Phylogenetic analysis of Nymphaeales using fast-evolving and noncoding chloroplast markers

CORNELIA LÖHNE1*, THOMAS BORSCH1 and JOHN H. WIERSEMA2

1Nees-Institut für Biodiversität der Pflanzen, Rheinische Friedrich-Wilhelms-Universität, 53115 Bonn, Germany
2Systematic Botany and Mycology Laboratory, Agricultural Research Service, United States Department of Agriculture, Beltsville, MD 20705-2350, USA

Received June 2006; accepted for publication March 2007

The Nymphaeales (water-lilies and relatives) represent one of the earliest branching lineages of angiosperms and comprise about 70 aquatic species. Here, we present a comprehensive study of phylogenetic relationships within the Nymphaeales from a dataset containing 24 representatives of the order, including all currently recognized genera and all subgenera of the genus Nymphaea, plus 5 outgroup taxa. Nine different regions of the chloroplast genome – comprising spacers, group II introns, a group I intron, and a protein coding gene – were analysed. This resulted in a character matrix of 6597 positions and an additional 369 characters obtained from coded length mutations. Maximum parsimony and Bayesian analyses of the complete dataset yielded congruent, fully resolved and well-supported trees. Our data confirm the monophyly of the Cabombaceae but do not provide convincing support for the monophyly of Nymphaeaceae with respect to Nuphar. Moreover, the genus Nymphaea is inferred to be paraphyletic with respect to Ondinea, Victoria and Euryale. In fact, the Australian endemic Ondinea forms a highly supported clade with members of the Australian Nymphaea subgenus Anecphya. In addition, Victoria and Euryale are inferred to be closely related to a clade comprising all night-blooming water-lilies (Nymphaea subgenera Hydrocallis and Lotos). An experimental approach showed taxon sampling to be of influence on the nodes reconstructed in core Nymphaeaceae. The results underscore that more diverse genera, if not clearly known to be monophyletic, should be represented by all major lineages. © 2007 The Linnean Society of London, Botanical Journal of the Linnean Society, 2007, 154, 141–163.

subgenus Hydrocallis, the palaeotropical subgenus Lotos, the pantropical subgenus Brachyceras, and the north-temperate subgenus Nymphaea.

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

Molecular and morphological data generally indicate a close relationship of Cabomba and Brasenia, thus substantiating the monophyly of the family Cabombaceae (see Williamson & Schneider, 1993). However, the monophyly of the family Nymphaeaceae does not gain much support in phylogenetic analyses. The rbcl study of Les et al. (1991) yielded a bootstrap support (BS) of 50 and the more recent analyses of trnT-trnF sequence data by Borsch et al. (2007) resulted in a jackknife support (Jk) of 76 and Bayesian posterior probability (PP) of 0.64, respectively. A recent analysis of basal angiosperm relationships (Löhne & Borsch, 2005) using petD group II intron sequences found three major clades within Nymphaeales, i.e. Cabombaceae, Nuphar and the remaining Nymphaeaceae. These changed their relative positions depending on the combination of data partitions and tree inference method used. Other studies of basal angiosperms show Nymphaeaceae as a monophylum but with low support (75% bootstrap support in Borsch et al., 2003), or do not even resolve generic relationships (rbcl + atpB +18S; Soltis et al., 2000). Analyses combining data from all three genomic compartments (Qiu et al., 1999; Zanis et al., 2002; Qiu et al., 2005) are usually short in taxon sampling, and comprise only one species each of Nymphaea, Nuphar, Cabomba and Brasenia. The heretofore most extensive phylogenetic analysis of Nymphaeales was carried out by Les et al. (1998), and was based on rbcl, matK, 18S rDNA, as well as a morphological matrix. It included one representative of each of the eight genera of Nymphaeales. However, Nymphaeaceae were rooted with Cabombaceae, thus assuming monophyly of Nymphaeaceae. Most recently, nuclear ITS sequences were employed to Nymphaeales phylogeny (Liu et al., 2005; Podoplelova & Ryzhakov, 2005) but could not clarify relationships, mainly due to limitations in taxon and character sampling.

