencer.

Sequence Alignment

Homology assessment followed the rules of Borsch et al. (2003), further extended by Löhne and Borsch (2005). Microstructural changes are understood as mutational events that can include one to many nucleotides at once (Gu and Li 1995; Benson 1997; Kelchner 2000). The alignment method therefore aims to recognize sequence motifs resulting from such mutational events (e.g., simple repeats) rather than apply gap costs or calculating global or local nucleotide similarities. Hotspots (i.e., parts of a genomic region with a high number of length-mutational events that prevent unambiguous alignment at a given level of distance; Borsch et al. 2003) were excluded from analyses. Microsatellites consisting of single nucleotide stretches (A's or T's) were also considered to be hotspots because motif recognition within these length-variable stretches is not possible. It was recently shown that length mutations in microsatellites can include several repeat units at once (Tesfaye et al., forthcoming). For four different groups of species within *Nymphaea*, separate alignments had to be made of terminal parts of the *trnL* intron P8 stem loop because no homology could be assessed across the complete data set. These species groups were (i) all members of subgg. *Anecphya*, *Brachyceras*, and *Ondinea*; (ii) subg. *Hydrocallis*; (iii) subg. *Lotos* and *Nymphaea petersiana*; and (iv) subg. *Nymphaea*. Respective four sequence blocks are placed next to each other in character set P8.

Coding Microstructural Changes

Indels that result from microstructural changes were coded according to the widely accepted simple indel coding method (Simmons and Ochoterena 2000). Adjacent independent gaps (see Löhne and Borsch 2005) were kept separate because they result from independent rather than stepwise microstructural mutations. The recognition of independent microstructural mutations is important for correctly describing patterns of molecular evolution. In most cases, adjacent gaps result from simple sequence repeats that occur in close proximity but do not involve the same structural elements. As a consequence, there is no primary homology for these repeats. In complex situations (i.e., presence of overlapping indels; Borsch et al. 2003), simple indel coding has been suggested to result in a loss of information because of a high frequency of states coded as inapplicable in many taxa. To overcome this, Simmons and Ochoterena (2000) suggested calculating stepwise matrices that consider the stepwise mutational process and applying a parsimony principle to reconstruct the simplest sequence of mutational steps, a concept recently extended and incorporated into software by Müller (2005*b*). However, Simmons et al. (forthcoming) recently compared several indel-coding methods in a simulation study and found simple indel coding to perform rather well. In view of this and our aim to

apply the binary model in MrBayes to the indel character set as well, using simple indel coding here seemed warranted.

Tree Reconstruction

All characters, including indels, were given equal weight. Gaps were treated as missing data in the sequence matrices. Initial parsimony analyses were first executed with PAUP* (Swofford 2002), using heuristic searches with simple stepwise addition, tree-bisection-reconnection branch swapping, and multiple trees saved, but did not swap to completion after a reasonable time. Subsequently, the parsimony ratchet (Nixon 1999), as implemented in PRAP (Müller 2004), was used to find shortest trees for all 12 data sets (A1–C4). Settings were 200 ratchet iterations, weight 2, weighted = 25%, and 10 random addition cycles. Heuristic search parameters in the ratchet were simple stepwise addition (no random addition cycles), no multrees saved, maxtrees automatically increased by 100. To evaluate node support in parsimony trees, jackknifing was carried out with 37% character deletion and 10,000 replicates, using heuristic searches as for parsimony but multrees not in effect (saving only one tree). Calculations could be easily completed with PAUP* (Swofford 2002) on a Pentium PC in a few minutes for each data set. A high replicate number was chosen because this has been shown to be most influential in reaching small confidence intervals for jackknife percentages (Müller 2005*a*), which is of importance when comparing relative performance of data partitions.

Bayesian inference (BI) of data sets A1, A2, and A4 was conducted with MrBayes, version 3.1 (Ronquist and Huelsenbeck 2003). Modeltest 3.06 (Posada and Crandall 1998) was used to infer the optimal models to describe substitutional patterns in the *trnT-trnF* sequence data sets based on Akaike Information Criterion. The optimal models were as follows: GTR (Rodríguez et al. 1990) for the *trnT-trnL* spacer and the P8 stem loops of the *Brachyceras-Anecphya-Ondinea* clade and subg. *Hydrocallis*, the GTR + G model for the *trnL* intron and the *trnL-trnF* spacer, and the F81 model (Felsenstein 1981) for the P8 character set of subgg. *Lotos* and *Nymphaea*. For the indel character set, the binary (restriction site) model implemented in MrBayes 3.1 was applied. All analyses were performed for 1,000,000 generations, creating posterior probability distributions of trees with Metropolis-coupled Markov Chain Monte Carlo. Four independent runs with four chains each were carried out, with heating temperature 0.2, saving one tree every 100 generations. For data sets A1 and A2, posterior probabilities reached a stable value after 20,000 generations. Thus, the burn-in was set to 200, and a consensus was calculated from 9800 trees sampled after the burn-in in each chain (39,200 trees total). For data set A4, probabilities converged to a stable value after 15,000 generations, and 9850 trees (39,400 trees total) were sampled for calculating the consensus after setting the burn-in to 150.

Analysis of Molecular Evolution

Sequence variability, GC content, and transition : transversion ratios were calculated with SeqState (Müller 2005*c*). The probability for microstructural changes is difficult to assess, and the simple combination of matrices describing substitutions and indels with equal weight may not reflect true circumstances of mutational dynamics. This study therefore aims to empirically investigate the evolution of microstructural changes. Using PAUP* (Swofford 2002), we determined ancestral states of all indel characters for one of the eight shortest trees found in the ratchet analysis of data set A4, and state changes for indels variable within Nymphaeales are shown on the tree. Visual examination of the alignment during indel coding allowed for discrimination between simple and overlapping indels and to record motifs and positions of assumed microstructural changes in the two spacers and the intron. Using ancestral states and data on indel motifs and size (app. A) permitted the assessment of patterns of indel homoplasy and its correlation with particular kinds and positions of microstructural changes.

Results

Variability of the trnT-trnF *Region*

The overall matrix, excluding hotspots and *trnL* exons (A4), is 2073 characters (*trnT-trnL* spacer = 1–759, *trnL* intron = 760–1339, *trnL-trnF* spacer = 1340–2073). In the overall alignment of 2547 positions (*trnL* 5′exon = 939–973, *trnL* 3′exon = 1744–1793), hotspot H1 (pos. 217–393) comprises long insertions in Austrobaileyales, which cannot be aligned among *Austrobaileya*, *Illicium*, and *Schisandra*. Hotspots H2–H5 and H7 are not present (annotation of hotspots following Borsch et al. 2003). Hotspot H6 (pos. 1321–1511) is the largest hotspot in Nymphaeales. It corresponds to the terminal part of the P8 stem loop and is composed of AT-rich parts, of particular length in *Nuphar*, *Barclaya*, *Victoria*, and *Nymphaea* subgg. *Hydrocallis* and *Lotos*, which cannot be aligned with each other. As a consequence, four individual files (in the following, called P8) were created that comprise sequences of either *Nymphaea* subg. *Nymphaea*, subg. *Hydrocallis*, subg. *Lotos*, or subg. *Brachyceras* + *Anecphya* + *Ondinea* (fig. 2). Surprisingly, almost no intraspecific variation is found in this AT-rich satellite-like sequence within any of the plants sampled. Hotspot H8 (pos. 1744–1793) contains a microsatellite in Nymphaeales with presumed inversions of stretches of A's and T's.

Sequence statistics were calculated for data sets A4 (all taxa, all characters) and C4 (only Nymphaeales, all characters), distinguishing the two spacers, the group I intron, and P8 (table 1). The amount of variable and potentially parsimony-informative characters is highest in the *trnT-trnL* spacer, followed by the *trnL-trnF* spacer, and is lowest in the *trnL* intron. The *trnT-trnL* spacer exhibits the highest amount of length variability in data set C due to large insertions and deletions in *Amborella* and Austrobaileyales (accordingly lowest length variability in data set A). Sequences of the *trnL* intron are the most length variable in Nymphaeales, caused by the P8 stem loop. Both spacers and the intron have similar GC contents (34%–37%), whereas the CG content of P8 is considerably lower (4%).

The spacers show a much higher number of indels (87 and 90 for *trnT-trnL* and *trnL-trnF*, respectively) than the *trnL* intron (58). A list of indels, including their size and motifs, is provided in appendix A and the resulting binary matrix in appendix B. A high proportion of indels appears as unique to *Amborella*, the *Austrobaileya-Illicium-Schisandra* lineage, or the Nymphaeales clade. Only 48% of the indels of the overall matrix (A4) are variable within Nymphaeales.

Fig. 1 Floral diversity in the different subgenera of *Nymphaea*. *A*, Subg. *Brachyceras*, *Nymphaea micrantha* (M. Koehnen s.n.); *B*, subg. *Anecphya*, *N. gigantea* (T. Borsch 3836); *C*, subg. *Lotos*, *N. lotus* var. *thermalis* (T. Borsch 3832); *D*, subg. *Hydrocallis*, *N. rudgeana* (M. Koehnen s.n.); *E*, subg. *Nymphaea*, *N. mexicana* (T. Borsch & B. Summers 3226); *F*, subg. *Nymphaea*, *N. tetragona* (T. Borsch 3155).

Parsimony Analyses

Heuristic searches on the combined data set (A4) saving multiple trees did not swap to completion. Parsimony ratchet searches yielded eight shortest trees of 1022 steps after a few minutes. Searches on data sets A2 and A1 yielded trees with fewer steps due to the smaller size of their matrices (table 2). Because effects of different outgroups were minimal, only the trees found when analyzing the full taxon set (searches A1–4, including *Amborella* and Austrobaileyales) are illustrated here (fig. 3). This also presents the statistical results for the monophyly of Nymphaeaceae, Cabombaceae, and Nymphaeales.

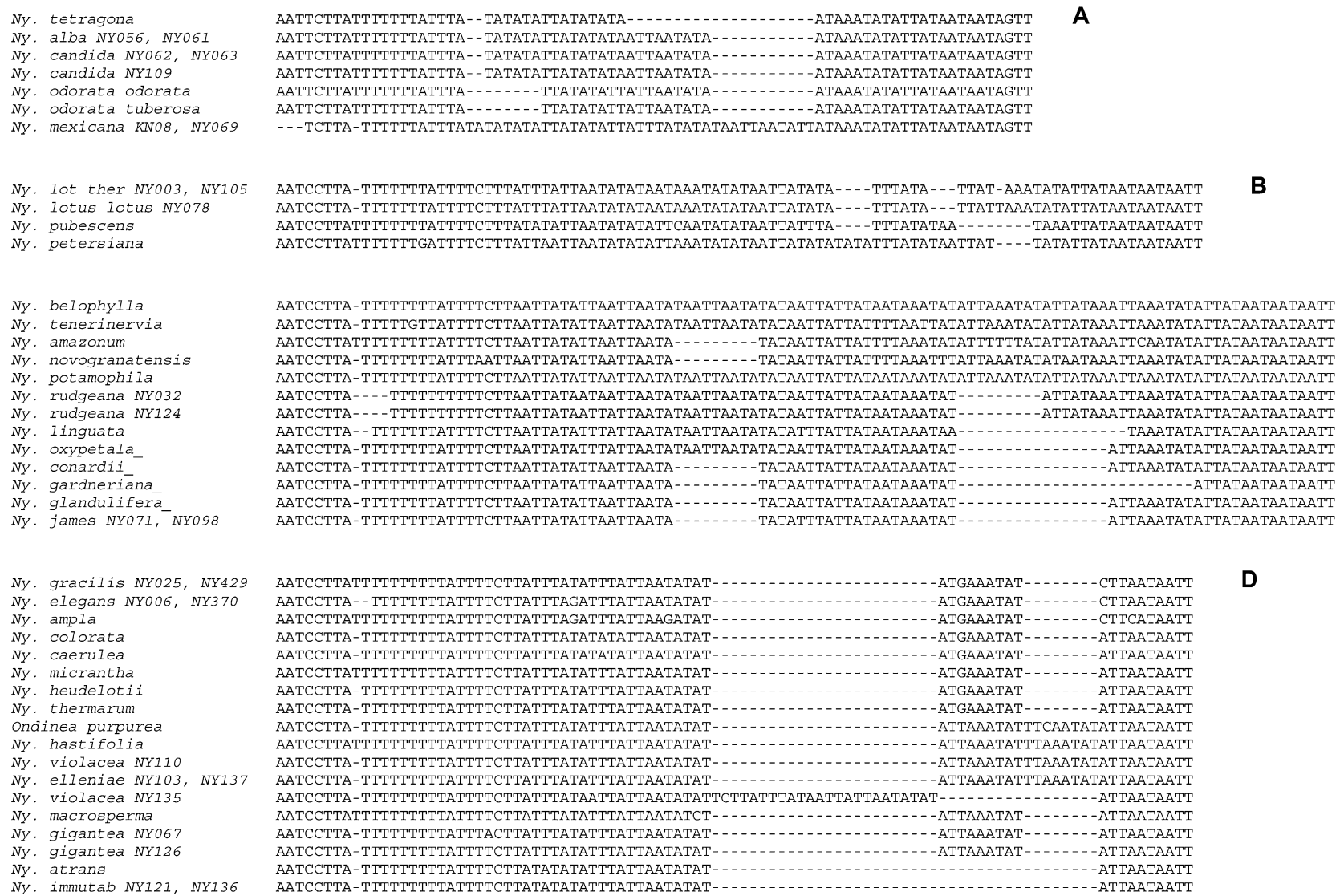


Fig. 2 Sequence alignment of the P8 loop of the *trnL* intron from major lineages of the genus *Nymphaea*. *A*, Subg. *Nymphaea* clade; *B*, subg. *Lotos–Nymphaea petersiana* clade; *C*, subg. *Hydrocallis* clade; *D*, subgg. *Brachyceras-Anecphya-Ondinea* clade. Individuals from one taxon with identical sequences are summarized on a single line.

Results of searches A1, A2, and A4 were highly congruent, differing only by the degree of resolution that increased from A1 over A2 to A4, due to additional characters (indels and *trnL* P8 stem loop region). Nodes unresolved in A1 or A2 are indicated as "n.r." in the strict consensus tree illustrated in figure 3. A single node (*Nymphaea violacea* NY135 sister to the *Nymphaea hastifolia–Ondinea–N. elleniae* clade) was resolved additionally in analyses A1 and A2 but weakly supported (63% and 64% jackknife [JK]). Only one weakly supported node (uniting *N. amazonum*, *N. conardii*, and *N. gardneriana*; 63% JK) was resolved differently in A1.

Whereas the Nymphaeales clade was inferred with 100% JK in all searches, the Nymphaeaceae gained only medium support (fig. 3; tables 2, 3). Maximum support was found for a core Nymphaeaceae clade consisting of *Euryale*, *Victoria*, *Nymphaea*, and *Ondinea* (100% JK; node 2 in fig. 3) and for *Barclaya* as its sister group (node 1 in fig. 3). In contrast, there is only weak indication for the monophyly of the genus *Nymphaea* with respect to the *Euryale-Victoria* clade as its sister (node 3 in fig. 3; tables 2, 3). The genus *Ondinea*, however, is shown with high confidence as being nested terminally within subg. *Anecphya* of *Nymphaea*. Species of subgg. *Brachyceras*, *Anecphya*, and *Ondinea* further share the same basic structure of the AT-rich satellite region in P8 of the *trnL* intron (fig. 2). A temperate subg. *Nymphaea* clade is inferred as sister to all remaining species of *Nymphaea* (including *Ondinea*), for which confidence is distinctly increased by indels. In addition to the subg. *Nymphaea* clade, there are two other major lineages, one consisting of subgg. *Hydrocallis* and *Lotos*, plus *N. petersiana* of subg. *Brachyceras*, which will be referred to as HL clade, and the other of subgg. *Brachyceras*, *Anecphya*, and the genus *Ondinea* (BAO clade). *Nymphaea petersiana* is only weakly supported as sister to subg. *Lotos*, but there is good evidence (82%–89% JK, 1.00 posterior probability [PP]) for its position within the HL clade. Resolution within the monophyletic subg. *Hydrocallis* is low and is caused only by indels and the P8 satellite region. However, each of the species in *Hydrocallis* (12 out of 14 were sampled) can be unambiguously recognized by its *trnT-trnF* sequence (fig. 2). There are three lineages in subg. *Brachyceras* (excepting *N. petersiana*), two of which contain species from the Old World tropics and a third that comprises only New World species (*N. elegans*, *N. gracilis*, *N. ampla*). The three lineages appear in a polytomy with *Anecphya-Ondinea*. Chloroplast data suggest several clades of species within subg. *Anecphya* and a polyphyletic nature of *N. violacea*. Signal provided by P8 alone for the four different lineages is illustrated in the form of unrooted networks (fig. 4).

Microstructural Changes

The occurrence of microstructural mutations in the *trnT-trnF* region during the diversification of Nymphaeales is illustrated in figure 5 using one of the shortest trees inferred with parsimony from data set A4. In figure 5, a distinction is made between entire indels and indels that are part of complex situations in the alignment. Inferred differences of ancestral states unravel mutational events. In conjunction with respective sequence motifs, frequencies and size distributions (i.e., the number of nucleotides involved in a respective mutational event) for four different kinds of microstructural mutations were determined (fig. 6). About 20% of the indel characters are homoplastic. Simple sequence repeats (SSRs) are the most frequent mutations, accounting for almost 90% of the reconstructed insertions. There is a heavy bias toward the maintenance of duplicated sequence motifs, once acquired during the evolution of the *trnT-trnF* region. The loss of SSR elements is inferred in only three cases, as compared with 68 gains. A smaller proportion of the inferred insertions do not show any recognizable motif and are thus considered to be of unknown origin (fig. 6).

Table 1

Sequence Statistics for the Two Spacers and the Intron Based on Data Sets A and C

	trnT-trnL spacer	*trnL* intron	P8 partition	*trnL-trnF* spacer	*trnT-trnF*
Data set A:					
Sequence length range, including hotspots (bp)	460–684	474–606	244–412	...	1323–1549
Average length (SD)	476 (33.7)	541 (24.6)	...	384 (21.1)	1402 (30.2)
No. characters[a]	761	579	...	737	2077
Variable characters (%)[a]	30.5	21.8	...	24.0	25.8
Informative characters (%)[a]	16.6	10.9	...	12.6	13.6
GC content (%)	37.01	34.59	...	34.01	35.22
ti : tv	1.97	.48	...	.62	.75
Data set C:					
Sequence length range, including hotspots (bp)	460–484	507–606	55–111	360–412	1360–1452
Average length (SD)	469 (5.6)	545 (20.4)	76 (15.7)	387 (11.9)	1401 (22.1)
No. characters[a]	761	579	[387]	737	2077
Variable characters (%)[a]	16.0	10.9	9.5	11.5	13.0
Informative characters (%)[a]	8.3	5.9	3.6	6.4	6.9
GC content (%)	37.4	34.33	4.33	34.04	35.28
ti : tv	2.07	.38	<.01	.55	.69

Note. The P8 partition was included only for data set C because it is not present in Austrobaileyales and *Amborella*. The number of characters for the P8 partition reflects matrices of the four major lineages (see fig. 2) in sum. ti : tv = transition : transversion ratio.

[a] Data refer to the matrix used in tree inference and exclude mutational hotspots.

Table 2

Results of Parsimony Analyses Using Distant and Close Outgroups and Different Character Sets

	A1	A2	A4	B1	B2	B4	C1	C2	C4
Characters variable	535	751	805	400	550	604	270	361	415
Characters pars. inv.	282	367	390	151	197	220	144	188	211
No. steps	716	963	1022	488	656	715	320	425	484
No. shortest trees	30	6	8	29	46	90	28	49	112
CI	.863	.866	.869	.889	.890	.892	.894	.887	.890
RI	.916	.917	.916	.933	.934	.933	.956	.954	.952
RC	.790	.794	.796	.830	.831	.832	.854	.846	.848
HI	.137	.134	.131	.111	.110	.108	.106	.113	.110

Note. Taxon set A = *Amborella*, *Austrobaileya*, *Illicium*, *Schisandra*; taxon set B = *Amborella*; taxon set C = *Brasenia*, *Cabomba*, *Nuphar* spp. 1 = substitutions; 2 = substitutions + indels; 4 = substitutions + indels + P8. pars. inv. = parsimony informative. CI = Consistency Index. RI = Retention Index. RC = Rescaled Consistency Index. HI = Homoplasy Index.

From the total of microstructural mutations, insertions account for 62%, although an insertion bias is less prominent in the AT-rich satellite-like part of the *trnL* P8 stem loop.

Bayesian Analysis

Results of BI are largely congruent with the trees found with parsimony. Similar to parsimony analysis, there is almost no support for the monophyly of Nymphaeaceae, and data set A1 (tree not shown) even resulted in *Nuphar* as sister to all remaining Nymphaeales rather than Cabombaceae. In congruence with the maximum parsimony trees, the core Nymphaeaceae gain high support (fig. 7) but not so for *Nymphaea* (including *Ondinea*). The BAO and HL lineages were found to constitute a clade, substantiating subg. *Nymphaea* as sister to all remaining species of *Nymphaea*. The sister group relationship between *N. petersiana* and subg. *Lotos* is confirmed with 0.98 PP. Within *Hydrocallis*, several groups of species are found after the addition of the satellite-like region in P8 of the *trnL* intron (fig. 7). The BI of the overall data set weakly (0.65 PP) indicates that subg. *Brachyceras* (except *N. petersiana*) could also be paraphyletic to subg. *Anecphya-Ondinea*. Branches for the divergence of *Euryale-Victoria* and the three *Nymphaea* clades are extremely short (fig. 8). Branches leading to the crown groups of the temperate subg. *Nymphaea*, BAO, and HL clades are long, whereas branches within these three clades are rather short. In Cabombaceae, a threefold-longer branch is found leading to *Cabomba*, as compared with *Brasenia* (fig. 8).

Outgroup Effects on the Trees Resolved

Tree topologies using distant versus close outgroups were identical, whereas jackknife support for 24% of the nodes varied (summarized in table 3). Only nodes with low to medium support were affected by different outgroups. Support for the monophyly of *Nymphaea* (including *Ondinea*; node 3 in fig. 3) decreased when using *Brasenia*, *Cabomba*, and *Nuphar* as outgroup (table 3). In comparison with the complete taxon set (67%–69% JK; searches A1-A4), node 3 gained hardly any support (50%–53% JK) in searches C1-C4. To the contrary, Nymphaeaceae were resolved with distinctly higher confidence when sequences of Austrobaileyales were excluded (table 3).

Discussion

Phylogenetic Utility and Molecular Evolution of the Spacers and the Group I Intron

Relative to the analysis of *trnT-trnF* sequences across basal angiosperms (Borsch et al. 2003), similar hotspots were found in this study. Nevertheless, due to the much smaller genetic distances covered, as compared with basal angiosperms as a whole, hotspots are smaller in extent or are not even found here (hotspots 2–5 and 7, which occur in basal angiosperms; Borsch et al. 2003). In particular, in the *trnT-trnL* spacer, length variability is considerably lower when considering the Nymphaeales alone. In the basal angiosperm data set, up to an additional 950 nt were present in hotspot H1, due to insertions in monocots, magnoliids, and eudicots.

The *trnL* intron is the most conserved of the three parts of *trnT-trnF*, in terms of both substitutions and microstructural changes. The number of coded indels in *trnL* in Nymphaeales is still lower than in the spacers, despite the variable P8 stem loop and the longer average length of the *trnL* intron sequences (table 1). This may be explained by much stronger structural constraints in the *trnL* group I intron, where substantial length variation in angiosperms is found only in the P6 and P8 stem loops (Borsch et al. 2003; Quandt et al. 2004). In Nymphaeales, the *trnT-trnL* spacer is distinctly more variable than the *trnL-trnF* spacer, whereas their level of variability at higher distances across basal angiosperms was found to be in the same range (Borsch et al. 2003).

Molecular Evolution and Phylogenetic Signal from the P8 Stem Loop of the trnL *Intron*

The P8 partition distinguished here corresponds to the highly variable, terminal AT-rich part of the P8 stem loop, as described by Borsch et al. (2003, annotated by two arrows in their fig. 2). The AT-rich terminal parts of P8 in Nymphaeales show almost no intraspecific variability, even in individuals from geographically distant populations (e.g., *N. jamesoniana* from Florida and Ecuador or *N. candida* from Scandinavia and Siberia). The same applies to individuals of *Nuphar* from Virginia and southern Florida (fig. 2). The only characters exhibiting intraspecific variability are substitutions but not indels. An explanation may be that the AT-rich sequences found in Nymphaeales are stabilized through hairpin

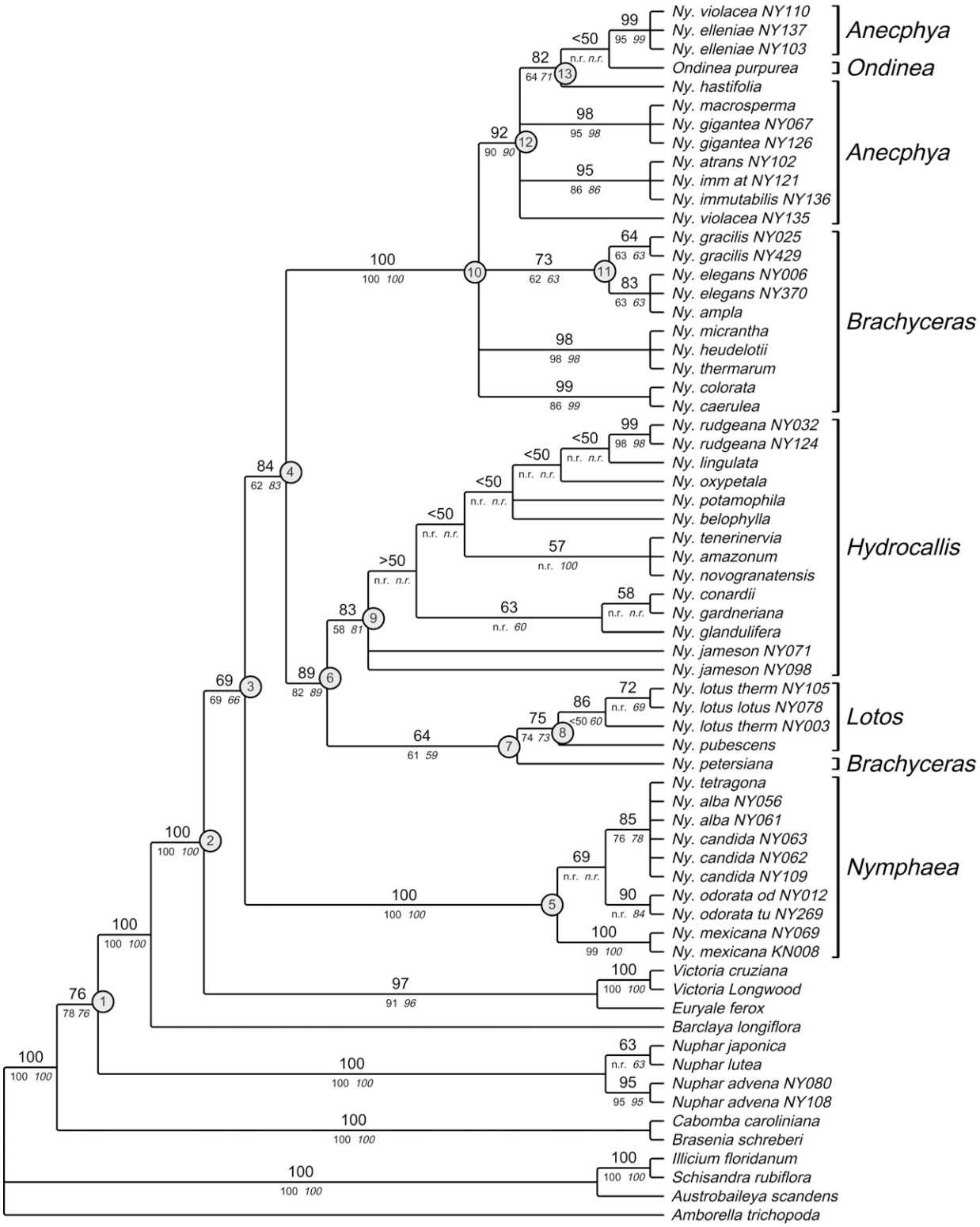


Fig. 3 Maximum parsimony tree (strict consensus of eight shortest trees found) of the combined analysis (data set A4) of *trnT-trnF* in *Nymphaea* and the Nymphaeales. Jackknife values are shown above branches. Support values for the same nodes found in parsimony analyses based on substitutions only (partition 1) and on substitutions + indels but excluding the satellite-like region in P8 (partition 2) are indicated below for comparison (left, roman, and right, italics, respectively). In case a node was not resolved in analyses A1 or A2, it is indicated by "n.r." Numbers refer to nodes discussed in table 3.

Table 3

Statistical Support of Major Nodes (10,000 jackknife replicates) with Respect to Different Outgroups and Partitions

Node	A1	A2	A4	B1	B2	B4	C1	C2	C4
1. Monophyly of Nymphaeaceae	78	76	76	93	90	91	...	...	...
2. Monophyly of core Nymphaeaceae	100	100	100	100	100	100	100	100	100
3. Monophyly of *Nymphaea* (including *Ondinea*)	69	66	69	68	57	60	51	50	52
4. Subg. *Nymphaea* sister to remaining *Nymphaea* (including *Ondinea*)	62	83	84	62	83	84	63	83	83
5. Monophyly of subg. *Nymphaea*	100	100	100	100	100	100	100	100	100
6. Subg. *Hydrocallis* sister to subg. *Lotos* (including *N. petersiana*)	82	89	89	82	89	90	85	89	89
7. *N. petersiana* sister to subg. *Lotos*	61	59	64	59	60	64	60	59	64
8. Monophyly of subg. *Lotos*	74	73	75	73	74	76	76	76	76
9. Monophyly of subg. *Hydrocallis*	58	81	83	58	82	83	55	80	80
10. Subgg. *Brachyceras-Anecphya-Ondinea* clade	100	100	100	100	100	100	100	100	100
11. New World clade within subg. *Brachyceras*	62	63	73	63	63	73	63	63	75
12. Monophyly of subg. *Anecphya* (including *Ondinea*)	90	90	92	90	90	92	89	89	90
13. *N. elleniae–N. hastifolia–N. violacea* p.p.–*Ondinea* clade	64	71	82	64	71	83	63	73	80

Note. The structure of this table follows that of table 2 (A1–C4).

formation in their secondary structures. The possibility of terminal P8 sequence elements to form hairpins was shown for *N. odorata* by Borsch et al. (2003) and for *Cabomba* and *Nuphar* by Quandt et al. (2004). Using *trnL* sequences from different tracheophyte, fern, and bryophyte lineages, Quandt et al. (2004) and Quandt and Stech (2005) hypothesized that the P8 stem loop has been independently prolongating in different land plant lineages. Extent and variability of AT-rich P8 elements as described for Nymphaeales in this study are highest for all angiosperms known so far. It seems that independent prolongation of P8 has even happened within *Nymphaea* (fig. 2), resulting in sequence elements diagnostic for four major clades.

Minisatellites are rare in the chloroplast genomes of flowering plants but have been encountered in orchids (Cafasso et al. 2001) and Rosaceae (King and Ferris 2002; Cozzolino et al. 2003) and also in the *trnK* group II intron of Aristolochiaceae (S. Wanke and D. Quandt, personal communication). In comparison with the AT-rich P8 sequences in Nymphaeales, the repetitive nature of these minisatellites is much more regular. As a consequence, slipped-strand mispairing (Levinson and Gutman 1987; Di Rienzo et al. 1994) can cause extreme length variability of such satellites, also within populations. The AT-rich sequence elements in the *trnL* intron appear to be much more stable, not being true minisatellites. Phylogeny inference based on the P8 character set (table 3) and indel evolution in P8 (fig. 5) show that homoplasy is higher when compared with the remaining parts of *trnT-trnF*. In *Nymphaea*, the AT-rich elements are of limited phylogenetic utility. Nevertheless, they provide a great amount of information for species identification (DNA bar coding).

History and Phylogenetic Utility of Microstructural Mutations

In order to efficiently apply information resulting from microstructural mutations in phylogeny reconstruction, three major issues need to be discussed. The first is how information from indels should best be coded. The second regards the mechanisms and probabilities for microstructural mutations to occur, and the third deals with possible weights of indel characters as compared with substitutions. The latter two issues are closely interrelated and of crucial importance for using likelihood approaches in tree inference. However, currently we are just beginning to understand how length-variable sequences evolve. Empirical studies on microstructural mutations in real data sets are one line of work in this field. We have reconstructed the evolutionary history of microstructural changes in the *trnT-trnF* region in Nymphaeales (figs. 5, 6) by using a parsimony approach.

The most frequently occurring mutations in this data set are SSRs, also called tandem repeats (Kelchner 2000). Compared with larger data sets of noncoding sequences analyzed so far, similar observations were made by Graham et al. (2000) for the slowly evolving chloroplast (cp) inverted repeat (IR) and by Löhne and Borsch (2005) for the rapidly evolving group II intron in chloroplast *petD*. Similar to this study (figs. 5, 6), Graham et al. (2000) encountered a deficit of 2–3-bp-long indels. Some universal patterns thus seem to be present in cp noncoding DNA. Based on *trnL-trnF* spacer data, van Ham et al. (1994) suggested that homoplasy decreased with increasing length of indels, but this is not the case either here or in other data sets (Müller and Borsch 2005*b*). Graham et al. (2000) found no difference in homoplasy among 162 indels of different size classes in the IR. For short inversions that are typically found in terminal loops of hairpin structures (Kelchner and Wendel 1996; Mes et al. 2000; Quandt et al. 2003), frequent changes between the two possible states have been demonstrated, even among closely related species. The same pattern occurs in a microsatellite in the *trnL-trnF* spacer in Nymphaeales that was excluded as hotspot 8 (details not shown, as this study does not focus on inversions). Moreover, Graham et al. (2000) described complex double inversions associated with inverted repeat sequences, which, in spite of their complexity, evolved in parallel in different angiosperm lineages. More recently, Provan et al. (2004) and Tesfaye et al. (forthcoming) provided evidence for cp microsatellites to represent a different class of length-variable DNA because rates of insertion and deletion at these loci are much higher as compared with those of the classes of microstructural mutations mentioned before. These examples show that rates and mechanisms of microstructural mutations obviously

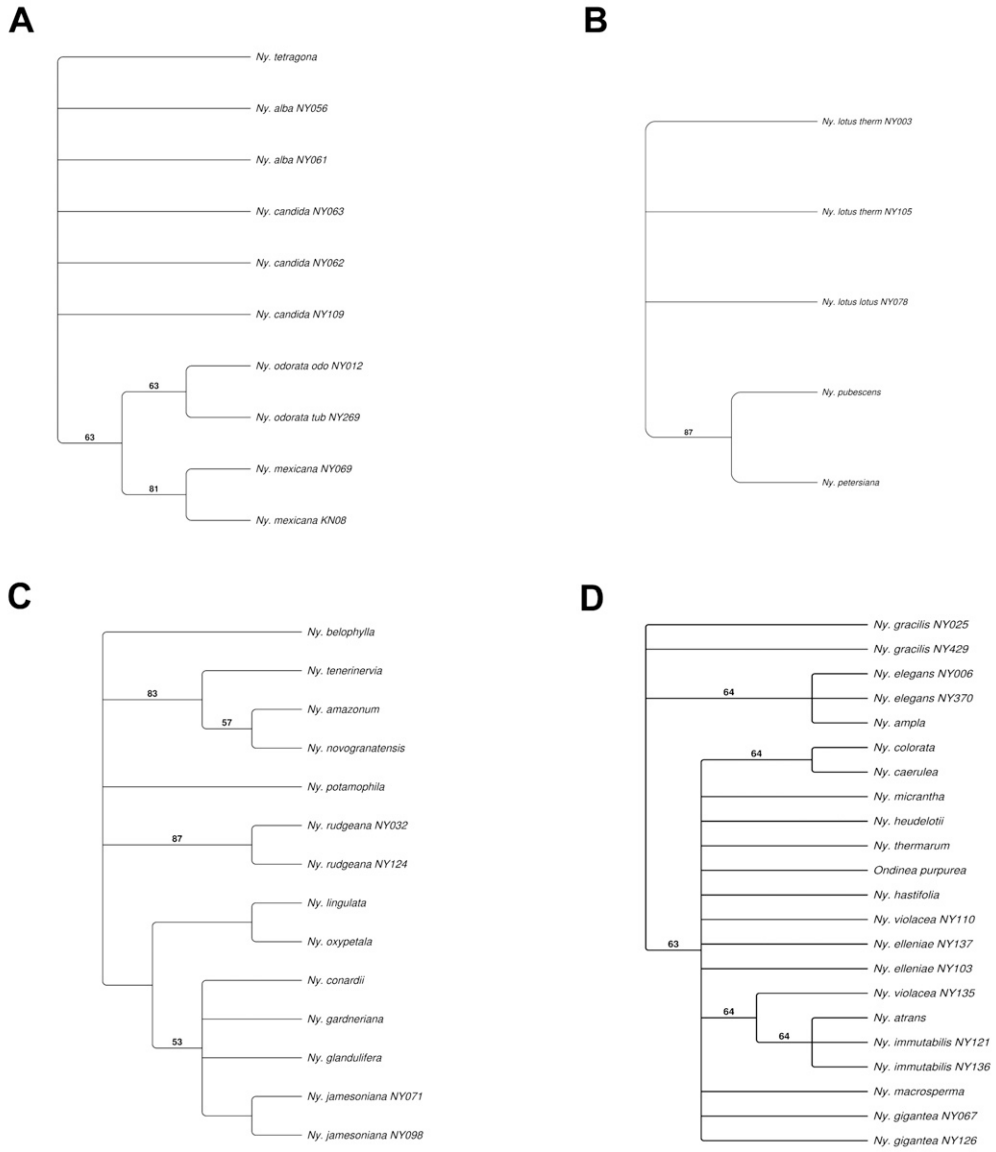
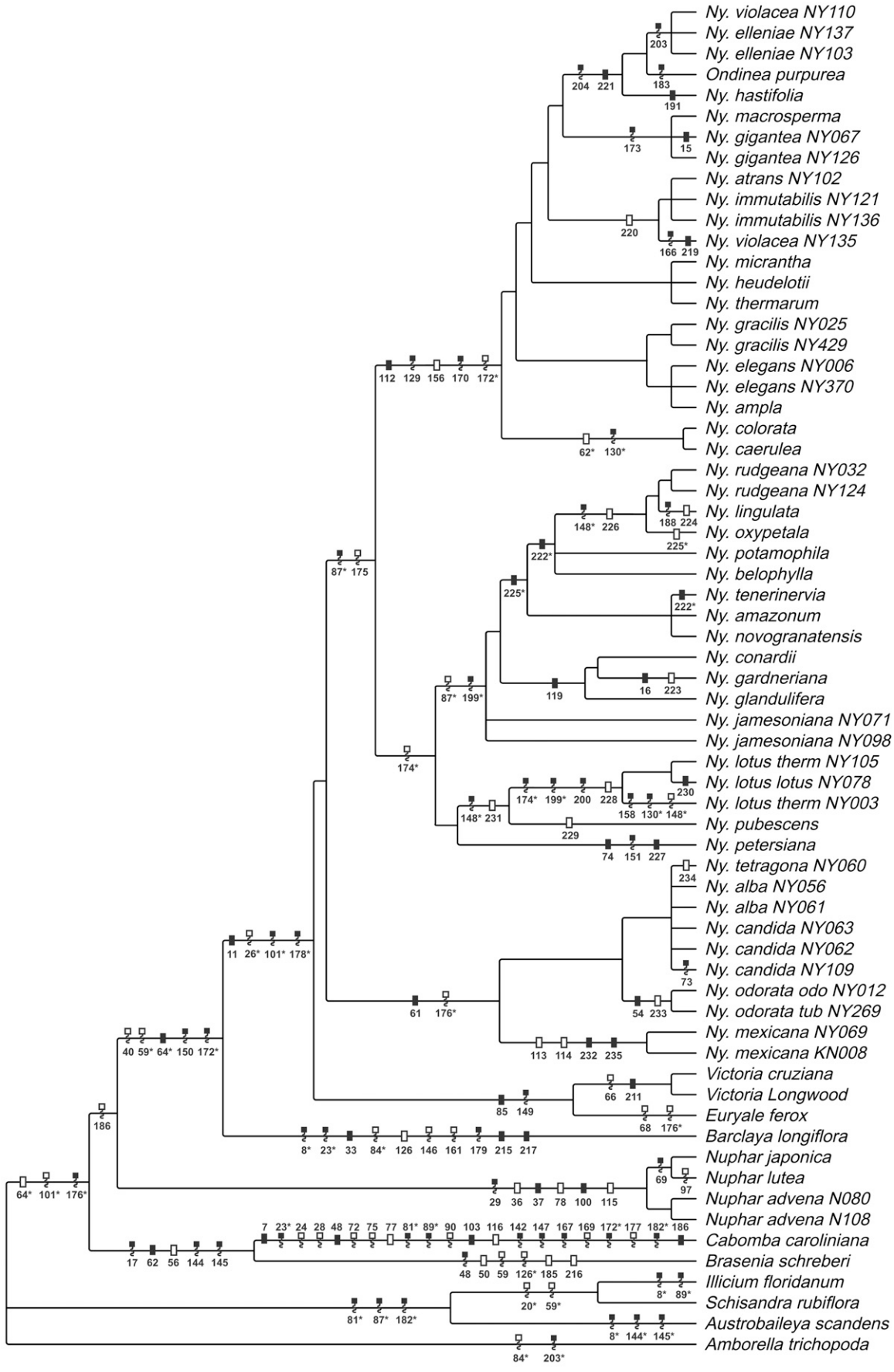


Fig. 4 Unrooted networks showing the signal of the P8 stem loop in the four clades (character set 3). In the data set of the subg. *Nymphaea* clade (*A*), one tree of six steps and CI = 1.0 was recovered; of the subg. *Lotos–Nymphaea petersiana* clade (*B*), two trees of 12 steps and CI = 1.0; of the subg. *Hydrocallis* clade (*C*), two trees of 25 steps and CI = 0.88; and of the *Brachyceras-Anecphya-Ondinea* clade (*D*), six trees of 14 steps and CI = 1.0.

depend strongly on specific characteristics of certain structural elements within a genomic region.