Uncertainties also remain concerning the relationships within Nymphaeaceae. Based on the analysis by Les et al. (1999) relationships of Nymphaeaceae genera can be hypothesized as: Barclaya (Ondinea (Nymphaea (Euryale, Victoria))). However, the sister-group relationship of Ondinea to a clade consisting of Nymphaea and Victoria plus Euryale was not convincingly supported by the molecular data (bootstrap support = 51) in the above-mentioned study, with the node being based on three morphological characters only [Bremer support (BS) = 3]. A very recent molecular phylogenetic study involving noncoding trnT-trnF sequences from a much broader sampling of Nymphaea species (Borsch et al., 2007) indeed resolved Ondinea as part of a Nymphaea subgen. Anechpaya clade. Furthermore, Borsch et al. (2007) could not find convincing support for the monophyly of Nymphaea with respect to Victoria and Euryale (bootstrap = 69, PP = 0.86). The species of Nymphaea were consistently found in three well-supported clades: a clade consisting of N. subgenera Anechpaya and Brachyceras; another consisting of N. subgenera Hydrocallis and Lotos; and a third of N. subgen. Nymphaea that appeared sister to the other two. The trnT-trnF dataset of Borsch et al. (2007) further indicated that the pantropical Nymphaea subgenus Brachyceras might be paraphyletic to the Australian subgenus Anechpaya.

In view of these recent results demonstrating the apparent effects of taxon sampling in Nymphaea, and considering the unresolved questions about phylogenetic relationships in Nymphaeales, we designed a comprehensive study using a multigene dataset as well as a more representative taxon sampling covering all genera of Nymphaeales. In this study, each subgenus of Nymphaea is represented by at least two species that are most distant in the trnT-trnF tree of Borsch et al., (2007). We use several fast-evolving regions of the chloroplast genome to address the phylogeny of Nymphaeales. Our dataset comprises a wide spectrum of structurally different markers: a group I intron (trnL), group II introns (petD, rpl16, trnK), spacers (petB-petD, trnK-psbA, trnT-trnL, trnL-trnF), and the matK gene. The phylogenetic utility of these fast-evolving regions has been substantiated in numerous studies for the rpl16 intron (e.g. Kelchner & Clark, 1997; Renner & Chanderbali, 2000; Zhang, 2000), for the trnK-matK region (e.g. Steele & Vigalys,
1994; Johnson & Soltis, 1995; Hilu & Liang, 1997; Hilu et al., 2003; Müller & Borsch, 2005b), and for the trnT-trnF region (e.g. Taberlet et al., 1991; Asmussen & Chase, 2001; Borsch et al., 2003; Sauquet et al., 2003; Neinhuis et al., 2005). Compared to rather slowly evolving genomic regions of conserved genes, rapidly evolving regions provide higher percentages of variable and informative characters because of higher overall rates of substitutions. The study of Borsch et al. (2003) on trnT-trnF indicated that noncoding regions such as spacers and group I introns are alignable across major lineages of basal angiosperms, and at the same time provide resolution within these lineages. Based on these results Löhne & Borsch (2005) characterized the intron in petD as a representative of the structurally different group II introns, and demonstrated its high efficiency for phylogeny inference among basal angiosperms. The petD intron, the trnT-trnF region, and a set of additional, fast-evolving chloroplast genome regions seemed to be promising markers for a comprehensive combined analysis of Nymphaeales phylogeny. Moreover, it has been shown that microstructural mutations, which are frequent in noncoding DNA, provide valuable additional information for phylogenetic inference (Simmons, Ochoterena & Carr, 2001; Hamilton, Braverman & Soria-Hernanz, 2003; Löhne & Borsch, 2005; Müller, 2006). A focus on fast-evolving spacers and introns was therefore expected to yield a large set of microstructural characters as well.

For the present study there are three principal objectives:

1. Generate a well-resolved and sufficiently supported chloroplast phylogeny of Nymphaeales based on an extensive taxon and character sampling.
2. Test the monophyly of Nymphaeales as well as the monophyly of the genus Nymphaea, while meeting the first objective.
3. Evaluate the phylogenetic utility of the structurally different chloroplast genomic regions used in this study.

Ultimately, it will be necessary to compare well-founded phylogenetic hypotheses obtained from the individual genomes to address putative effects arising from ancient hybridization. Thus, selecting for genomic regions with maximum phylogenetic utility will be necessary to obtain well-supported trees for all three compartments.