The reconstruction of state transformations in Nymphaeales (fig. 5) therefore distinguishes between indels that are part of complex situations (overlapping indels) and simple (=entire) indels. The percentage of indels in complex situations is highest in the *trnL-trnF* spacer (82%), medium in the *trnT-trnL* spacer (60%), and lowest in the *trnL* intron including P8 (45%). However, close examination of individual indels being part of a complex situation but coded as individual characters (according to simple indel coding) shows that these complex indels are not necessarily the result of a complex mutational process (i.e., two or more subsequent length-mutational events affecting the same site). We will explain this using the following two examples. In the case of the overlapping indels 146–151 in the *trnL-trnF* spacer (app. A; fig. 5), a 17-nt SSR (indel 148) was inserted independently in two different clades and later lost in one individual of *N. lotus* var. *thermalis* of one clade. Indel 149 is a 5-nt SSR that occurred completely independent in the common ancestor of *Euryale* and *Victoria* and also indel 147 as an autapomorphic SSR of *Cabomba* (6-nt SSR). Indel 150 is a 16-nt gain in core Nymphaeaceae and *Barclaya* of unknown origin. Indel 151 is a 4-nt SSR in *N. petersiana* reconstructed to have evolved in parallel to the gain of indel 148 in subg. *Lotos*, and the respective sequence motifs do not indicate any stepwise process. The only microstructural change that was inferred to have occurred on a successive node is 146 (a big deletion in *Barclaya* that could involve indel


Ny. violacea NY110
Ny. elleniae NY137
Ny. elleniae NY103
Ondinea purpurea
Ny. hastifolia
Ny. macrosperma
Ny. gigantea NY067
Ny. gigantea NY126
Ny. atrans NY102
Ny. immutabilis NY121
Ny. immutabilis NY136
Ny. violacea NY135
Ny. micrantha
Ny. heudelotii
Ny. thermarum
Ny. gracilis NY025
Ny. gracilis NY429
Ny. elegans NY006
Ny. elegans NY370
Ny. ampla
Ny. colorata
Ny. caerulea
Ny. rudgeana NY032
Ny. rudgeana NY124
Ny. lingulata
Ny. oxypetala
Ny. potamophila
Ny. belophylla
Ny. tenerinervia
Ny. amazonum
Ny. novogranatensis
Ny. conardii
Ny. gardneriana
Ny. glandulifera
Ny. jamesoniana NY071
Ny. jamesoniana NY098
Ny. lotus therm NY105
Ny. lotus lotus NY078
Ny. lotus therm NY003
Ny. pubescens
Ny. petersiana
Ny. tetragona NY060
Ny. alba NY056
Ny. alba NY061
Ny. candida NY063
Ny. candida NY062
Ny. candida NY109
Ny. odorata odo NY012
Ny. odorata tub NY269
Ny. mexicana NY069
Ny. mexicana KN008
Victoria cruziana
Victoria Longwood
Euryale ferox
Barclaya longiflora
Nuphar japonica
Nuphar lutea
Nuphar advena N080
Nuphar advena N108
Cabomba caroliniana
Brasenia schreberi
Illicium floridanum
Schisandra rubiflora
Austrobaileya scandens
Amborella trichopoda

150), which is autapomorphic and not informative. Another easier example is a deletion in *Cabomba* (indel 72) that overlaps with a 4-nt gain in one individual of *N. candida* from Siberia. In the latter case, the actual deletion in *Cabomba* is only 5 nt, whereas the gap appears to be nine positions long. Both events involve similar sequence positions but are on remote parts of the tree.

Based on the distribution of state transformations (fig. 5), the vast majority of mutational events at overlapping positions in the alignment occurred independently in distant parts of the tree. Thus, overlapping indels mostly are not the result of subsequent mutational events; i.e., they are not truly overlapping. Simple indel coding can therefore be expected to perform well because the extensive assumption of missing data in this strategy (Simmons et al. 2001) largely leads to missing signals in parts of the tree where respective nodes are already supported by a majority of other characters. Simulation studies currently under way (Simmons et al., forthcoming) comparing different indel-coding strategies also arrive at the conclusion that the simple indel-coding method works quite well. On the other hand, our empirical data on the evolution of *trnT-trnF* sequences indicate that locally calculating step matrices to be applied in complex indel coding may be misleading. Further empirical analyses of other length-variable genomic regions and other clades are needed.

Second, our empirical data show (fig. 6) a striking insertion bias for SSRs in all three parts of the *trnT-trnF* region in Nymphaeales. Almost all (65 of 68; 96%) simple sequence repeats were reconstructed as insertions. Finding tandem motifs in sequence alignments is hard when these motifs are short, and, thus, deletions in repeat regions may have been overlooked. However, flanking regions of gaps were always carefully examined during alignment and indel coding to prevent this, and the extreme excess of gains over losses as found in this study is not likely to be the result of ambiguity. This means that once acquired, there is little probability that a simple sequence repeat is lost again. The three exceptions (indels 172, 174, 228) are all located in parts of the *trnL-trnF* spacer where substitutions in the evolutionary history of Nymphaeales have led to several adjacent mono- or dinucleotide repeat motifs. These situations have been characterized as emerging satellites by Levinson and Gutman (1987). Like in more expanded satellite DNA, the mutational dynamics appears to be much more flexible in these emerging satellites (i.e., higher rates of microstructural mutations and high levels of homoplasy). Therefore, these few cases of lost SSRs might not be readily comparable. These empirical data on *trnT-trnF* molecular evolution further suggest that there might not be a single universal mechanism to explain length mutations, such as slipped-strand mispairing (Levinson and Gutman 1987). It appears that there are different mechanisms connected to different classes of microstructural mutations. For noncoding sequences of the slowly evolving chloroplast inverted repeat, Graham et al. (2000) also found that more tandem repeats were associated with insertions than with deletions but not as extreme as in this study. Graham et al. (2000) suggested that an insertion bias exists because the repair of mismatches after strand mispairing (Levinson and Gutmann 1987) largely involves adding additional nucleotides. So far, there are few studies reconstructing historical pathways of microstructural mutations in rapidly evolving chloroplast DNA. Van Ham et al. (1994) and Mes and 't Hart (1994) investigated the *trnL-trnF* spacer in Crassulaceae and in *Sedum*. A large matrix was analyzed by Löhne and Borsch (2005) for the *petD* intron. Neither of these studies focused on the specific evolutionary history of distinct classes of microstructural mutations.

It is obvious from this *trnT-trnF* data set that microstructural mutations (outside satellite DNA) are less frequent than substitutions (table 1), as has also been found in many other studies. This frequency bias has motivated suggestions to give indels a higher weight than substitutions. However, this is still under dispute. Gu and Li (1995) assumed a logarithmic distribution of indel size classes in human and rodent pseudogenes and suggested a logarithmic gap penalty for alignment algorithms. Other workers (e.g., Vogt 2002) proposed a priori weighing of indel characters, giving increased weight to longer indels. The idea is that rare microstructural mutation events (i.e., occurring with low probability) have higher phylogenetic information content than do frequent mutations (i.e., occurring with high probability). The distribution of reconstructed microstructural mutations (fig. 6) in *trnT-trnF* of Nymphaeales shows that there is a peak of 4–6-nt-long SSRs. Because there are many indels unique to *Amborella* and the Austrobaileyales, we have analyzed only those microstructural mutations that occurred within the radiation of the Nymphaeales crown group. The question of which microstructural changes occurred along the stem of Nymphaeales before its radiation into the extant genera will be dealt with elsewhere. Similar to the chloroplast genome inverted repeat (Graham et al. 2000), SSRs of 2 and 3 nt are particularly rare. Moreover, different classes of microstructural mutations show different size distributions, hindering any a priori weight assignment relative to length. High probability for particular microstructural mutations can also result in their multiple occurrences on different branches of the tree. Whereas most indels in this data set have a Consistency Index [CI] = 1 (details not shown), there are also homoplasious indels (fig. 5, annotated with asterisk). Homoplasious indels are often long, e.g., the 17-nt SSR (indel 148) that occurs in only one individual of *N. lotus*. Thus, based on empirical evidence, the probability for microstructural mutations seems to be determined by site-specific structural constraints and strongly differs among different classes. The above-mentioned high frequency of short inversions associated with hairpins is a good example. Further

Fig. 5 History of microstructural mutations in the *trnT-trnF* region of Nymphaeales (for one of the eight shortest trees of the overall maximum parsimony analysis). Because of the high number of indels that are variable only at the level of the three major lineages (*Amborella*, Austrobaileyales, Nymphaeales), only those microstructural characters with state changes within the Nymphaeales are displayed. Indel numbers following appendix A are shown above symbols, with 7–87 corresponding to the *trnT-trnL* spacer, 89–126 to the *trnL* intron, 129–218 to the *trnL-trnF* spacer, and 219–235 to the satellite region in P8 of the *trnL* intron. A wavy line with a square illustrates microstructural changes that are part of complex situations, and a rectangle indicates simple changes. Filled symbols are insertions; open symbols are deletions.

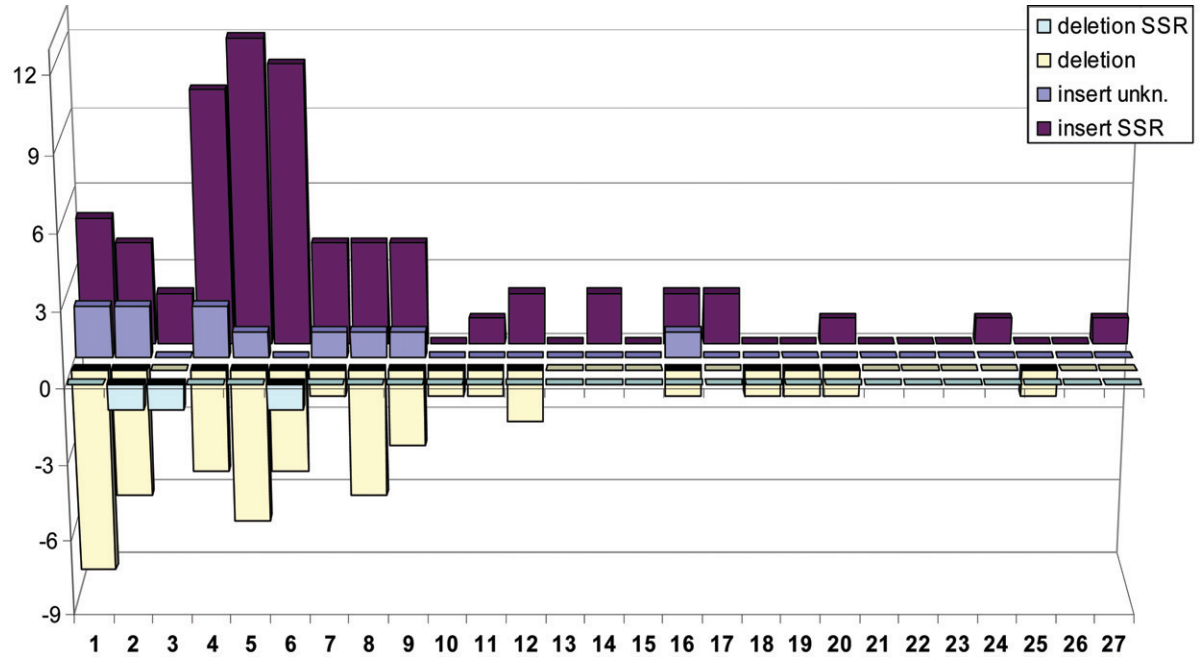


Fig. 6 Frequency and size distribution of microstructural changes within Nymphaeales. The number of nucleotides involved in a mutational event is shown (*X*-axis) relative to the absolute number of mutational events. Blue and violet bars refer to insertions (gain of sequence), whereas green and yellow bars refer to deletions (loss of sequence). The left bar illustrates simple sequence repeats (SSRs) and the right bar mutational events with unclear motifs, including deletions.

work needs to improve the understanding of molecular evolutionary patterns within specific genomic regions. In this study, we are reluctant to apply any differential weighting scheme for indels. However, it may be noted that there are globally synapomorphic indels (occurring only once), in contrast to homoplastic indels (fig. 5). Globally synapomorphic indels may deserve higher weight in phylogeny inference because of a lower probability to occur.

Signal from Indels and Substitutions in Nymphaeales

There is a large number of nodes for which support increased significantly when indels were added to the substitution-based matrix, such as the sister group relationship of the BAO and HL clades (62% to >83% JK; node 4 in fig. 3), making the temperate subg. *Nymphaea* sister to the remainder of *Nymphaea* species or the monophyly of subg. *Anecphya* (including *Ondinea*), where JK support increased from 51% to >90% (node 12 in fig. 3; table 3). The benefit of including indel characters (i.e., information derived from microstructural changes) is therefore obvious for the phylogeny of *Nymphaea* and the Nymphaeales. High phylogenetic information content of indel characters has been emphasized in an increasing number of studies on plants (Graham et al. 2000; Simmons et al. 2001; Geiger 2002; Hamilton et al. 2003; Leebens-Mack et al. 2005; Löhne and Borsch 2005; Müller and Borsch 2005*b*) and other organisms such as insects (Kawakita et al. 2003), bacteria (Griffiths et al. 2005), and hominoids (Lloyd and Calder 1991). Microstructural mutations have been shown to be less homoplastic than substitutions in an early study of *trnL-trnF* spacer evolution in Crassulaceae (van Ham et al. 1994) and for several data sets of group II introns in *trnK* and *petD*, such as for Lentibulariaceae (Müller and Borsch 2005*b*), Amaranthaceae (Müller and Borsch 2005*a*), and basal angiosperms (Löhne and Borsch 2005). Evidence provided here for the two spacers and the group I intron in *trnL* are along the same line and suggest generally lower homoplasy levels for indel as compared with substitution characters in data sets of the rapidly evolving spacers and introns of the cp genome LSU and SSU. Graham et al. (2000) show a similar pattern for the slowly evolving plastome inverted repeat regions.

Outgroup Effects on Nymphaea *Phylogeny Inference*

Although Austrobaileyales are not the sister group of Nymphaeales but the next higher lineage of the basal angiosperm grade, it was considered representative of the remaining angiosperms sister to Nymphaeales. This approach was deemed valid because tree reconstruction methods used here calculate relationships globally and root a posteriori. Choosing derived groups as outgroups would be dangerous if a priori polarization of character states is used that assumes the plesiomorphic state in the outgroup. If the ingroup is monophyletic, effects of distant outgroups are thus mostly effects of long-branch attraction. Graham et al. (2002) showed in simulation experiments that distant outgroups preferentially attracted long internal branches. For the case of Nymphaeales, all possible outgroups (*Amborella*, Austrobaileyales) exhibit distant sequences (fig. 8), and the same applies to possible outgroups for *Nymphaea* (*Brasenia*, *Cabomba*, *Nuphar*). Because an outgroup has to be chosen subjectively among the extant taxa, our rooting experiments are a means to evaluate the resolved topology. Table 3 shows that most nodes are not influenced by different outgroups (thus receiving confidence), whereas nodes 1 (monophyly of

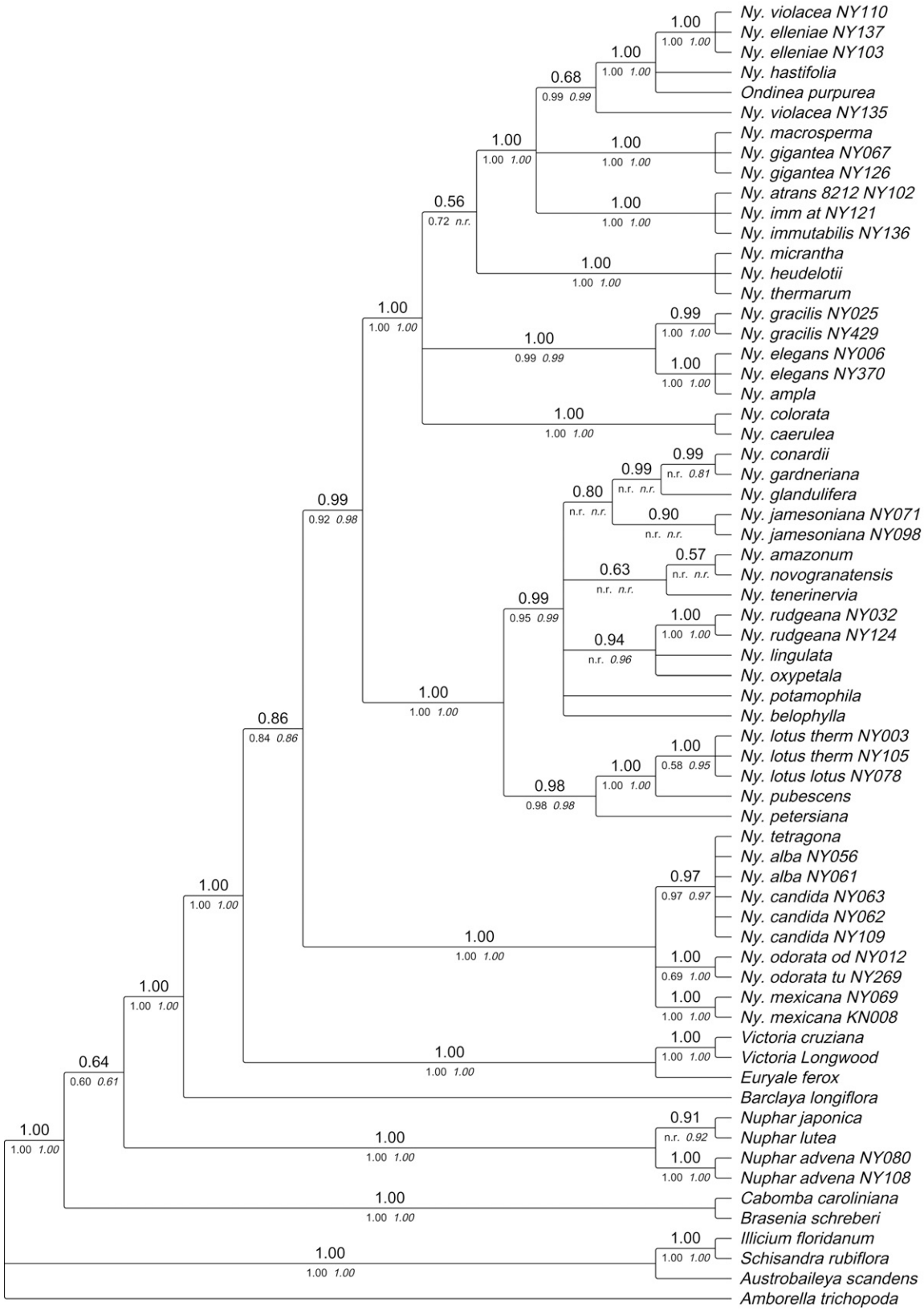


Fig. 7 Bayesian tree (strict consensus of all trees found with four independent runs) of the combined analysis (data set A4) of *trnT-trnF* in *Nymphaea* and the Nymphaeales. The corresponding confidence values shown above branches. Posterior probabilities of analyses based on the two data sets A1 and A2 are indicated below (left, roman, and right, italics, respectively).

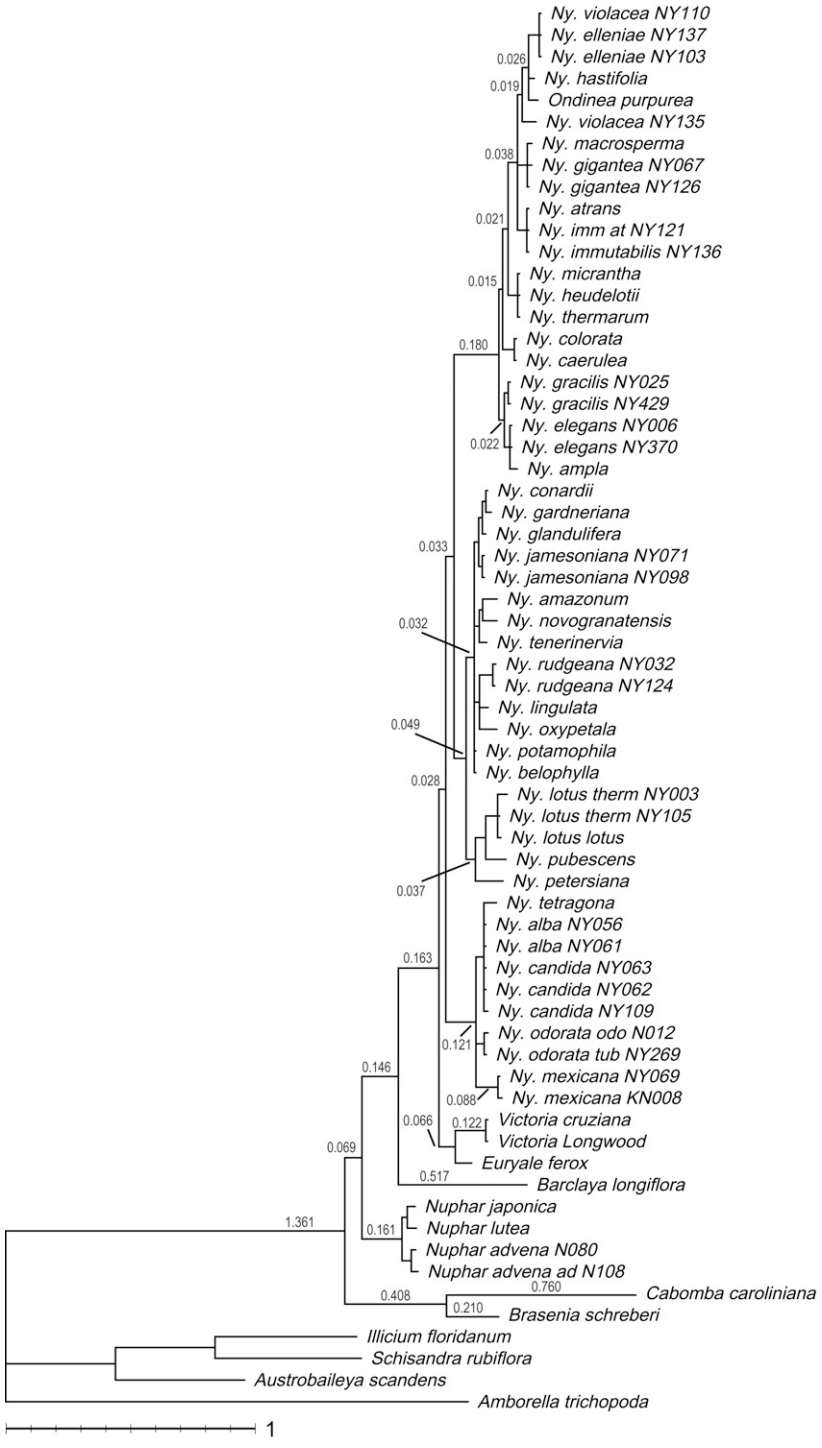


Fig. 8 Bayesian phylogram of the combined analysis (substitutions + indels + P8).

Nymphaeaceae) and 3 (monophyly of *Nymphaea*) are affected. The latter two nodes should be discussed with care.

Nymphaea *Monophyly and Position within Nymphaeales*

The clade consisting of *Nymphaea*, *Ondinea*, *Victoria*, and *Euryale* is one of the best-supported clades in both the parsimony and Bayesian analyses. This clade was also found in all data partitions by Les et al. (1999). A number of morphological characters are synapomorphic for *Nymphaea*, *Ondinea*, *Victoria*, and *Euryale*, such as the more or less strongly protruding floral axis (Moseley 1961), tetramerous outer perianth, and order of initiation of sepals and petals (Schneider

et al. 2003). In this study, we call these four closely related genera the core Nymphaeaceae. No clear phenotypic characters are known to date that would support the monophyly of the *Nymphaea* clade (node 3 in fig. 3). Considering *trnT-trnF* sequence data, it is one of the nodes with very low jackknife values (table 3) and complete lack of any supporting indel characters. Whereas the inclusion of *Ondinea* is highly substantiated, additional sequence data are needed to test relationships of *Victoria* and *Euryale* with members of the genus *Nymphaea* as currently circumscribed.

Relationships among Major Lineages of Nymphaea

The *Brachyceras-Anecphya-Ondinea* clade (BAO clade) gains maximum confidence in both parsimony and Bayesian analyses and, with five indels, also has high support from microstructural mutations (fig. 5). Four of the resulting indels are globally synapomorphic, and the internal branch leading to the *Brachyceras-Anecphya-Ondinea* crown group (fig. 8) is the second-longest branch in Nymphaeaceae, after the branch leading to *Barclaya*. There are several morphological character states shared among the species of subgg. *Brachyceras* and *Anecphya*, such as incomplete carpellary fusion (Caspary 1865, 1888; Conard 1905; Moseley 1961), small to absent capillary appendages, mostly violet flower colors, and slightly sculptured tectum of their pollen grains (Wiersema 1987; T. Borsch and M. Hesse, unpublished data). In this phylogenetic context, the apetalous condition in flowers of *Ondinea* (Den Hartog 1970; Williamson and Moseley 1989) appears to be derived as reductions in the number of floral organs. Kenneally and Schneider (1983) found still-petalous individuals of *Ondinea*. This further indicates that genetic changes from an *Anecphya*-like ancestor to *Ondinea* may, in fact, be small. A clarification of the origin of *Ondinea* within subg. *Anecphya* requires further sequence data from all genomes and further sampling of species within subg. *Anecphya*.

The *Hydrocallis-Lotus* clade (HL clade), as inferred from *trnT-trnF* data, also is well supported by several nonmolecular features. Its species have anthers embedded medially on the stamens (Wiersema 1987), conspicuous linear to clavate capillary appendages, nocturnal flowering (Prance and Arias 1975; Wiersema 1988; Hirthe and Porembski 2003), and a completely psilate tectum of their pollen grains (Wiersema 1987). Whether *N. petersiana* shares these character states or retains other plesiomorphic states in the genus requires further study. In contrast to close affinities between subgg. *Brachyceras* and *Anecphya* that were already suggested by Conard (1905), the exact affinities between the subgg. *Lotos*, *Hydrocallis*, and *Nymphaea* had remained obscure, depending on which morphological characters were emphasized (Wiersema 1987).

Currently, no phenotypic synapomorphies uniting the BAO and HL clades are known. Despite medium to low support values, both parsimony and Bayesian approaches converge on inferring their close relationship, rendering the subg. *Nymphaea* clade as sister to all remaining species of *Nymphaea*. There are a number of features restricted to the species of the temperate subg. *Nymphaea* clade, including distinctly verrucate pollen (T. Borsch and M. Hesse, unpublished data), seeds lacking hairlike protrusions (Weberbauer 1894; Wiersema 1987; although this may have evolved after the divergence of *N. mexicana*), and leaf morphology with exclusively entire margins. It may be noted that the branch leading to the subg. *Nymphaea* crown group is quite long (fig 8; coming third in Nymphaeaceae), and many morphological features may be derived in this temperate clade.

Radiations within Major Lineages of Nymphaea

The *trnT-trnF*-based trees resolve the monophyly of several lineages within *Nymphaea* with high confidence, whereas relationships within their crown groups (of the temperate subg. *Nymphaea* clade, *Hydrocallis* clade, *Lotus* clade, and *Brachyceras-Anecphya-Ondinea* clade) are much more difficult to resolve. In addition, evolutionary patterns may be more complex as a result of reticulate evolution. It has been well known for a long time that species of *Nymphaea* can hybridize within subgenera but not between them (Conard 1905; Wood 1959). However, Doran et al. (2004) and Les et al. (2004) recently confirmed the artificial generation of an intersubgeneric hybrid involving *N. colorata* and *N. gigantea*. The two parental species belong to subgg. *Brachyceras* and *Anecphya*, respectively, and this hybrid may be anticipated in light of the close relationship between the two subgenera.

Subg. *Nymphaea*. There are three lineages within temperate subg. *Nymphaea*, and each are well supported, but relationships among them are unclear. One comprises *N. mexicana* (sect. *Xanthantha*) and is depicted as sister to the remaining temperate species (figs. 3, 7). Given that pollen grains of all remaining subg. *Nymphaea* species have distinct, cylindrical protrusions on all surfaces of the ektexine, including the operculum (T. Borsch and M. Hesse, unpublished data), while *N. mexicana* has only small protrusions and a glabrous operculum (Wiersema 1987; Gabarayeva and El-Ghazaly 1997), and that all other species of *Nymphaea* have a psilate tectum, pollen morphology might provide support for the exclusion of *N. mexicana* from a core temperate clade. The adaptation to subtropical climates and long pedunculate flowers in *N. mexicana* might then be viewed as plesiomorphic character states. However, its large seeds (Wiersema 1987) and extensive development of long stolons seem to be synapomorphic in nature. The second lineage (figs. 3–5) comprises *N. odorata* s.l., a highly variable species occurring throughout North America to northern Central America (Wiersema 1996). Extensive sampling using chloroplast *trnL* intron and nuclear ITS sequences, as well as ISSR fingerprints (Woods et al. 2005*a*, 2005*b*) unraveled two chloroplast haplotypes of different geographical distribution, largely corresponding to subsp. *odorata* and subsp. *tuberosa* (Paine) Wiersema & Hellq. In addition, the results of Woods et al. (2005*a*, 2005*b*) indicate ancient and recent gene flow among subspecies and provide evidence for possible ancient introgression of an *N. mexicana*–type ancestral genome into subsp. *tuberosa*. Within the third lineage, there is no sequence divergence among European populations of *N. alba* and *N. candida* (sect. *Nymphaea*; figs. 4, 5, 8), although both taxa can be distinguished morphologically when comparing individuals from central and north-central (Scandinavia) Europe. The exact distributions of both taxa in Eastern Europe and temperate Asia are not clear, and further species have been accepted in recent floristic surveys such as *N. colchica* (Gagnidze 2005). The individual of *N. candida* from Siberia sampled in

this study shows an autapomorphic SSR (fig. 5) in the *trnT-trnL* spacer. Further studies need to determine whether different cp haplotypes exist among the Eurasian populations of *N. alba*, *N. candida*, and allies. The three dwarf species of sect. *Chamaenymphaea* (Wiersema 1996), which extend to subboreal and boreal regions of the Northern Hemisphere, were represented by only one individual of *N. tetragona* from Finland in this study. Sequences of *trnT-trnF* are different in this species (figs. 2, 5, 8), and it is likely that small erect rhizomes, flowers with few tepals (eight to 17), and filaments widest above the middle will be shown as synapomorphic for sect. *Chamaenymphaea* once other species are included.

Subg. *Hydrocallis*. Although every species of the subg. *Hydrocallis* clade possesses several autapomorphic substitutions and often also indels (figs. 3, 5, 7), relationships are hardly resolved with *trnT-trnF* sequence data. Internal branches in the *Hydrocallis* crown group are extremely short (fig. 8), indicating potential rapid radiation. Nevertheless, it is obvious that most characters variable within *Hydrocallis* are located in the terminal AT-rich part of the P8 stem loop of the *trnL* intron (figs. 2, 4). On the other hand, synapomorphic indels are rare in the AT-rich part of the P8 stem loop (fig. 5), in spite of considerable length variability, so that a lack of resolution could, in addition, be caused by high levels of homoplasy in P8. A clade of *N. conardii*, *N. gardneriana*, and *N. glandulifera* appears in the Bayesian and parsimony trees (figs. 3, 7), albeit with low support. In the Bayesian tree, *N. jamesoniana* is further resolved as sister to the three species. On the basis of floral biology, overall floral morphology, distinctive weblike leaf venation, similar chromosome number (2n=28), seed morphology, and flavonoid profile, a relationship among *N. conardii*, *N. gardneriana*, and *N. jamesoniana* would be expected (Wiersema 1987). The association of *N. glandulifera*, for which chromosome and flavonoid data are lacking, with these species is supported by its floral morphology and biology but not its leaf venation or seed morphology. For the clade of *N. rudgeana*, *N. lingulata*, and *N. oxypetala*, as evidenced with *trnT-trnF*, little support from morphology, flavonoid chemistry, or floral biology can be found. A presumed relationship among *N. oxypetala* and both *N. belophylla* and *N. potamophila* based on their sagittate leaf morphology is not supported. Nevertheless, this clade does include two polyploid species (*N. rudgeana* and *N. oxypetala*) with a chromosome number higher than 2n=28 (Wiersema 1987) and may represent a lineage of derived species. Wiersema (1987) hypothesized that *N. rudgeana*, because of its unusual chromosome number (2n=42) and similarities to subg. *Lotos* in leaf morphology and seed anatomy, might be the result of an ancient hybridization event between an ancestor of subg. *Lotos* and a former member of the *Hydrocallis* clade. Sequences of *trnT-trnF* reveal the *N. rudgeana* chloroplast genome as clearly nested within *Hydrocallis*. Further studies using nuclear genes will be needed to clarify whether *N. rudgeana* is of reticulate origin, with a paternal parent from another lineage. The grouping of *N. amazonum*, *N. novogranatensis*, and *N. tenerinervia* (figs. 3–5, 7) is well supported by morphology. These three species share a number of features, in their phytochemistry, seed morphology, and floral biology, that have been postulated as ancestral within the subgenus (Wiersema 1987). They also share two other presumably derived characters, a powdery stigma and a granulate seed surface topography.

Subg. *Lotos*. This subgenus constitutes the smallest of the five clades. *Nymphaea lotus*, represented here by one specimen from West Africa (var. *lotus*) and two specimens from the hot springs of Hungary (var. *thermalis*), shows the highest intraspecific variability in the genus, in terms of both substitutions and indels. The typical variety is distributed in Africa and Madagascar (Conard 1905), whereas var. *thermalis* was primarily separated for geographical reasons. However, there is no clear association between the European individuals in this study. *Nymphaea pubescens* is shown as sister to *N. lotus* and is well separated genetically, as indicated by indels (fig. 5) and distinct branches (fig. 8). According to Conard (1905), *N. pubescens* occurs from India to the Philippines, Java, and Australia and differs from *N. lotus* by ovate leaf blades (orbicular in *N. lotus*) that are much more densely pubescent beneath than in *N. lotus*. Sequence divergence between *N. lotus* and *N. pubescens* considerably exceeds the variation typically found within species, and because all individuals of *N. lotus* form a statistically supported clade, *trnT-trnF* data provide clear evidence for the distinctness of *N. pubescens*.

The emergence of *N. petersiana* as sister to the remainder of species in subg. *Lotos* is surprising but has also been confirmed with a number of other chloroplast (Löhne et al., unpublished data) and nuclear ITS sequences (Borsch 2000). Leaf morphology of *N. petersiana* sampled from Malawi is similar to subg. *Lotos*, with margins being toothed and veins distinctly raised from the blade beneath but without the pubescence characteristic of that subgenus. However, the floral morphology of *N. petersiana* strongly resembles subg. *Brachyceras* and contrasts with subg. *Lotos* in its staminal appendages, short-triangular carpellary appendages, and blue pigmentation (Mendonça 1960). This incongruous morphology was commented on nearly a century ago by Gilg (1908), who mentioned that Conard (1905), who had treated this species in synonymy under *N. capensis* Thunb. of subg. *Brachyceras*, had annotated a sheet of this plant as a mixture of these two groups. Verdcourt (1989) treated this taxon under *N. nouchali* of subg. *Brachyceras*, using its leaf characters to distinguish his var. *petersiana* (Klotzsch) Verdc. from the other varieties of that species. The Malawan plants have large round tubers that are used as food (Chawanje et al. 2001), but so far, important information on the floral biology, degree of syncarpy, and pollen morphology is lacking for *N. petersiana* that could shed further light on its subgeneric affinities.

Subg. *Brachyceras*. Within this subgenus, all New World species sampled are resolved in a clade (node 11 in fig. 3). *Nymphaea elegans* (from Florida and Texas) and *N. ampla* (from Veracruz, Mexico) appear closely related and distinct from the Mexican Plateau endemic *N. gracilis*. The latter species exhibits a distinct seed morphology and more emergent flowers with broader filaments that support this dichotomy. It would be interesting to evaluate further the relationships among these taxa with additional material of *N. elegans* from western Mexico and the widespread *N. ampla* and *N. pulchella* DC. from throughout their Neotropical ranges. Until recently (Bonilla-Barbosa 2001; Wiersema 2001, 2003), *N. pulchella* was generally subsumed under *N. ampla*. Molecular analysis of further samples of *N. ampla*/*N. pulchella* will be required to substantiate this classification. Species limits and nomenclature of Old World members of subg. *Brachyceras* are complex and have been studied only floristically. Of these studies, the treatment by Verdcourt (1989) for East

Africa is perhaps the broadest in its scope but is nonetheless regional in focus and based largely on herbarium study and thus may not completely reflect natural groups. It is interesting that the studied species from central and western Africa (*N. micrantha*, *N. heudelotii*, *N. thermarum*) are resolved in a different clade as compared with *N. caerulea*/*N. colorata*, of a largely eastern African lineage. Using additional sequence data from both chloroplast and nuclear genomes, it remains to be seen whether there is a single African radiation of *Brachyceras* or whether the two lineages described here have different origins. Two rare yellow-flowered African species of subg. *Brachyceras*, *N. stuhlmannii* (Engl.) Schweinf. & Gilg and *N. sulphurea* Gilg, unfortunately were not available for study.

Subg. *Anecphya*. Three clades are resolved by *trnT-trnF* data within *Nymphaea* subg. *Anecphya* (fig. 3). The two best-supported of these clades consist of *N. atrans* and *N. immutabilis* and *N. macrosperma* and *N. gigantea*, respectively. Both of these species pairs are characterized by large seeds, a distinctive gap between petals and stamens, and toothed leaf margins. Jacobs (1992) considered *N. immutabilis* and *N. atrans* as close relatives and reported frequent natural hybrids with intermediate character states and reduced fertility in areas where both species grow sympatrically. The third, more weakly supported, clade comprising *N. violacea*, *N. elleniae*, and *N. hastifolia* (and *Ondinea*), is characterized by relatively small seeds, petals grading into stamens, and entire-to-sinuate leaf margins (e.g., Jacobs and Porter, forthcoming). Remarkably, the two samples of *N. violacea* do not form a clade in the present analysis (figs. 3, 5). Ongoing studies on the subgenus *Anecphya* involving an extended taxon sampling and additional information from the nuclear ITS region (C. Löhne, T. Borsch, S. W. L. Jacobs, C. B. Hellquist, and J. H. Wiersema, unpublished data) confirm the polyphyletic nature of *N. violacea* and provide evidence for ancient and recent hybridization and introgression within subg. *Anecphya* and, especially, within the small-seeded group of species.