MATERIAL AND METHODS

TAXON SAMPLING AND PLANT MATERIAL

The dataset used in this study comprises 24 species of Nymphaeales, representing both genera of the Cabombaceae (Brasenia, Cabomba), each genus of the Nymphaeaceae (Barclaya, Euryale, Nuphar, Nymphaea, Ondinea, Victoria), and within the genus Nymphaea each of the five subgenera (Anechoya, Brachycheras, Hydrocallis, Lotos, Nymphaea), as well as both sections of Nuphar (Astylus and Nuphar). Additionally, sequences of Amborella trichopoda (Amborellaceae) and four representatives of Austrobaileyales (Austrobaileya, Illicium, Kadsura, Schisandra) are included. Most of the sequence data were generated during this study, with only a few sequences, especially for trnT-trnF and petD, taken from previous studies (Borsch et al., 2003; Löhne & Borsch, 2005). All sequences of Nymphaea alba were taken from GenBank (AJ627251; Goremykin et al., 2004). Information on all investigated species, on the origin of material, as well as on deposited vouchers and GenBank accessions are summarized in Table 1.

DNA ISOLATION, AMPLIFICATION AND SEQUENCING

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

Using different sets of primers (see primer list in Appendix 1), we amplified four regions of the chloroplast genome: (1) the petD region, comprising the petB-petD spacer, the petD 5’ exon (only 8 bp) and the petD intron, (2) the rpl16 intron, (3) the trnK-matK region, comprising the complete trnK intron, the matK gene and the trnK-psbA spacer, and (4) the trnT-trnF region, comprising the trnT-trnL spacer, the trnL gene with its intron, and the trnL-trnF spacer. PCR was conducted on a T3 Thermocycler (Biometra, Göttingen, Germany), using Taq-DNA-polymerase, buffer and dNTPs from Peqlab (Erlangen, Germany). See Appendix 2 for a detailed description of PCR conditions and reaction mixes optimized for each genomic region. PCR products were purified using a QiaQuick gel extraction kit (QIAGEN Inc., Valencia/CA, USA) and sequenced either with an ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit v. 1.1 (Applied Biosystems, Foster City/CA, USA) on ABI 310 or 377 automated sequencers, or with a CEQ DTCS Quick Start Kit (Beckman Coulter, Fullerton/CA, USA) on a CEQ 8000 sequencer.

SEQUENCE ALIGNMENT AND INDEL CODING

Sequences were aligned manually following the rules described in Löhne & Borsch (2005), using BIOEDIT v.5.0.9 (Hall, 1999). For each of the genomic regions a
### Table 1. Taxa used in the study, their respective families, source of material, location of voucher specimens and GenBank accession numbers of deposited sequences. Most of the sequence data were generated for this study with only some sequences (marked *) of the *trnT-trnF* and the *petD* regions taken from our own previous studies (Borsch et al., 2003; Löhne & Borsch, 2005; Borsch et al., 2007). All sequences of *Nymphaea alba* were taken from GenBank (Goremykin et al., 2004).