Classification

In his monograph of the water lilies, Conard (1905) established a classification system of five subgenera within *Nymphaea*. He used the name *Castalia* DC. (1821) for the temperate subgenus, which is to be called subg. *Nymphaea*, as it includes the type of the genus (*N. alba* L.). The name *Hydrocallis* was originally published by Planchon (1852) as a section, *Lotos* by De Candolle (1821) as a section, and *Anecphya* and *Brachyceras* by Caspary (1865) as a subsection. Conard maintained the circumscription of these five groups but leveled their classification at the same rank within *Nymphaea*. The results of this first molecular phylogenetic analysis of *Nymphaea* indicate that, with the possible exception of subg. *Brachyceras*, all subgenera are monophyletic and should be maintained.

To eliminate one element contributing to the paraphyly of subg. *Brachyceras*, we propose to shift *N. petersiana* to subg. *Lotos*. As already mentioned, although accepted by both Gilg (1908) and Mendonça (1960), *N. petersiana* was treated as a synonym of *N. capensis* by Conard (1905) and by Verdcourt (1989) as one of five African varieties under the Indian *N. nouchali* Burm. f. In addition to var. *petersiana*, he also recognized var. *ovalifolia* (Conard) Verdc., var. *caerulea* (Savigny) Verdc., var. *mutandaensis* Verdc., and var. *zanzibariensis* (Casp.) Verdc. Unlike for *N. petersiana*, our *trnT-trnF* data indicate that *N. colorata* (*sensu* Verdcourt = *N. nouchali* var. *zanzibariensis*) and *N. caerulea* Savigny are closely related, thus supporting Verdcourt's (1989) grouping of these two species. However, the status of his var. *ovalifolia* and var. *mutandaensis*, which were not sampled, remains unknown.

In subg. *Nymphaea*, both sect. *Chamaenymphaea* (dwarf allies of *N. tetragona*) and sect. *Xanthantha* (*N. mexicana*) seem to reflect natural groups, while the typical sect. *Nymphaea*, as currently circumscribed (*N. odorata*, *N. alba*, *N. candida*), might be paraphyletic to sect. *Chamaenymphaea*. The nontypical subgroups in subg. *Nymphaea* were first introduced by Planchon (1853; as *Nymphaea* sect. *Castalia* b. *Chamaenymphaea*) and Caspary (1888; as *Nymphaea* sect. *Symphytopleura* c. *Xanthantha*) and later formalized as sections by Wiersema (1996). To clarify their status, further molecular work is needed to increase resolution of the subg. *Nymphaea* clade. Our chloroplast data also provide evidence for new groupings in the other subgg. *Hydrocallis*, *Brachyceras*, and *Anecphya*. While these may eventually prove worthy of formal recognition, it would be premature to do so at this time. To fully unravel the extent of hybridization and introgression within subg. *Anecphya* (C. Löhne, T. Borsch, S. W. L. Jacobs, C. B. Hellquist, and J. H. Wiersema, unpublished data) and also within the subgg. *Brachyceras* and *Hydrocallis*, the additional study of nuclear markers and multiple individuals from different populations is required.

Intraspecific Variability and Species Identification with trnT-trnF *Sequences*

The actual information content of *trnT-trnF* sequences is much higher than is reflected in hypotheses on species relationships. Many species have a number of autapomorphic nucleotide substitutions and/or indels, which allow their unambiguous identification, even in cases where phylogenetic relationships are not resolved. Species identification with molecular markers functioning as bar codes is now being intensely discussed (e.g., Kress et al. 2005). However, the accuracy of such an approach and the kind of genomic region to be selected require further study. Bar coding in *Nymphaea* seems particularly interesting, as many species are grown as ornamentals and are difficult to identify in vegetative state. In addition, hybridization within subgenera is frequent, and many cultivars have been obtained through extensive hybrid breeding. For several of the older cultivars (e.g., Director G. T. Moore), the exact origin and possible parents are not definitely known, so separate molecular identification of chloroplast haplotypes and nuclear genotypes could provide an important perspective.

In *Nymphaea*, this study distinguishes 29 different *trnT-trnF* sequences (including those parts located in hotspots and excluded from tree inference). Given that 35 species are sampled, the resolving power of *trnT-trnF* sequences in terms of species identification is 83%. Intraspecific variability has been encountered in *N. mexicana* (one substitution in the *trnL-trnF* spacer; seven to eight A's in the microsatellite of H8), *N. odorata* (one substitution in the *trnL* intron), *N. lotus* (several substitutions and indels in all three partitions of *trnT-trnF*), *N. rudgeana* (one substitution in the satellite-like part of P8), and *N. gigantea* (one substitution and one 4-nt SSR in the *trnL-trnF* spacer). In all species

represented by several individuals of different geographical origin, such variable positions are outnumbered by substitution and indel characters synapomorphic for all individuals of a species. Intraspecific variability, therefore, is not a barrier to species identification with *trnT-trnF* sequences. In the case of *N. odorata*, one variable position in the *trnL* intron characterizes chloroplast haplotypes corresponding to morphologically differing subspecies (Woods et al. 2005*a*). Strikingly different *trnT-trnF* sequences in individuals currently identified as *N. violacea* based on morphology are the result of complex and not yet fully understood evolutionary patterns in subg. *Anecphya* (C. Löhne, T. Borsch, S. W. L. Jacobs, C. B. Hellquist, and J. H. Wiersema, unpublished data). The species names currently used thus reflect a situation that needs clarification through further taxonomic work. For those species that cannot be distinguished by their *trnT-trnF* sequences but that appear to be morphologically distinct, further sequence and perhaps also fingerprint data need to be generated. These are *N. belophylla*–*N. potamophila*, *N. heudelotii*–*N. micrantha*–*N. thermarum*, and *N. atrans*–*N. immutabilis*. Additional molecular work in conjunction with a careful analysis of phenotypic and autecological characters will be required in the case of *N. alba*–*N. candida* and *N. caerulea*–*N. colorata*, where the circumscription of biological entities and the application of names is uncertain.

Conclusions and Future Directions

This phylogenetic analysis of *Nymphaea* and the Nymphaeales based on *trnT-trnF* sequence data underscores the importance of a dense taxon sampling. This is exemplified by the surprising finding that *Ondinea* is nested within subg. *Anecphya*. Further research is needed to clarify the nearest relatives of *Ondinea* within subg. *Anecphya*. In particular, the monophyly of *Nymphaea* with respect to the *Euryale-Victoria* clade also needs to be further tested by additional sequence characters, particularly from different genomic regions. This *trnT-trnF* study will provide the basis to select a range of appropriate species to be analyzed with a high number of characters that represent all major lineages of water lilies.

In agreement with molecular clock dating (Yoo et al. 2005), it can be accepted that the extant genera of the Nymphaeales represent a comparatively recent radiation, beginning in the Tertiary. However, the high support of core Nymphaeaceae from molecular and morphological characters and the short branches above the core Nymphaeaceae node indicate a second phase of rapid radiation in Nymphaeales, after initial radiation of the Nymphaeales crown group. This second phase involves the three major lineages now clearly recognized in *Nymphaea* and also the *Euryale-Victoria* clade. Recognizing *Nymphaea* as a morphologically diverse and biogeographically complex genus changes the picture of major clades within Nymphaeales. Because a considerable time span exists between the divergence of the Nymphaeales stem from the angiosperm backbone and its radiation into its extant diversity, the question of what the innovations of the Nymphaeales stem are, as compared with synapomorphies that evolved later within the crown group, becomes relevant. Reconstructing phenotypic character evolution will therefore have to be based on all major lineages that can now be recognized within Nymphaeales.

Acknowledgments

This article was originally prepared to fulfill part of the requirements of a PhD dissertation of T. Borsch but has been considerably extended by additional plant material that became available over the years. Financial support through a scholarship from the Studienstiftung des deutschen Volkes and research grants from the International Water Lily and Water Gardening Society to T. Borsch are greatly acknowledged. T. Borsch acknowledges a Heisenberg Scholarship from the Deutsche Forschungsgemeinschaft. We thank Bill Summers (St. Louis), C. Barre Hellquist (North Adams), Fred Rich (Georgia Southern University), Kristi Woods (Virginia Tech), Alejandro Novelo (Universidad Nacional Autónoma de Méxcio), Charles Horn (Newberry College), and Pertti Uotila (Helsinki) and the staff of the Lammi Biological Station of the University of Helsinki for help during fieldwork. Chrissie C. Chawanje (Blyntare, Malawi), Surrey W. L. Jacobs (Sydney), Stefan Porembski (Rostock), Nur Ritter (Fresno), Udo Schmidt-Mumm (Colombia), and Eberhard Fischer (Koblenz) provided additional material. Thanks are due to the curator of the Bonn University Botanical Gardens, Wolfram Lobin, and Michael Neumann and Bernd Reinken, in charge of the *Nymphaea* living collections in the greenhouses, for their support. Moreover, this study benefitted from fruitful discussions and comments on earlier parts of the manuscript by David Dilcher (Gainesville), Jim Doyle (Davis), Kim Govers (Bonn), Dietmar Quandt (Dresden), and Kai Müller (Bonn).

Appendix A

List of Indels Found in the *trnT-trnF* Region

The *trnT-trnL* Spacer

1. "TGGG" indel in *Amborella*. According to the matrix from Borsch et al. (2003), it is also present in the gymnosperms. Thus, it appears to be a 4-nt deletion in the angiosperms above *Amborella*.
2. "TCWAC" present in all Nymphaeales, which may be the result of an early duplication event with subsequent substitutions.
3. Overlapping with 2. The gap in *Amborella* is 4 nt.
4. "CACATA"-SSR, completely present only in *Austrobaileya* (motif from pos. 2–13).
5. "A"-SSR, present in *Illicium floridanum* and overlapping with 4.
6. "ATATCTATCTATAT" indel in *Amborella*, which may be the result of several-length mutational events in the *Amborella* lineage involving adjacent sequence motifs. Nevertheless, they cannot be inferred.

7. "CTCACATAACATAA"-SSR in *Cabomba*.
8. "CATAA"-SSR in *Illicium*, *Austrobaileya*, and *Barclaya*, overlapping with 7.
9. Eight nucleotides missing in *Amborella* relative to *Nymphaea*, which has "AAATAAAA." Not possible to explain in one step. There is also no clear repeat motif.
10. Seven nucleotides missing in *Illicium*, *Austrobaileya*, and *Schisandra*; overlapping but not applicable in *Amborella*.
11. Six nucleotides missing in *Cabomba*, *Brasenia*, *Nuphar* spp., and *Barclaya*; overlapping but not applicable in *Amborella*, *Illicium*, *Austrobaileya*, and *Schisandra*.
12. Six nucleotides missing in *Amborella* relative to *Cabomba*, *Brasenia*, *Nuphar*, and *Barclaya*. No motif recognizable.
13. Three nucleotides missing in *Illicium*, *Austrobaileya*, and *Schisandra* relative to *Cabomba*, *Brasenia*, *Nuphar*, and *Barclaya*, overlapping with 12.
14. Two nucleotides missing in *Amborella*, whereas all remaining taxa have "KA." From this data set, it is not possible to infer whether this indel is apomorphic to *Amborella* Austrobaileyales plus Nymphaeales.
15. "TTAG"-SSR specific to *Nymphaea gigantea* (NY067).
16. "AAAATAAGTGAGTTAGTTCA"-SSR specific to *N. gardneriana* (NY026).
17. "GGGATCTTAGMTTATT"-SSR specific to *Brasenia schreberi* but overlapping with *Cabomba*.
18. "T"-SSR specific to *Cabomba* but overlapping with *Brasenia*.
19. "AATT" present in *Amborella*.
20. "CCGATCGGA" present in *Amborella* but a 9-nt gap in most other angiosperms but overlapping with 20 in *Illicium*, *Austrobaileya*, and *Schisandra*.
21. "C" present in *Illicium*, *Austrobaileya*, and *Schisandra*, resulting in an 8-nt gap in these three taxa.
22. "TATGAATATSAAT" present in *Illicium*, *Austrobaileya*, and *Schisandra*. Because similar motifs to reversed adjacent sequences are present, the indel could be the result of an ancient inverted repeat.
23. "CTTATTAT"-SSR in *Cabomba*.
24. "CTTATTA"-SSR in *Barclaya*, overlapping with 23.
25. Big gap in *Amborella* (length cannot be determined exactly).
26. A 12-nt gap in many taxa except *Nuphar*, overlapping with 25.
27. Gap unique in *Schisandra* but overlapping with 25.
28. Gap unique in *Cabomba* but overlapping with 25–27.
29. Gap in *Barclaya*, not present in *Nuphar* spp. but overlapping with 25–28.
30. Gap of 3 nt in *Illicium* but extending into hotspot; overlapping with 25.
31. Gap of 19 nt in Nymphaeales, overlapping with 25 (but both terminals different).
32. Gap of 8 nt in *Austrobaileya* and *Schisandra*, overlapping with 25 and 31.
33. "GGAT"-SSR in *Barclaya*.
34. "RTAAT"-SSR in *Austrobaileya*; the terminal nucleotide is substituted.
35. Five-nucleotide gap in *Amborella*, *Austrobaileya*, *Illicium*, and *Schisandra*. Nymphaeales have "TTATG."
36. One-nucleotide gap in *Nuphar*.
37. "KATTT"-SSR in *Nuphar*. A substitution in the flanking nucleotide in *Nuphar japonica*.
38. "GGAGA"-SSR in *Schisandra*; it is clearly a repeat structure, although the repeat motif has an additional "G" as compared with the adjacent sequence.
39. One-nucleotide gap in *Schisandra*.
40. A large gap in *Barclaya*, *Victoria*, *Euryale*, and *Nymphaea*.
41. A 1-bp shorter gap in *Cabomba*, *Brasenia*, and *Nuphar*, overlapping with 40. Interestingly, it is a fixed length difference in a polyAT strand.
42. "GAAAA"-SSR in *Illicium*, overlapping with 40 and 41.
43. "AA"-SSR, present in all angiosperms above *Amborella*. PolyA situation.
44. "A"-SSR in *Illicium*; polyA situation.
45. Three- to four-nucleotide gap in *Schisandra*, overlapping with 44.
46. One nucleotide in Nymphaeales ("R" in *Aborella* and Austrobaileyales).
47. One-nucleotide deletion in *Schisandra* (polyA situation).
48. "G"-SSR in *Cabomba*.
49. "TGGATATTC" present in *Amborella* but not in the other angiosperms. Motif not clear at this point.
50. One-nucleotide gap in *Brasenia*.
51. "CTATATTG"-SSR in *Amborella*.
52. "AATCATT"-SSR in *Amborella*.
53. "CTGATT"-SSR in Nymphaeales.
54. "TGAAC"-SSR in *N. odorata*.
55. "G"-SSR in *Illicium*.
56. "(TA)KATARAG"-SSR in *Cabomba* and *Brasenia*.
57. "AAGAK" present in *Amborella* and Nymphaeales, perhaps deletion in Nymphaeales.
58. "TAA + AAA" present in *Amborella*. Could be the result of a double repeat event.
59. "AGAAAGAA" present in *Amborella*.
60. Gap 1 nt shorter present in *Austrobaileya*, *Cabomba*, and *Nuphar*, overlapping with indel 59.
61. "GTKMAAA"-SSR present in *Nymphaea* subg. *Nymphaea*.
62. Obviously deletion of 8 nt in *Cabomba* and *Brasenia* and also in *N. colorata* and *N. caerulea*.
63. Obviously deletion of 7 nt in *Schisandra*, overlapping with 62.
64. Obviously deletion of 1 nt in polyT situation in *Cabomba*, *Brasenia*, and *Nuphar*.
65. Gap of 2 nt in *Amborella*, *Illicium*, *Austrobaileya*, and *Schisandra*.
66. Gap of one "C," which is obviously deleted adjacent to a polyT in *Victoria*; overlapping with 65.
67. Gap of 1 nt in *Amborella*.
68. Obviously deletion in *Euryale*, at the minimum 6 nt.
69. "TTTAA"-SSR in *Nuphar japonica* and *Nuphar lutea*, overlapping with 68.
70. Gap of at minimum 5 nt in *Amborella*.
71. "TGAATT"-SSR in *Austrobaileya*, overlapping with 70.
72. Deletion of at minimum 5 nt in *Cabomba*.
73. "GGAA"-SSR in *N. candida* NY109; overlapping with 72.
74. "AAAGAG+G"-SSR in *N. petersiana*; the downstream "G" can be explained only by a second SSR step.
75. At minimum 5-nt deletion in *Cabomba*.
76. "AAAT" present in *Amborella*, overlapping with 75.
77. A 4-nt deletion in *Cabomba*.
78. A 2-nt deletion in *Nuphar*.
79. "CTTA" present in *Amborella*.
80. At least 4-nt deletion in *Austrobaileya*.
81. "CAAA"-SSR in *Cabomba*, overlapping with 80.
82. An "A"-SSR in *Illicium*, overlapping with 81.
83. An at least 3-nt deletion in *Schisandra*, overlapping with 82–80.
84. An at least 5-nt deletion in *Amborella* and *Barclaya*.
85. "GGAAA" in *Victoria* and *Euryale*. Probably an insertion but of unknown origin.
86. "G"-SSR in *Illicium*, *Austrobaileya*, and *Schisandra*.
87. "RAAG" present in *Lotos*, *Brachyceras*, and *Anecphya*.

The *trnL* Intron

1. "T"-SSR in *Illicium*, overlapping with 2.
2. "TGTTT"-SSR in *Cabomba*, overlapping with 1.
3. Probably deletion of at least 8 nt in *Cabomba*, overlapping with 4.
4. "AAGTATTTCT" insertion in *Illicium*, of unknown origin and overlapping with 3.
5. A 2-nt deletion in *Illicium* (other taxa have "GG"); check whether this indel has something to do with 4.
6. A 7-nt indel in *Illicium*, *Austrobaileya*, and *Schisandra*, overlapping with 7.
7. A 5-nt indel in *Amborella*, overlapping with 6.
8. "TAGAA"-SSR in *Illicium*.
9. An 8- to 10-nt indel in *Amborella*, overlapping with 10 and 11.
10. *Nuphar lutea* is 3 nt shorter relative to *Illicium* and *Austrobaileya* and 1 nt relative to other Nymphaeales (1-nt deletion, in polyA situation).
11. "AA"-SSR in *Illicium* and *Austrobaileya*.
12. "GATAGG"-SSR in *Schisandra*.
13. "AATG"-SSR in *Nuphar*.
14. A 2-nt indel in *Cabomba*, *Brasenia*, *Nuphar*, and *Barclaya* and "CY" in *Amborella* + Austrobaileyales, overlapping with 15.
15. A 1-nt indel in *Victoria*, *Euryale*, and *Nymphaea*, overlapping with 14.
16. "AAGAW"-SSR in *Cabomba*.
17. A single "A" present in *Illicium*, origin unknown.
18. "GRAA"-SSR in *Amborella*.
19. A 4-nt deletion in *Austrobaileya*.
20. A 1-nt gap in *Amborella*.
21. "G" in *Amborella* and "A" in Austrobaileyales, whereas Nymphaeales have a gap.
22. "AGAA"-SSR in *Illicium*.
23. "TATR" probably ancient SSR in *Amborella* and Austrobaileyales; deletion in Nymphaeales?
24. "GGTATTG" in *Amborella*, gap in all other taxa; similar motifs occur in adjacent sequence parts, but repeat character is not evident.
25. "CTGAAATATCAA"-SSR in *Brachyceras*, *Ondinea*, and *Anecphya*.
26. Deletion of "TTAAT" in *Nymphaea mexicana*.
27. Deletion of "CGAAT" in *N. mexicana*; analyze structure and check whether 26 and 27 are connected.
28. Deletion of 2 nt in *Nuphar* (the first position is still in the P8 hotspot).
29. A 1-nt gap in *Cabomba*, probably deletion.
30. "TTGTG"-SSR in Nymphaeales.
31. Gap of at least 4 nt in *Amborella*; other taxa have "AATC"; overlapping with 32–34.
32. "AAAATA"-SSR in *N. conardii*, *N. gardneriana*, and *N. glandulifera*, overlapping with 31.
33. "AAATAT+AATATTA"-double SSR present in *Illicium*, overlapping with 31 and 34.
34. "AAATAT"-SSR, present in *Schisandra*, overlapping with 31.
35. "ATTAAG"-SSR in *Illicum*.
36. "AGRCTGGGAK" in Nymphaeales but gap in *Amborella* and Austrobaileyales; motif of microstructural change unclear.
37. Probably 13-nt deletion in Nymphaeales, overlapping with 38.
38. "TGAAGA"-SSR in Austrobaileyales, overlapping with 37.
39. Eight-nucleotide deletion in *Barclaya*.
40. "AGAAT"-SSR in *Amborella*.
41. "TTTG"-SSR in Nymphaeales.