<table>
<thead>
<tr>
<th>Species</th>
<th>Family</th>
<th>Origin</th>
<th>Voucher</th>
<th>GenBank accession numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><em>trnT-trnF</em></td>
</tr>
<tr>
<td><em>Amborella trichopoda</em> Baill.</td>
<td>AMB</td>
<td>University of California, Sta. Catarina Bot. Gard.</td>
<td>Borsch 3480 (VPI)</td>
<td>AY145324*</td>
</tr>
<tr>
<td><em>Austrobaileya scandens</em> C. White</td>
<td>AUS</td>
<td>Bonn Bot. Gard.</td>
<td>Borsch 3464 (BONN)</td>
<td>AY145326*</td>
</tr>
<tr>
<td><em>Cabomba caroliniana</em> Gray</td>
<td>CAB</td>
<td>USA, VA</td>
<td>Ludwig s.n. (VPI)</td>
<td>AY145328*</td>
</tr>
<tr>
<td><em>Cabomba sp.</em></td>
<td>CAB</td>
<td>Bought from a water gardening company</td>
<td>Löhne 59 (BONN)</td>
<td>AM489712</td>
</tr>
<tr>
<td><em>Brasenia schreberi</em> J.F.Gmel.</td>
<td>CAB</td>
<td>USA, VA</td>
<td>Borsch &amp; Wieboldt 3298 (VPI, FR)</td>
<td>AY145329*</td>
</tr>
<tr>
<td><em>Brasenia schreberi</em> J.F.Gmel.</td>
<td>CAB</td>
<td>Canada, Rice Lake [Saskatchewan]</td>
<td>Borsch, Wiersema &amp; Hellquist 3390 (BONN)</td>
<td>AM489713</td>
</tr>
<tr>
<td><em>Illicium floridanum</em> J. Ellis (N241)</td>
<td>ILL</td>
<td>Bonn Bot. Gard.</td>
<td>Borsch 3552 (BONN)</td>
<td>–</td>
</tr>
<tr>
<td>I. floridanum J. Ellis (N117)</td>
<td>ILL</td>
<td>USA, FL</td>
<td>Borsch &amp; Wilde 3104 (BONN)</td>
<td>AY145325*</td>
</tr>
<tr>
<td><em>Barclaya longifolia</em> Wall.</td>
<td>NYM</td>
<td>Bought from a water gardening company</td>
<td>Löhne 60 (BONN)</td>
<td>AM422019*</td>
</tr>
<tr>
<td><em>Euryale ferox</em> Salisb. (N379)</td>
<td>NYM</td>
<td>Bonn Bot. Gard. 14010</td>
<td>Borsch 3830 (BONN)</td>
<td>–</td>
</tr>
<tr>
<td>E. ferox Salisb. (N015)</td>
<td>NYM</td>
<td>Bonn Bot. Gard. 14010</td>
<td>Borsch 3830 (BONN)</td>
<td>AM422020*</td>
</tr>
<tr>
<td><em>Nuphar advena</em> (Aiton) W.T. Aiton</td>
<td>NYM</td>
<td>USA, FL</td>
<td>Borsch &amp; Wilde 3298 (FR)</td>
<td>AM422031*</td>
</tr>
<tr>
<td><em>Nuphar japonica</em> DC.</td>
<td>NYM</td>
<td>Bought from a water gardening company</td>
<td>Löhne 61 (BONN)</td>
<td>AM422022*</td>
</tr>
<tr>
<td><em>Nuphar lutea</em> (L.) Sibth &amp; Sm.</td>
<td>NYM</td>
<td>Germany, Hesse</td>
<td>Borsch 3337 (FR)</td>
<td>AY145330*</td>
</tr>
<tr>
<td><em>Nymphaea amazonum</em> Mart. &amp; Zucc.</td>
<td>NYM</td>
<td>Mexico, Oaxaca</td>
<td>Novelo, Wiersema, Hellquist &amp; Horn 1281 (MEXU)</td>
<td>AM422026*</td>
</tr>
<tr>
<td>Species</td>
<td>Family</td>
<td>Origin</td>
<td>Voucher</td>
<td>GenBank accession numbers</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>--------</td>
<td>-----------------------</td>
<td>---------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Nymphaea elkeniae</td>
<td>NYM</td>
<td>Australia, Queensland</td>
<td>Hellquist &amp; Leu 16757 (NASC, NSW, BRI)</td>
<td>AM489714 DQ185539 AM492175 AM421610</td>
</tr>
<tr>
<td>Nymphaea gracilis Zucc.</td>
<td>NYM</td>
<td>Mexico, Jalisco</td>
<td>Novelo, Wiersema, Hellquist &amp; Horn 1314 (MEXU)</td>
<td>AM422050* DQ185542 AM492177 AM421601</td>
</tr>
<tr>
<td>Nymphaea jamesoniana Planch.</td>
<td>NYM</td>
<td>USA, FL</td>
<td>Borsch &amp; Summers 3220 (FR, MO)</td>
<td>AM422032* DQ185544 AM492179 AM421603</td>
</tr>
<tr>
<td>Nymphaea lotus var. thermalis</td>
<td>NYM</td>
<td>Bonn Bot. Gard</td>
<td>Borsch 3832 (BONN)</td>
<td>AM422040* DQ185547 AM492182 AM421614</td>