The *trnL-trnF* Spacer

1. "CCCA"-SSR in *Nymphaea lotus thermarum* (NY003), repeated from *trnL* exon.
2. "AG+TCCCCA"-SSR in *N. colorata* and *N. caerulea*.
3. "CTAAAAAACA" in *Schisandra*, of unknown origin.
4. "AGAAAAAAGAATTTTTTTWAAAAGK" in *Austrobaileya*, of unknown origin.
5. "ATAAA"-SSR in *Illicium*, overlapping with 5.
6. "A"-SSR in *Amborella*, overlapping with 4.
7. "WAAAW"-SSR in *Illicium*, overlapping with 8 and 9.
8. "AAAG"-SSR in *Amborella*, overlapping with 7 and 9.
9. "AAA"-SSR in *Schisandra*, overlapping with 7 and 8.
10. A 2-nt deletion in *Austrobaileya*, overlapping with 11.
11. A 1-nt deletion in *Illicium* and *Schisandra*, overlapping with 10.
12. A 1-nt deletion in *Austrobaileya*.
13. At minimum 1-nt deletion in *Austrobaileya*, overlapping with 14 and 15.
14. "ATTTTCT" similar motifs in adjacent sequences but origin unclear.
15. "C"-SSR in *Amborella*.
16. "TTTTK"-SSR in *Austrobaileya* but "TTKCK"-SSR in *Cabomba* and *Brasenia*, overlapping with 17.
17. "T"-SSR in *Illicium* and *Nuphar*, overlapping with 16.
18. Big deletion in *Barclaya longiflora*.
19. "TTTTCK"-SSR in *Cabomba*, overlapping with 18.
20. "AKCCTCTTTTTTCGCCA"-SSR in *N. rudgeana*, *N. lingulata*, *N. oxypetala*, *N. lotus* subsp. *lotus*, *N. lotus* subsp. *thermalis* (N105), and *N. pubescens* (N406). Infraspecific variation in *N. lotus* points to homoplasy. Mechanism?
21. "AAATW"-SSR in *Euryale* and *Victoria*, overlapping with 18 and 22.
22. "GCGCTTCC()AAATTAGA" in *Victoria*, *Euryale*, *Nymphaea*, and *Ondinea* but missing in other taxa.
23. "AGAA"-SSR in *N. petersiana*, overlapping with 18.
24. "TAAACTAAAAC" present in *Amborella*; similar motifs in adjacent sequence but mechanism unclear.
25. "CAAATT"-SSR in *Schisandra*.
26. "GATAT"-SSR in Nymphaeales, overlapping with 27.
27. "T"-SSR in *Schisandra*. Could be a polyT situation (upstream are three T's) and thus independent from 26.
28. Probably 5-nt deletion in *Brachyceras*, *Ondinea*, and *Anecphya*; others have "STTTC."
29. Big deletion in *Illicium*.
30. "GATATGTTTATCATTC"-SSR in *N. lotus thermalis* (NY003), overlapping with 29.
31. Probably 4- to 5-nt deletion in Nymphaeales, overlapping with 29 and 32.
32. A 1-nt gap in *Amborella* relative to *Austrobaileya* and *Schisandra*.
33. A 2-nt deletion in *Barclaya*, overlapping with 29.
34. A 2-nt gap in *Amborella*, overlapping with 29.
35. A large gap in *Austrobaileya* and *Schisandra*, overlapping with 29.
36. A 2-nt gap in *Amborella*, overlapping with 29 and 35.
37. "TGTTGTTATTGTGAT" in *Amborella*, overlapping with 29 and 35 but gap in Nymphaeales.
38. "GCAGTAT"-SSR in *N. violacea* NY135, overlapping with 29 and 35.
39. "T"-SSR in *Cabomba*, overlapping with 29 and 35.
40. Big gap in *Amborella*, termini specific to *Amborella*, overlapping with 41, 29, +.
41. Gap of 1 nt in *Cabomba*, overlapping with 40, 29, 35.

42. "MCATAA"-SSR in *Brachyceras*, *Ondinea*, and *Anecphya*, overlapping with 40 and 29.
43. A 6- to 8-nt gap in *Austrobaileya* and *Schisandra*, overlapping with 29 and 40.
44. A 2-nt gap (other taxa have "AT"), overlapping with 29, 40, and 43.
45. "ACATAACCATAACATATGTA+TATGGTA"-SSR in *N. macrosperma* and *N. gigantea*.
46. Gaps in *N. petersiana*, *N. pubescens*, and *Hydrocallis*, overlapping with 29 and 40.
47. A gap of 2 nt.
48. An about 6-nt gap in *Austrobaileya*, *Schisandra*, *Euryale*, and subg. *Nymphaea*, overlapping with 29, 40, and 46.
49. A 5-nt gap present only in *Cabomba*, overlapping with 29, 40, 46, 48, and 49.
50. A gap of 4 nt in *Brasenia* and *Nuphar.*
51. "TG"-SSR in *Barclaya*, overlapping with 29, 40, 46, 47, and 48.
52. "WTGATT"-SSR in *Cabomba*, overlapping with 29 and 40.
53. "ATAWAT" in *Austrobaileya* and *Schisandra*, overlapping with 29, 40, and 54.
54. "WGKATA" in *Austrobaileya*, *Schisandra*, and *Cabomba*, overlapping with 29 and 40.
55. "TGTA"-SSR in *Ondinea*, overlapping with 29 and 40.
56. Gap of 2 nt in *Schisandra*, overlapping with 29 and 40.
57. At minimum a gap of 4 nt in *Brasenia*, overlapping with 29 and 40.
58. "GAASATAK" in *Austrobaileya*, *Schisandra*; *Amborella* has only "AACATAT" (gap 40) and *Cabomba* only "C."
59. Relative to 58, a 7-nt gap in *Cabomba*.
60. "TTTGTATACAAG"-SSR in *N. lingulata*, overlapping with 29 and 40.
61. "YCCCCA"-SSR in *Schisandra*, overlapping with 29.
62. A 1-nt gap in Nymphaeales, whereas *Amborella*, *Austrobaileya*, and *Schisandra* have a "C," a polyC situation, overlapping with 29.
63. "TTTAAT"-SSR in *N. hastifolia*.
64. A 14-nt gap in *Illicium*, overlapping with 65.
65. A 10-nt gap in *Amborella*, overlapping with 64.
66. Gap of 14 nt in Nymphaeales, overlapping with 67–69.
67. Gap of 6 nt in *Illicium*, overlapping with 66, 68, and 69.
68. Gap of 3 nt in *Amborella* and Austrobaileya, overlapping with 66 and 67.
69. Gap of 1 nt in *Amborella*, overlapping with 66 and 67.
70. "YAAA"-SSR in *Illicium*, *Austrobaileya*, and *Schisandra*.
71. "ACAAAG"-SSR in subg. *Hydrocallis* and subg. *Lotos*, except *N. pubescens* and *N. petersiana*.
72. "A"-SSR in *N. lotus*, overlapping with 71.
73. "ABAAACA" in *Illicium* and *Schisandra*, overlapping with 74.
74. "TAAACA" in Austrobaileya, overlapping with 73; could be of repeat origin but highly modified.
75. "AAATAAAGG"-SSR in *N. violacea* (NY110) and *N. elleniae*, overlapping with 76.
76. "AAAGG"-SSR in *Amborella*, overlapping with 75.
77. "GTCAAATCCAA" present in *Amborella*; motif could be derived from repeats but highly modified.
78. Large gap in *Amborella*, overlapping with 79, 81, and 82.
79. A 7-nt indel with "AGAAAA" in Nymphaeales, which could be of repeat origin because similar motifs are adjacent.
80. A large gap in Nymphaeales, overlapping with 81 and 82; 79 also overlaps but has two different terminals.
81. A 7-nt gap in *Austrobaileya*, overlapping with 78, 80, and 82.
82. A 1-nt gap in *Illicium*, overlapping with 78, 80, and 81.
83. "TTAGTACCTTTAAA"-SSR in *Victoria*.
84. "ATTTAGTTTAG" in *Amborella*, similar motifs occur adjacently, so that some repeat events might have occurred in generating this sequence.
85. "AAWTG"-SSR in Nymphaeales.
86. "CACA"-SSR in *Barclaya*.
87. An 8-nt deletion in *Brasenia*.
88. "CTCCA"-SSR in *Barclaya*.
89. "ATRCA"-SSR in *Amborella*.

The P8 Stem Loop in the *trnL* Intron

1. "TCTTATTTATAATTATTAATATAT"-SSR in *Nymphaea violacea* (NY135).
2. "ATKAAATAT" in all BOA species except *N. violaea* 135, *N. atrans*, and *N. immutabilis*.
3. "TTMAATAT"-SSR in *Ondinea*, *N. hastifolia*, *N. violacea* 110, and *N. elleniae*.
4. "TAATTAATA"-SSR in *N. belophylla*, *N. tenerinervia*, *N. potamophila*, *N. rudgeana*, *N. lingulata*, and *N. oxypetala.*
5. A 25-nt gap in *N. gardneriana*.
6. An 18-nt gap in *N. lingulata*, overlapping with 5.
7. A 16-nt gap in *N. oxypetala*, *N. conardii*, *N. glandulifera*, and *N. jamesoniana*, overlapping with 5 and 6.
8. A 9-nt gap in *N. rudgeana*, overlapping with 5–7.
9. "TATA"-SSR in *N. petersiana*.
10. "TAA" in *N. pubescens* and *N. petersiana* but lacking in *N. lotus*; motif unclear except inverted repeat.
11. An 8-nt gap in *N. pubescens*, overlapping with 12 and 13.
12. A 1-nt gap in *N. lotus thermalis*; could be a "T"-SSR in *N. lotus*.
13. A 4-nt gap in *N. petersiana*, overlapping with 11; motif not really clear. (Indel 27 from *trnL* extends into P8 by 3 nt; it is not coded here, as it was already coded for *trnL*.)
14. "TA"-SSR in *N. mexicana*, overlapping with 15.
15. "TATATA" present in all but *N. odorata*.
16. A 20-nt gap in *N. tetragona*, overlapping with 17; probably deletion.
17. "TAATTAATATT" in *N. mexicana*; can be partially explained by a repeated motif.

Appendix B

Table B1

Binary Matrix of Indels Coded

Nymphaea trnT-trnL, 87 indels:

Amborella_trichopoda_N116_AY145324	1100-10-0--0-0000-11100-0-----1-000100111000111011110000111101110-0100-100111111---00--
Illicium_floridanum_N117_AY145325	00-1100010-101000-01010-101--011000100111111111001000010100-01110-110101001011011001
Austrobaileya_scandens_N115_AY145326	00-1000010-101000-01010-111--110010100111010111001000000101001110-11011100101100--11001
Schisandra_rubi_N151_AY145327	00-0-00-10-101000-01010-1-0--11000010101101-010001000000100-01010-11010100101101--01001
Cabomba_caroliniana_N112_AY145328	0110-011110111001000-0111110-10-001100110-101011010010010010000-011110100-00-010111110-0
Brasenia_schreb_N106_AY145329	0110-00-110111001100-00-101--10-001100110-101010000010010000-00-011110101001011010111 0-0
Nuphar_japonica_N400	0110-00-110111000-00-00-1111110-001010110-101010010010000010011011111110100101001011 10-0
Nuphar_lutea_N107_AY145330	0110-00-110111000-00-00-1111110-001010110-101010010010000010011011111110100101001011 10-0
Nuphar_advena_N080_AY145351	0110-00-110111000-00-00-1111110-001010110-101010010010000010011011111010010010010111 0-0
Nuphar_specBoWieb_N108	0110-00-110111000-00-00-1111110-001010110-101010010010000010011011111010100100101011 10-0
Barclaya_longiflora_N376	0110-000110111000-00-0101111010-10110010--101010010010000000-011111110101001011010 1-00-0
Victoria_N316	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011110101010010110101111-0
Victoria_Longwood_N378	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011110101010010110101111-0
Euryale_ferox_N015	0110-00-111111000-00-00-101--10-00110010--101010010010000000-01111110-1010010110101111-0
Nymphaea_tetragona_NY060	0110-00-111111000-00-00-101--10-00110010--101010010010000000-111111110101001011010111 0-0
Nymphaea_alba_NY056	0110-00-111111000-00-00-101--10-00110010--101010010010000000-111111110101001011010111 0-0
Nymphaea_alba_NY061	0110-00-111111000-00-00-101--10-00110010--101010010010000000-111111110101001011010111 0-0
Nymphaea_candida_NY063	0110-00-111111000-00-00-101--10-00110010--101010010010000000-111111110101001011010111 0-0
Nymphaea_candida_NY062	0110-00-111111000-00-00-101--10-00110010--101010010010000000-111111110101001011010111 0-0
Nymphaea_candida_NY109	0110-00-111111000-00-00-101--10-00110010--101010010010000000-111111110101101011010111 0-0
Nymphaea_odorata_odo_N012_AY145333	0110-00-111111000-00-00-101--10-00110010--101010010011000000-111111110101001011010111 0-0
Nymphaea_odorata_tub_NY269	0110-00-111111000-00-00-101--10-00110010--101010010011000000-111111110101001011010111 0-0
Nymphaea_mexicana_NY069	0110-00-111111000-00-00-101--10-00110010--101010010010000000-111111110101001011010111 0-0
Nymphaea_mexicana_KN008	0110-00-111111000-00-00-101--10-00110010--101010010010000000-111111110101001011010111 0-0
Nymphaea_belophylla_NY027	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_tenerinervia_NY140	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_amazonum_NY428	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_novogranatensis_NY021	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_potamophila_NY389	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_rudgeana_NY032	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_rudgeana_NY124	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_lingulata_NY029	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_oxypetala_NY387	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_conardii_NY022	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_gardneriana_NY026	0110-00-111111010-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_glandulifera_NY390	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_jamesoniana_NY071	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_jamesoniana_NY098	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111 0-0
Nymphaea_lotus_therm_NY003	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111011
Nymphaea_lotus_therm_NY105	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111011
Nymphaea_lotus_lotus_NY078	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111011
Nymphaea_pubescens_NY406	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111011
Nymphaea_petersiana_NY058	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101011011010111011
Nymphaea_gracilis_NY025	0110-00-111111000-00-00-101--10-00110010--101010010010000000-011111110101001011010111011

Nymphaea_gracilis_NY429	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_elegans_NY006	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_elegans_NY370	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_pulchella_NY100	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_colorata_NY122	0110-00-111111000-00-00-101--10-00110010--10101001001000000-00-1111101010010110101111011
Nymphaea_caerulea_NY113	0110-00-111111000-00-00-101--10-00110010--10101001001000000-00-1111101010010110101111011
Nymphaea_micrantha_NY007	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_cf_nouchalii_NY066	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_thermarum_NY065	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Ondinea_purpurea_N377	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_hastifolia_NY134	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_violacea_8230_NY110	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_elleniae_8227_NY137	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_elleniae_NY103	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_violacea_NY135	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_macrosperma_NY127	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_gigantea_NY067	0110-00-111111100-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_gigantea_NY126	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_atrans_8212_NY102	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_imm_at_NY121	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea_immutabilis_NY136	0110-00-111111000-00-00-101--10-00110010--10101001001000000-0111111101010010110101111011
Nymphaea trnL intron, 41 indels:	
Amborella_trichopoda_N116_AY145324	-01011000--0011001101011011110 0---0010110
Illicium_floridanum_N117_AY145325	011100-1111001101011111100111101011101 1100
Austrobaileya_scandens_N115_AY145326	-01010-0111001100001101001111010 0-0011100
Schisandra_rubi_N151_AY145327	-01010-0110101100011101001111010100011100
Cabomba_caroliniana_N112_AY145328	110-1110110000-1001100000111011 00-010-101
Brasenia_schreb_N106_AY145329	-0101110110000-0001100000111111 00-010-101
Nuphar_japonica_N400	-0101110110010-0001100000111011 00-010-101
Nuphar_lutea_N107_AY145330	-010111010-010-0001100000111011 00-010-101
Nuphar_advena_N080_AY145351	-0101110110010-0001100000111011 00-010-101
Nuphar_specBoWieb_N108	-0101110110010-0001100000111011 00-010-101
Barclaya_longiflora_N376	-0101110110000-0001100000111111 00-010-001
Victoria_N316	-0101110110001000011000001111 1100-010-101
Victoria_Longwood_N378	-0101110110001000011000001111 1100-010-101
Euryale_ferox_N015	-0101110110001000011000001111 1100-010-101
Nymphaea_tetragona_NY060	-0101110110001000011000001111 1100-010-101
Nymphaea_alba_NY056	-0101110110001000011000001111 1100-010-101
Nymphaea_alba_NY061	-0101110110001000011000001111 1100-010-101
Nymphaea_candida_NY063	-0101110110001000011000001111 1100-010-101
Nymphaea_candida_NY062	-0101110110001000011000001111 1100-010-101
Nymphaea_candida_NY109	-0101110110001000011000001111 1100-010-101
Nymphaea_odorata_odo_N012_AY145333	-0101110110001000011000001111 1100-010-101
Nymphaea_odorata_tub_NY269	-0101110110001000011000001111 1100-010-101
Nymphaea_mexicana_NY069	-0101110110001000011000000011 1100-010-101
Nymphaea_mexicana_KN008	-0101110110001000011000000011 1100-010-101
Nymphaea_belophylla_NY027	-0101110110001000011000001111 1100-010-101
Nymphaea_tenerinervia_NY140	-0101110110001000011000001111 1100-010-101
Nymphaea_amazonum_NY428	-0101110110001000011000001111 1100-010-101

Table B1

(*Continued*)

Nymphaea_novogranatensis_NY021	-0101110110001000011000001111111100-010-101
Nymphaea_potamophila_NY389	-0101110110001000011000001111111100-010-101
Nymphaea_rudgeana_NY032	-0101110110001000011000001111111100-010-101
Nymphaea_rudgeana_NY124	-0101110110001000011000001111111100-010-101
Nymphaea_lingulata_NY029	-0101110110001000011000001111111100-010-101
Nymphaea_oxypetala_NY387	-0101110110001000011000001111111100-010-101
Nymphaea_conardii_NY022	-0101110110001000011000001111111110-010-101
Nymphaea_gardneriana_NY026	-0101110110001000011000001111111110-010-101
Nymphaea_glandulifera_NY390	-0101110110001000011000001111111110-010-101
Nymphaea_jamesoniana_NY071	-0101110110001000011000001111111100-010-101
Nymphaea_jamesoniana_NY098	-0101110110001000011000001111111100-010-101
Nymphaea_lotus_therm_NY003	-0101110110001000011000001111111100-010-101
Nymphaea_lotus_therm_NY105	-0101110110001000011000001111111100-010-101
Nymphaea_pubescens_NY406	-0101110110001000011000001111111100-010-101
Nymphaea_petersiana_NY058	-0101110110001000011000001111111100-010-101
Nymphaea_gracilis_NY025	-0101110110001000011000001111111100-010-101
Nymphaea_gracilis_NY429	-0101110110001000011000001111111100-010-101
Nymphaea_elegans_NY006	-0101110110001000011000001111111100-010-101
Nymphaea_elegans_NY370	-0101110110001000011000001111111100-010-101
Nymphaea_pulchella_NY100	-0101110110001000011000001111111100-010-101
Nymphaea_colorata_NY122	-0101110110001000011000001111111100-010-101
Nymphaea_caerulea_NY113	-0101110110001000011000001111111100-010-101
Nymphaea_micrantha_NY007	-0101110110001000011000001111111100-010-101
Nymphaea_cf_nouchalii_NY066	-0101110110001000011000001111111100-010-101
Nymphaea_thermarum_NY065	-0101110110001000011000001111111100-010-101
Ondinea_purpurea_N377	-0101110110001000011000001111111100-010-101
Nymphaea_hastifolia_NY134	-0101110110001000011000001111111100-010-101
Nymphaea_violacea_8230_NY110	-0101110110001000011000001111111100-010-101
Nymphaea_elleniae_8227_NY137	-0101110110001000011000001111111100-010-101
Nymphaea_elleniae_NY103	-0101110110001000011000001111111100-010-101
Nymphaea_violacea_NY135	-0101110110001000011000001111111100-010-101
Nymphaea_macrosperma_NY127	-0101110110001000011000001111111100-010-101
Nymphaea_gigantea_NY067	-0101110110001000011000001111111100-010-101
Nymphaea_gigantea_NY126	-0101110110001000011000001111111100-010-101
Nymphaea_atrans_8212_NY102	-0101110110001000011000001111111100-010-101
Nymphaea_imm_at_NY121	-0101110110001000011000001111111100-010-101
Nymphaea_immutabilis_NY136	-0101110110001000011000001111111100-010-101
Nymphaea trnL-trnF, 90 indels:	
Amborella_trichopoda_N116_AY145324	-000000111111010-100-00100-1101010111000-----------------11-01010110000-0-1010-1--00100101
Illicium_floridanum_N117_AY145325	-0001111110110000100-00000-10---1-----------------------------00-1---10-110-01011010000100
Austrobaileya_scandens_N115_AY145326	-0010-00-0-00-011100-00000-11011110----1-00-0110---0110111100101111011O-100-01010-10000100
Schisandra_rubi_N151_AY145327	-0100-0001011000-100-000110110111O----1-00-0110---0110011101101111111O-110-01011110000100
Cabomba_caroliniana_N112_AY145328	-0000-00-11111011110-0000111100-1110001100110111O-01010111000-0110---00-0-0-0110--10010100
Brasenia_schreb_N106_AY145329	-0000-00-11110011100-0000111100-11100001101001111000-0010--00-0110---00-0-0-0110--10010000
Nuphar_japonica_N400	-0000-00-11110000100-0000111100-11100001101001111000-00110-00-0110---00-0-0-0110-10010100
Nuphar_lutea_N107_AY145330	-0000-00-11110000100-0000111100-11100001101001111000-00110-00-0110---00-0-0-0110--10010100
Nuphar_advena_N080_AY145351	-0000-00-11110000100-0000111100-11100001101001111000-00110-00-0110---00-0-0-0110--10010100

```
Nuphar_specBoWieb_N108                -0000-00-11110000100-0000111100-11100001101001111000-00110-00-0110---00-0-0-0110--10010100
Barclaya_longiflora_N376              -0000-00-1111000-0-----00111100-01100001101101111010-00110-00-0110---00-0-0-0110--10011110
Victoria_N316                         -0000-00-1111000-10011000111100-11100001101101111100-00110-00-0110---00-0-0-0110--11010100
Victoria_Longwood_N378                -0000-00-1111000-10011000111100-11100001101101111100-00110-00-0110---00-0-0-0110--11010100
Euryale_ferox_N015                    -0000-00-1111000-10011000111100-1110000110110110---0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_tetragona_NY060              -0000-00-1111000-10001000111100-1110000110110110---0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_alba_NY056-0000              -00-1111000-10001000111100-1110000110110110---0-00110-00-0110---00-0-0-0110-10010100
Nymphaea_alba_NY061                   -0000-00-1111000-10001000111100-1110000110110110---0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_candida_NY063                -0000-00-1111000-10001000111100-1110000110110110---0-00110-00-0110---00-0-0-0110-10010100
Nymphaea_candida_NY062                -0000-00-1111000-10001000111100-1110000110110110---0-00110-00-0110---00-0-0-0110-10010100
Nymphaea_candida_NY109                -0000-00-1111000-10001000111100-1110000110110110---0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_odorata_odo_N012_AY145333    -0000-00-1111000-10001000111100-1110000110110110---0-00110-00-0110---00-0-0-0110-10010100
Nymphaea_odorata_tub_NY269            -0000-00-1111000-10001000111100-1110000110110110---0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_mexicana_NY069               -0000-00-1111000-10001000111100-1110000110110110---0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_mexicana_KN008               -0000-00-1111000-10001000111100-1110000110110110---0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_belophylla_NY027             -0000-00-1111000-10001000111100-111000011011OO-----0-00110-00-0110---0100-0-0110--10010100
Nymphaea_tenerinervia_NY140           -0000-00-1111000-10001000111100-111000011011OO-----0-00110-00-0110---0100-0-0110--10010100
Nymphaea_amazonum_NY428               -0000-00-1111000-10001000111100-111000011011OO-----0-00110-00-0110---0100-0-0110--10010100
Nymphaea_novogranatensis_NY021        -0000-00-1111000-10001000111100-111000011011OO-----0-00110-00-0110---0100-0-0110--10010100
Nymphaea_potamophila_NY389            -0000-00-1111000-10001000111100-111000011011OO-----0-00110-00-0110---0100-0-0110--10010100
Nymphaea_rudgeana_NY032               -0000-00-1111000-10101000111100-111000011011OO-----0-00110-00-0110---0100-0-0110-10010100
Nymphaea_rudgeana_NY124               -0000-00-1111000-10101000111100-111000011011OO-----0-00110-00-0110---0100-0-0110--10010100
Nymphaea_lingulata_NY029              -0000-00-1111000-10101000111100-111000011011OO-----0-00110-10-0110---0100-0-0110--10010100
Nymphaea_oxypetala_NY387              -0000-00-1111000-10101000111100-111000011011OO-----0-00110-00-0110---0100-0-0110--10010100
Nymphaea_conardii_NY022               -0000-00-1111000-10001000111100-111000011011OO-----0-00110-00-0110---0100-0-0110--10010100
Nymphaea_gardneriana_NY026            -0000-00-1111000-10001000111100-111000011011OO-----0-00110-00-0110---0100-0-0110--10010100
Nymphaea_glandulifera_NY390           -0000-00-1111000-10001000111100-111000011011OO-----0-00110-00-0110---0100-0-0110--10010100
Nymphaea_jamesoniana_NY071            -0000-00-1111000-10001000111100-111000011011OO-----0-00110-00-0110---0100-0-0110--10010100
Nymphaea_jamesoniana_NY098            -0000-00-1111000-10001000111100-111000011011OO-----0-00110-00-0110---0100-0-0110--10010100
Nymphaea_lotus_therm_NY003            01000-00-1111000-10001000111110-1110000110110101O1-0-00110-00-0110---0110-0-0110--10010100
Nymphaea_lotus_therm_NY105            -0000-00-1111000-10101000111100-1110000110110101O1-0-00110-00-0110---0110-0-0110--10010100
Nymphaea_lotus_lotus_NY078            -0000-00-1111000-10101000111100-1110000110110101O1-0-00110-00-0110---0110-0-0110--10010100
Nymphaea_pubescens_NY406              -0000-00-1111000-10101000111100-111000011011OO-----0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_petersiana_NY058             -0000-00-1111000-10001100111100-111000011011OO-----0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_gracilis_NY025               -0000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_gracilis_NY429               -0000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_elegans_NY006                -0000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-0-0110-10010100
Nymphaea_elegans_NY370                -0000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_pulchella_NY100              -0000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_colorata_NY122               11000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_caerulea_NY113               11000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_micrantha_NY007              -0000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_cf_nouchalii_NY066           -0000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-0-0110--10010100
Nymphaea_thermarum_NY065              -0000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-0-0110--10010100
Ondinea_purpurea_N377                 -0000-00-1111000-10001000110100-1110000111100101O1-0-01110-00-0110---00-0-0-0110-10010100
Nymphaea_hastifolia_NY134             -0000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-1110---00-0-0-0110-10010100
Nymphaea_violacea_8230_NY110          -0000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-110110-10010100
Nymphaea_elleniae_8227_NY137          -0000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-110110-10010100
Nymphaea_elleniae_NY103               -0000-00-1111000-10001000110100-1110000111100101O1-0-00110-00-0110---00-0-110110-10010100
Nymphaea_violacea_NY135               -0000-00-1111000-10001000110100-1110010111100101O1-0-00110-00-0110---00-0-000110-10010100
```

Table B1

(*Continued*)

Nymphaea_macrosperma_NY127	-0000-00-1111000-10001000110100-11100001111011011 1-0-00110-00-0110---00-0-000110-10010100
Nymphaea_gigantea_NY067	-0000-00-1111000-10001000110100-11100001111011011 1-0-00110-00-0110---00-0-000110-10010100
Nymphaea_gigantea_NY126	-0000-00-1111000-10001000110100-11100001111011011 1-0-00110-00-0110---00-0-000110-10010100
Nymphaea_atrans_8212_NY102	-0000-00-1111000-10001000110100-11100001111001011 1-0-00110-00-0110---00-0-000110-10010100
Nymphaea_imm_at_NY121	-0000-00-1111000-10001000110100-11100001111001011 1-0-00110-00-0110---00-0-000110-10010100
Nymphaea_immutabilis_NY136	-0000-00-1111000-10001000110100-11100001111001011 1-0-00110-00-0110---00-0-000110-10010100
Indels from P8, 17 indels:	
Amborella_trichopoda_N116_AY145324	----------
Illicium_floridanum_N117_AY145325	----------
Austrobaileya_scandens_N115_AY145326	----------
Schisandra_rubi_N151_AY145327	----------
Cabomba_caroliniana_N112_AY145328	----------
Brasenia_schreb_N106_AY145329	----------
Nuphar_japonica_N400	----------
Nuphar_lutea_N107_AY145330	----------
Nuphar_advena_N080_AY145351	----------
Nuphar_specBoWieb_N108	----------
Barclaya_longiflora_N376	----------
Victoria_N316	----------
Victoria_Longwood_N378	----------
Euryale_ferox_N015	----------
Nymphaea_tetragona_NY060	----------010-
Nymphaea_alba_NY056	----------0110
Nymphaea_alba_NY061	----------0110
Nymphaea_candida_NY063	----------0110
Nymphaea_candida_NY062	----------0110
Nymphaea_candida_NY109	----------0110
Nymphaea_odorata_odo_N012_AY145333	----------010
Nymphaea_odorata_tub_NY269	----------010
Nymphaea_mexicana_NY069	----------1111
Nymphaea_mexicana_KN008	----------1111
Nymphaea_belophylla_NY027	---11111---------
Nymphaea_tenerinervia_NY140	---11111---------
Nymphaea_amazonum_NY428	---01111---------
Nymphaea_novogranatensis_NY021	---11111---------
Nymphaea_rudgeana_NY032	---11110---------
Nymphaea_rudgeana_NY124	---11110---------
Nymphaea_lingulata_NY029	---110----------
Nymphaea_oxypetala_NY387	---1110----------
Nymphaea_conardii_NY022	---0110----------
Nymphaea_gardneriana_NY026	---00-----------
Nymphaea_glandulifera_NY390	---0110----------
Nymphaea_jamesoniana_NY071	---0110----------
Nymphaea_jamesoniana_NY098	---0110----------
Nymphaea_lotus_therm_NY003	--------00101----
Nymphaea_lotus_therm_NY105	--------00101----
Nymphaea_lotus_lotus_NY078	--------00111----
Nymphaea_pubescens_NY406	--------010------
Nymphaea_petersiana_NY058	--------111-0----

Appendix C

Table C1

Samples Included in This Study

Taxon	Field/garden origin	Voucher	Code	GenBank no.
Angiosperms (other than Nymphaeales):				
Amborella trichopoda Baill.	University of California, Santa Catarina Bot Gard	T. Borsch 3480 (BONN, VPI)	N116	AY145324[a]
Austrobaileya scandens C. T. White	Bonn Bot Gard 09789 ex BG Zürich	T. Borsch 3464 (BONN)	N115	AY145326[a]
Illicium floridanum J. Ellis	USA, Florida	T. Borsch & V. Wilde 3104 (VPI, FR)	N117	AY145325[a]
Schisandra rubriflora Rehder & E. H. Wilson	Bonn Bot Gard 0727 ex BG Munich	T. Borsch 3477 (BONN)	N151	AY145327[a]
Nymphaeales (other than *Nymphaea*):				
Barclaya longifolia Wall.	Water gardening source	C. Löhne 60 (BONN)	NY376	AM422019
Brasenia schreberi J. F. Gmel.	USA, Virginia	T. Borsch & T. Wieboldt 3311 (VPI, FR)	NY106	AY145329[a]
Cabomba caroliniana A. Gray	USA, Virginia	J. C. Ludwig s.n. (VPI)	NY112	AY145328[a]
Euryale ferox Salisb.	Bonn Bot Gard (14010)	T. Borsch 3830 (BONN)	NY015	AM422020
Nuphar advena (Aiton) W. T. Aiton	USA, Florida	T. Borsch & V. Wilde 3093 (FR)	NY080	AY145351[a]
Nuphar lutea (L.) Sm.	Germany, Hesse	. Borsch 3337 (FR)	NY107	AY145330[a]
Nuphar advena (Aiton) W. T. Aiton subsp. *advena*	USA, Virginia	T. Borsch & T. Wieboldt 3298 (VPI, BONN)	NY108	AM422021
Nuphar japonica DC.	Bonn Bot Gard [aquarium plant]	C. Löhne 61 (BONN)	NY400	AM422022
Ondinea purpurea Hartog	Australia, Western Australia	S. W. L. Jacobs & C. B. Hellquist 8853 (NSW)	NY377	AM422023
Victoria cruziana A. D. Orb.	Bonn Bot Gard	C. Löhne 55 (BONN)	NY316	AM422024
Victoria "Longwood Hybrid"	Bonn Bot Gard	T. Borsch 3831 (BONN)	NY378	AM422025
Nymphaea subg. *Hydrocallis*:				
N. amazonum Mart. & Zucc.	Mexico, Veracruz	A. Novelo R., J. H. Wiersema, C. B. Hellquist & C. N. Horn 1281 (MEXU)	NY428	AM422026
N. belophylla Trickett	Colombia, Meta	U. Schmidt-Mumm 942 (no voucher)	NY027	AM422027
N. conardii Wiersema	Mexico, Veracruz	A. Novelo R., J. H. Wiersema, C. B. Hellquist & C. N. Horn 1306 (MEXU)	NY022	AM422028
N. gardneriana Planch.	Guyana, Upper Takutu–Upper Essequibo Distr.	C. N. Horn & J. H. Wiersema 10084 (US, BRG, NBYC)	NY026	AM422029
N. glandulifera Rodschied Guyana.	Pomeroon Distr.	C. N. Horn & J. H. Wiersema 4523 (US, BRG, NBYC)	NY390	AM422030
N. lingulata Wiersema	Guyana, Upper Takutu–Upper Essequibo Dist.	C. N. Horn & J. H. Wiersema 11000 (US, BRG, NBYC)	NY029	AM422031
N. jamesoniana Planch.	USA, Florida	T. Borsch & B. Summers 3220 (BONN, MO)	NY071	AM422032
N. jamesoniana Planch.	Ecuador	M. Schwerdtfeger (BONN, GOET)	NY098	AM422033
N. novogranatensis Wiersema	Mexico, Oaxaca	A. Novelo R. & J. H. Wiersema 1187 (MEXU)	NY021	AM422034
N. oxypetala Planch.	Bolivia, Santa Cruz	N. Ritter, G. E. Crow, M. Garvizu, & C. Crow 4491 (NHA)	NY387	AM422035
N. potamophila Wiersema Guyana	Upper Takutu–Upper Essequibo Distr.	C. N. Horn & J. H. Wiersema 11090 (US, BRG, NBYC)	NY389	AM422036
N. rudgeana G. Mey.	Guyana, Mahaica–Berbice Distr.	C. N. Horn, S. Hill, & D. Gopaul 10045 (US, BRG, NBYC)	NY032	AM422037
N. rudgeana G. Mey.	BG Bonn 1088 [Guyana]	Koehnen s.n. (BONN)	NY124	AM422038
N. tenerinervia Caspary	Guyana, Upper Takutu–Upper Essequibo Distr.	C. N. Horn & J. H. Wiersema 11086 (US, BRG, NBYC)	NY140	AM422039
Nymphaea subg. *Lotos*:				
N. lotus L. var. *thermalis* (DC.) Tuzson.	Bonn Bot Gard 11547-11 [Romania]	T. Borsch 3832 (BONN)	NY003	AM422040
N. lotus L. var. *thermalis* (DC.) Tuzson.	Bonn Bot Gard 05553 [Romania]	T. Borsch 3833 (BONN)	NY105	AM422041
N. lotus L. var. *lotus*	Ivory Coast	S. Porembski s.n. (no voucher)	NY078	AM422042
N. pubescens Willd.	Australia, Northern Territory	S. W. L. Jacobs 8798 (NSW)	NY406	AM422043
Nymphaea subg. *Brachyceras*:				
N. ampla (Salisb.) DC.	Mexico, Veracruz	A. Novelo R., J. H. Wiersema, C. B. Hellquist, & C. N. Horn 1295 (MEXU)	NY100	AM422044
N. caerulea Savign.	Bonn Bot Gard 13783	T. Borsch 3834 (BONN)	NY113	AM422045
N. cf. colorata Peter	Bonn Bot Gard 1073	T. Borsch 3835 (BONN)	NY122	AM422046

Table C1

(*Continued*)

Taxon	Field/garden origin	Voucher	Code	GenBank no.
N. elegans Hook.	USA, Florida, Collier Co.	T. Borsch & V. Wilde 3084 (FR)	NY006	AM422047
N. elegans Hook.	USA, Louisiana, Cameron Parish	T. Borsch & K. Woods 3424 (BONN, VPI)	NY370	AM422048
N. gracilis Zucc.	Mexico, Michoacan	A. Novelo R., J. H. Wiersema, C. B. Hellquist, & C. N. Horn 1346 (MEXU)	NY025	AM422049
N. gracilis Zucc.	Mexico, Jalisco	A. Novelo R., J. H. Wiersema, C. B. Hellquist, & C. N. Horn 1314 (MEXU)	NY429	AM422050
N. micrantha Guill. & Perr.	Bonn Bot Gard 5830 [Zimbabwe]	M. Koehnen s.n. (BONN)	NY007	AM422051
N. heudelotii Burm. f.	Bonn Bot Gard 14244 [Rwanda]	E. Fischer s.n. (BONN)	NY066	AM422052
N. petersiana Klotzsch	Malawi	C. Chawanje s.n. (BONN, FR)	NY058	AM422053
N. thermarum Eb. Fisch.	Bonn Bot Gard 12088 [Rwanda]	E. Fischer s.n. (BONN)	NY065	AM422054
Nymphaea subg. *Anecphya*:				
N. atrans S. W. L. Jacobs	Australia, Queensland	S. W. L. Jacobs, C. B. Hellquist, & J. H. Wiersema 8212 (NASC, NSW, BRI)	NY102	AM422055
N. elleniae S. W. L. Jacobs	Australia, Queensland	S. W. L. Jacobs, C. B. Hellquist, & J. H. Wiersema 8224 (NASC, NSW, BRI)	NY103	AM422056
N. elleniae S. W. L. Jacobs	Australia, Queensland	S. W. L. Jacobs, C. B. Hellquist, & J. H. Wiersema 8227 (NASC, NSW, BRI)	NY137	AM422057
N. gigantea Hook.	Bonn Bot Gard 1728	T. Borsch 3836 (BONN)	NY067	AM422058
N. gigantea Hook.	Australia, Queensland	S. W. L. Jacobs & C. B. Hellquist 8357 (NASC, NSW, BRI)	NY126	AM422059
N. hastifolia Domin	Australia, Northern Territory, Darwin	J. H. Wiersema & C. B. Hellquist s.n. (no voucher)	NY134	AM422060
N. immutabilis S. W. L. Jacobs	Australia, Queensland, Cabbage Creek	S. W. L. Jacobs, C. B. Hellquist, & J. H. Wiersema s.n. (no voucher)	NY121	AM422061
N. immutabilis S. W. L. Jacobs	Australia, Queensland, Mt. Molloy	S. W. L. Jacobs, C. B. Hellquist, & J. H. Wiersema s.n. (no voucher)	NY136	AM422062
N. macrosperma Merr. & L. M. Perry	Australia, Northern Territory	C. B. Hellquist, J. H. Wiersema, & K. Brennan 16181 (MASS)	NY127	AM422063
N. violacea Lehm.	Australia, Queensland	S. W. L. Jacobs, C. B. Hellquist, & J. H. Wiersema 8230 (NASC, NSW, BRI)	NY110	AM422064
N. violacea Lehm.	Australia, Queensland	S. W. L. Jacobs, C. B. Hellquist, & J. H. Wiersema 8213 (NASC, NSW, BRI)	NY135	AM422065
Nymphaea subg. *Nymphaea*:				
N. alba L.	Germany, Bavaria, Luttensee	T. Borsch 3339 (BONN)	NY056	AM422066
N. alba L.	Finland, Nylandia, Porvoo	T. Borsch 3151 (BONN, H)	NY061	AM422067
N. candida C. Presl	Finland, Tavastia australis, Katloisteryärvi	T. Borsch 3154 (BONN, H)	NY062	AM422068
N. candida C. Presl	Finland, Tavastia australis, Maaranyärvi	T. Borsch 3152 (BONN, H)	NY063	AM422069
N. candida C. Presl	Russia, Siberia	C. B. Hellquist s.n. (MASS)	NY109	AM422070
N. mexicana Zucc.	USA, Florida	T. Borsch & B. Summers 3226 (BONN, VPI)	NY069	AM422071
N. mexicana Zucc.	USA, Texas	K. Woods & T. Borsch 701 (BONN, VPI)	KN008	AM422072
N. odorata Aiton subsp. *odorata*	USA, Georgia, Okefenokee Swamp	T. Borsch & V. Wilde 3132 (BONN, VPI)	NY012	AY145333[a]
N. odorata Aiton subsp. *tube rosa* (Paine) Wiersema & Hellq.	Canada, Manitoba	T. Borsch, J. H. Wiersema, & C. B. Hellquist 3389 (BONN, NASC)	NY269	AM422073
N. tetragona Georgi	Finland, *Tavastia australis*, Kanajärvi	T. Borsch 3155 (BONN, H)	NY060	AM422074

Note. For field origin, the countries and smaller units (states, etc.) are provided. In the case samples were taken from a documented wild-collected individual that is held in a living collection, the respective origin is provided in brackets. Vouchers are deposited in a number of herbaria (acronyms follow *Index Herbarium*), including TB (at BONN). Code numbers specify a DNA accession and should facilitate comparison, as the same DNAs are also used in other studies. GenBank numbers were obtained for sequences generated here except as noted.

[a] Sequences taken from Borsch et al. (2003).

Literature Cited

Barkman TJ, G Chenery, JR McNeal, J Lyons-Weiler, W Elisens, G Moore, AD Wolfe, CW De Pamphilis 2000 Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. Proc Natl Acad Sci USA 97:13166–13171.

Bayer RJ, JR Starr 1998 Tribal phylogeny of the Asteraceae based on two non-coding chloroplast sequences, the *trn*L intron and *trn*L/*trn*F intergenic spacer. Ann Mo Bot Gard 85:242–256.

Benson G 1997 Sequence alignment with tandem duplication. J Comput Biol 4:351–367.

Böhle UR, H Hilger, R Cerff, WF Martin 1994 Noncoding chloroplast DNA for plant molecular systematics at the infrageneric level. Pages 391–403 *in* B Schierwater, B Streit, GP Wagner, R de Salle, eds. Molecular ecology and evolution: approaches and applications. Birkhauser, Basel.

Bonilla-Barbosa JR 2001 Nymphaeaceae. Flora Guerrero 13:1–17.

Borsch T 2000 Phylogeny and evolution of the genus *Nymphaea* (Nymphaeaceae). PhD diss. Friedrich-Wilhelms Universität, Bonn.

Borsch T, KW Hilu, D Quandt, V Wilde, C Neinhuis, W Barthlott 2003 Non-coding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms. J Evol Biol 16: 558–576.

Borsch T, C Löhne, K Müller, KW 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 Leopold 92:85–110.

Cafasso D, G Pellegrino, A Musacchio, A Widmer, S Cozzolino 2001 Characterization of a minisatellite repeat locus in the chloroplast genome of *Orchis palustris*. Curr Genet 39:394–398.

Caspary R 1865 Nymphaeaceae. Ann Mus Bot Lugduno-Batavum 2: 241–253.

——— 1888 Nymphaeaceae. Pages 1–10 *in* A Engler, K Prantl, eds. Die natürlichen Pflanzenfamilien. Vol 3. Engelmann, Leipzig.

Cevalloz-Ferriz SRS, RA Stockey 1989 Permineralized fruits and seeds from the Princeton Chert (Middle Eocene) of British Columbia: Nymphaeaceae. Bot Gaz 150:207–217.

Chawanje CM, WE Barbeau, I Grun 2001 Nutrient and antinutrient content of an underexploited Malawian water tuber *Nymphaea petersiana*. Ecol Food Nutrit 40:347–366.

Chen I, SR Manchester, Z Chen 2004 Anatomically preserved seeds of *Nuphar* (Nymphaeaceae) from the early Eocene of Wutu, Shandong Province, China. Am J Bot 91:1265–1272.

Collinson ME 1980 Recent and Tertiary seeds of the Nymphaeaceae sensu lato with a revision of *Brasenia ovula* (Brong.) Reid and Chandler. Ann Bot 46:603–632.

Conard HS 1905 The waterlilies: a monograph of the genus *Nymphaea*. Publ Carnegie Inst Wash 4:1–279.

Cozzolino S, D Cafasso, G Pellegrino, A Musacchio, A Widmer 2003 Molecular evolution of a plastid tandem repeat locus in an orchid lineage. J Mol Evol 57:S41–S49.

De Candolle AP 1821 *Nymphaea*. Pages 49–59 *in* AP de Candolle, ed. Regni vegetabilis systema naturale. Vol 2. Treuttel & Würz, Paris.

Den Hartog C 1970 *Ondinea* a new genus of Nymphaeaceae. Blumea 18:413–417.

Di Rienzo A, AC Peterson, JC Garza, AM Valdes, M Slatkin, NB Freimer 1994 Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170.

Doran AS, DH Les, ML Moody, WE Phillips 2004 Nymphaea "William Phillips," a new intersubgeneric hybrid. Hortscience 39: 446–447.

Ervik F, JT Knudsen 2003 Water lilies and scarabs: faithful partners for 100 million years? Biol J Linn Soc 80:539–543.

Felsenstein J 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376.

Friis EM, KR Pedersen, PR Crane 2001 Fossil evidence of water lilies (Nymphaeales) in the Early Cretaceous. Nature 410:357–360.

Gabarayeva NI, G El-Ghazaly 1997 Sporoderm development in *Nymphaea mexicana* (Nymphaeaceae). Plant Syst Evol 204:1–19.

Gagnidze R 2005 Vascular plants of Georgia: a nomenclatural checklist. Georgian Academy of Sciences, Tiblisi.

Gandolfo MA, KC Nixon, WL Crepet 2004 Cretaceous flowers of Nymphaeaceae and implications for complex insect entrapment pollination mechanisms in early angiosperms. Proc Natl Acad Sci USA 101:8056–8060.

Geiger DL 2002 Stretch coding and block coding: two new strategies to represent questionable aligned DNA sequences. J Mol Evol 54: 191–199.

Gielly L, YM Yuan, P Küpfer, P Taberlet 1996 Phylogenetic use of noncoding regions in the genus *Gentiana* L.: chloroplast *trnL* (UAA) intron versus nuclear ribosomal internal transcribed spacer sequences. Mol Phylogenet Evol 5:460–466.

Gilg E 1908 Nymphaeaceae africanae. Bot Jahrb Syst 41:351–366.

Graham SW, RG Olmstead 2000 Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. Am J Bot 87: 1712–1730.

Graham SW, RG Olmstead, SCH Barrett 2002 Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. Mol Biol Evol 19:1769–1781.

Graham SW, PA Reeves, ACE Burns, RG Olmstead 2000 Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. Int J Plant Sci 161(suppl):S83–S96.

Griffiths E, AK Petrich, RS Gupta 2005 Conserved indels in essential proteins are distinctive characteristics of Chlamydiales and provide novel means for their identification. Microbiology 151:2647–2657.

Gu X, WH Li 1995 The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment. J Mol Evol 40:464–473.

Gupta PP 1978 Cytogenetics of aquatic ornamentals. II. Cytology of Nymphaeas. Cytologia 43:477–484.

——— 1980 Cytogenetics of aquatic ornamentals. VI. Evolutionary trends and relationships in the genus *Nymphaea*. Cytologia 45: 307–314.

Hamilton MB, JM Bravermann, DF 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. Mol Biol Evol 20:87–92.

Hesse M, R Zetter 2005 Ultrastructure and diversity of recent and fossil zona-aperturate pollen grains. Plant Syst Evol 255:145–176.

Hilu KW, T Borsch, K Müller, DS Soltis, PS Soltis, V Savolainen, MW Chase, et al 2003 Angiosperm phylogeny based on *matK* sequence information. Am J Bot 90:1758–1776.

Hirthe G, S Porembski 2003 Pollination of *Nymphaea lotus* (Nymphaeaceae) by rhinoceros beetles and bees in the northeastern ivory coast. Plant Biol 5:670–675.

Ito M 1987 Phylogenetic systematics of Nymphaeales. Bot Mag Tokyo 100:17–35.

Jacobs SWL 1992 New species, lectotypes, and synonyms of Australasian *Nymphaea* (Nymphaeaceae). Telopea 4:635–641.

Jacobs SWL, CL Porter Forthcoming Nymphaeaceae. *In* AS George, ed. Flora of Australia. Vol 2. Australian Biological Resources Study, Canberra.

Kadono Y, EL Schneider 1987 The life history of *Euryale ferox* Salisb. in southwestern Japan with special reference to reproductive ecology. Plant Species Biol 2:109–115.

Kanno A, A Hirai 1993 A transcription map of the chloroplast genome from rice (*Oryza sativa*). Curr Genet 23:166–174.

Kawakita A, T Sota, JS Ascher, M Ito, H Tanaka, M Kato 2003 Evolution and phylogenetic utility of alignment gaps within intron sequences of three nuclear genes in bumble bees (*Bombus*). Mol Biol Evol 20:87–92.

Kelchner SA 2000 The evolution of noncoding chloroplast DNA and its application in plant systematics. Ann Mo Bot Gard 87:482–498.

Kelchner SA, JF Wendel 1996 Hairpins create minute inversions in non-coding regions of chloroplast DNA. Curr Genet 30:259–262.

Kenneally KF, EL Schneider 1983 The genus *Ondinea* (Nymphaeaceae) including a new subspecies from the Kimberley Region, Western Australia. Nuytsia 4:359–365.

King RA, C Ferris 2002 A variable minisatellite sequence in the chloroplast genome of *Sorbus* L. (Rosaceae: Maloideae). Genome 45:570–576.

Kress WJ, KJ Wurdack, EA Zimmer, LA Weight, DH Janzen 2005 Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 102:8369–8374.

Langlet O, E Söderberg 1929 Über die Chromosomenzahlen einiger Nymphaeaceen. Acta Hortic Bergiani 9:85–104.

Leebens-Mack J, LA Raubeson, L Cui, JV Kuehl, MH Fourcade, TW Chumley, JL Boore, RK Jansen, CW dePamphilis 2005 Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein Zone. Mol Biol Evol 22: 1948–1963.

Les DH, DK Garvin, CF Wimpee 1991 Molecular evolutionary history of ancient aquatic angiosperms. Proc Natl Acad Sci USA 88: 10119–10123.

Les DH, ML Moody, AS Doran, WE Phillips 2004 A genetically confirmed intersubgeneric hybrid in *Nymphaea* L. (Nymphaeaceae Salisb.). Hortscience 39:219–222.

Les DH, EL Schneider, DJ Padgett, PS Soltis, DE 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. Syst Bot 24:28–46.

Levinson G, GA Gutmann 1987 Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221.

Lloyd DG, VL Calder 1991 Multi-residue gaps, a class of molecular characters with exceptional reliability for phylogenetic analyses. J Evol Biol 4:9–21.

Löhne C, T Borsch 2005 Phylogenetic utility and molecular evolution of the *petD* group II intron in basal angiosperms. Mol Biol Evol 22:317–332.

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

——— 2000 Basal angiosperm phylogeny inferred from duplicate phytochromes A and C. Int J Plant Sci 161(suppl):S41–S55.

Mendonça FA 1960 Nymphaeaceae. Flora Zambesiaca 1:175–180.

Mes THM, P Kuperus, J Kirschner, J Stepanek, P Oosterveld, H Storchova, JCM den Nijs 2000 Hairpins involving both inverted and direct repeats are associated with homoplasious indels in non-coding chloroplast DNA of *Taraxacum* (Lactuceae: Asteraceae). Genome 43:634–641.

Mes THM, H 't Hart 1994 *Sedum surculosum* and *S. jaccardianum* (Crassulaceae) share a unique 70 bp deletion in the chloroplast DNA *trn*L (UAA)/*trn*F (GAA) intergenic spacer. Plant Syst Evol 193:213–221.

Moseley MF 1961 Morphological studies of the Nymphaeaceae. II. The flowers of *Nymphaea*. Bot Gaz 122:233–259.

Moseley MF, EL Schneider, PS Williamson 1993 Phylogenetic interpretations from selected floral vasculature characters in the Nymphaeaceae sensu lato. Aquat Bot 44:325–342.

Müller K 2004 PRAP: computation of Bremer support for large data sets. Mol Phylogenet Evol 31:780–782.

——— 2005*a* The efficiency of different search strategies in estimating parsimony jackknife, bootstrap, and Bremer support. BMC Evol Biol 5:58.

——— 2005*b* Incorporating information from length mutational events into phylogenetic analysis. Mol Phylogenet Evol 38:667–676.

——— 2005*c* SeqState: primer design and sequence statistics for phylogenetic DNA data sets. Appl Bioinform 4:65–69.

Müller K, T Borsch 2005*a* Phylogenetics of Amaranthaceae based on matK/trnK sequence data: evidence from parsimony, likelihood, and Bayesian analyses. Ann Mo Bot Gard 92:66–102.

——— 2005*b* Phylogenetics of *Utricularia* (Lentibulariaceae) and molecular evolution of the *trnK* intron in a lineage with high substitutional rates. Plant Syst Evol 250:39–67.

Müller K, T Borsch, KW Hilu 2006 Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting *matK*, *trnT-F*, and *rbcL* in basal angiosperms. Mol Phylogenet Evol 41:99–117.

Nixon KC 1999 The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 15:407–414.

Okada H, M Tamura 1981 Karyomorphological study of the Nymphaeales. J Jpn Bot 56:367–374.

Planchon JE 1852 Enumeration succinctedes especes de la famille des Nymphéacées. Fl Ser Jard Eur 8:117–120.

——— 1853 Etudes sur les Nymphéacées. Ann Sci Nat Bot, ser 3, 19, 17–63.

Posada D, KA Crandall 1998 Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.

Prance GT, GR Arias 1975 A study of the floral biology of *Victoria amazonica* (Poepp.) Sowerby (Nymphaeaceae). Acta Amazonica 5: 109–129.

Provan J, P Wolters, KH Caldwell, W Powell 2004 High-resolution organellar genome analysis of *Triticum* and *Aegilops* sheds new light on cytoplasm evolution in wheat. Theor Appl Genet 108: 1182–1190.

Qiu Y-L, O Dombrovska, J Lee, L Li, BA Whitlock, F Bernasconi-Quadroni, JS Rest, et al 2005 Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. Int J Plant Sci 166:815–842.

Qiu Y-L, J Lee, F Bernasconi-Quadroni, DE Soltis, PS Soltis, M Zanis, EA Zimmer, Z Chen, V Savolainen, MW Chase 1999 The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature 402:404–407.

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 Biol 5:400–410.

Quandt D, K Müller, M Stech, JP Frahm, W Frey, KW Hilu, T Borsch 2004 Molecular evolution of the chloroplast trnL-F region in land plants. *In* B Goffinet, V Hollowell, R Magill, eds. Molecular systematics of Bryophytes. Monogr Syst Bot Mo Bot Gard 98: 13–37.

Quandt D, M Stech 2005 Molecular evolution of the *trnL* (UAA) intron in bryophytes. Mol Phylogenet Evol 36:429–443.

Renner SS 1999 Circumscription and phylogeny of the Laurales: evidence from molecular and morphological data. Am J Bot 86: 1301–1315.

Rodríguez F, JL Oliver, A Marin, JR Medina 1990 The general stochastic model of nucleotide substitution. J Theor Biol 142: 485–501.

Ronquist F, JP Huelsenbeck 2003 MrBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Sauquet H, JA Doyle, T Scharaschkin, T Borsch, KW Hilu, LW Chatrou, A Le Thomas 2003 Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple datasets: implications for character evolution. Bot J Linn Soc 142:125–186.

Schneider EL 1983 Gross morphology and floral biology of *Ondinea purpurea* den Hartog. Aust J Bot 31:371–382.

Schneider EL, SC Tucker, PS Williamson 2003 Floral development in the Nymphaeales. Int J Plant Sci 164(suppl):S279–S292.

Schneider EL, PS Williamson 1993 Nymphaeaceae. Pages 486–493 *in* K Kubitzki, ed. Families and genera of flowering plants. Vol 2. Springer, Berlin.

Simmons MP, KF Müller, AD Norton Forthcoming The relative performance of indel-coding methods in simulations. Mol Phylogenet Evol.

Simmons MP, H Ochoterena 2000 Gaps as characters in sequence-based phylogenetic analyses. Syst Biol 49:369–381.

Simmons MP, H Ochoterena, TG Carr 2001 Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. Syst Biol 50:454–462.

Soltis PS, DE Soltis, MW Chase 1999 Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature 402:402–404.

Swofford DL 2002 PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer, Sunderland, MA.

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

Tesfaye GK, K Govers, E Bekele, T Borsch Forthcoming Characterisation of *Coffea* chloroplast microsatellites and evidence for the recent divergence of *C. arabica* and *C. eugenioides* cp genomes. Genome.

van Ham RCHJ, H 't Hart, THM Mes, JM Sandbrink 1994 Molecular evolution of noncoding regions of the chloroplast genome in the Crassulaceae and related species. Curr Genet 25:558–566.

Verdcourt B 1989 Nymphaeaceae. Pages 1–12 *in* RM Polhill, ed. Flora of tropical East Africa. Balkema, Rotterdam.

Vogt L 2002 Weighting indels as phylogenetic markers of 18S rDNA sequences in Diptera and Strepsiptera. Org Divers Evol 2: 335–349.

Weberbauer A 1894 Beiträge zur Samenanatomie der Nymphaeaceae. Engler's Jahrb 18:213–374.

Weidlich WH 1976*a* The organization of the vascular system in the stems of Nymphaeaceae: *Nymphaea* subgenera *Anecphya*, *Lotos*, and *Brachyceras*. Am J Bot 63:1365–1379.

——— 1976*b* The organization of the vascular system in the stems of Nymphaeaceae: *Nymphaea* subgenera *Castalia* and *Hydrocallis*. Am J Bot 63:499–509.

Wiersema JH 1987 A monograph of *Nymphaea* subgen. *Hydrocallis* (Nymphaeaceae). Syst Bot Monogr 16:1–112.

——— 1988 Reproductive biology of *Nymphaea* (Nymphaeaceae). Ann Mo Bot Gard 75:795–804.

——— 1996 *Nymphaea tetragona* and *Nymphaea leibergii* (Nymphaeaceae): two species of diminutive water lilies in North America. Brittonia 48:520–531.

——— 2001 Nymphaeaceae: flora de Nicaragua. Syst Bot Mo Bot Gard 85:1592–1596.

——— 2003 Nymphaeaceae. Flora Venezuel Guayana 7:118–124.

Williamson PS, MF Moseley 1989 Morphological studies of the Nymphaeaceae sensu lato. XVII. Floral anatomy of *Ondinea purpurea* subsp. *purpurea* (Nymphaeaceae). Am J Bot 76:1779–1794.

Williamson PS, EL Schneider 1993 Cabombaceae. Pages 157–161 *in* K Kubitzki, ed. Families and genera of flowering plants. Vol 2. Springer, Berlin.

Won H, S S Renner 2005 The chloroplast *trnT-trnF* region in the seed plant lineage Gnetales. J Mol Evol 61:425–436.

Wood CE 1959 The genera of Nymphaeaceae and Ceratophyllaceae in the southeastern United States. J Arnold Arbor Harv Univ 40: 94–112.

Woods K, KW Hilu, T Borsch, JH Wiersema 2005*a* Pattern of variation and systematics of *Nymphaea odorata*. II. Sequence information from ITS and *trnL-trnF*. Syst Bot 30:481–493.

Woods K, KW Hilu, JH Wiersema, T Borsch 2005*b* Pattern of variation and systematics of *Nymphaea odorata*. I. Evidence from morphology and inter-simple sequence repeats (ISSRs). Syst Bot 30: 471–480.

Yoo M-J, CD Bell, PS Soltis, DE Soltis 2005 Divergence times and historical biogeography of Nymphaeales. Syst Bot 30:693–704.

Zanis M, DE Soltis, PS Soltis, S Mathews, MJ Donoghue 2002 The root of the angiosperms revisited. Proc Natl Acad Sci USA 99: 6848–6853.