</tr>
<tr>
<td>Nymphaea macroperma</td>
<td>NYM</td>
<td>Australia, Northern Territories (NSW)</td>
<td>Jacobs &amp; Hellquist 8796</td>
<td>AM489715 DQ185540 AM492176 AM421611</td>
</tr>
<tr>
<td>Nymphaea micrantha</td>
<td>NYM</td>
<td>Bonn Bot. Gard 5830</td>
<td>Koehnen s.n. (BONN)</td>
<td>AM422051* DQ185541 AY590874* AM421615</td>
</tr>
<tr>
<td>Nymphaea novogranatensis</td>
<td>NYM</td>
<td>Mexico, Oaxaca</td>
<td>Novelo &amp; Wiersema 1187 (MEXU)</td>
<td>AM422034* DQ185545 AM492180 AM421612</td>
</tr>
<tr>
<td>Nymphaea odorata Ait. ssp. tuberosa</td>
<td>NYM</td>
<td>Canada, Manitoba</td>
<td>Borsch, Hellquist &amp; Wiersema 3389 (BONN, NASC)</td>
<td>AM422073* DQ185549 AY590873* AM421605</td>
</tr>
<tr>
<td>Nymphaea oxypetala Planch.</td>
<td>NYM</td>
<td>Bolivia, Santa Cruz</td>
<td>Ritter, G. Crow, Garvius &amp; C. Crow 4491 (NHA)</td>
<td>AM422035* DQ185546 AM492181 AM421613</td>
</tr>
<tr>
<td>Nymphaea petersiana Klotzsch</td>
<td>NYM</td>
<td>Malawi</td>
<td>Chawanje s.n. (FR, BONN)</td>
<td>AM422053* DQ185548 AM492183 AM421604</td>
</tr>
<tr>
<td>Ondinea purpurea Hartog</td>
<td>NYM</td>
<td>Western Australia</td>
<td>Jacobs &amp; Hellquist 8853 (NSW)</td>
<td>AM422023* DQ185538 AM492174 AM421600</td>
</tr>
<tr>
<td>Victoria cruziana A.D.Orb.</td>
<td>NYM</td>
<td>Bonn Bot. Gard.</td>
<td>Löhne 55 (BONN)</td>
<td>AM422024* DQ185535 AY590870* AM421598</td>
</tr>
<tr>
<td>Victoria 'Longwood Hybrid'</td>
<td>NYM</td>
<td>Bonn Bot. Gard.</td>
<td>Borsch 3831 (BONN)</td>
<td>AM422025* DQ185536 AM492172 AM421609</td>
</tr>
<tr>
<td>Kadsura japonica (L.) Dun.</td>
<td>SCH</td>
<td>Bonn Bot. Gard.</td>
<td>Borsch 3411 (BONN)</td>
<td>AM489711 DQ185525 AM492167 AM421616</td>
</tr>
<tr>
<td>Schisandra chinensis (Turcz.)</td>
<td>SCH</td>
<td>Bonn Bot. Gard.</td>
<td>Borsch &amp; Löhne 3492 (BONN)</td>
<td>– DQ185526 AY590866* AM421608</td>
</tr>
<tr>
<td>Schisandra rubriflora Rehder &amp; E. H. Wilson</td>
<td>SCH</td>
<td>Bonn Bot. Gard. 0727 ex BG Munich</td>
<td>Borsch 3477 (BONN)</td>
<td>AY145327* – – –</td>
</tr>
</tbody>
</table>

AMB, Amborellaceae; AUS, Austrobaileyaceae; CAB, Cabombaceae; ILL, Illiciaceae; NYM, Nymphaeaceae; SCH, Schisandraceae;
separate alignment was initially produced. Mutational hotspots, i.e. regions of uncertain homology, were excluded from analysis. Also, all exon parts flanking the petD, rpl16, trnK and trnL introns were excluded from analysis as they are short, and because they were often incomplete owing to the placement of sequencing primers. After alignment, gaps were coded automatically in a binary matrix using SEQSTATE v.1.21 (Müller, 2005). The 'simple indel coding' strategy based on Simmons & Ochoterena (2000) was applied. Alignments of all genomic regions, plus the respective indel matrices, were then combined to a single nexus file comprising several data partitions. Alignments and nexus data files are available from the Corresponding author upon request.

PHYLOGENETIC ANALYSES

Prior to phylogenetic analyses sequence divergence (uncorrected p distance), G/C content and other sequence statistics were calculated using SEQSTATE (Müller, 2005). In order to compare similarity of phylogenetic signal between genomic regions, partition homogeneity tests were performed in 1000 replicates using PAUP* v.4.0b10 (Swofford, 2002). For phylogeny reconstruction the following data partitions were analysed and compared: (1) all genomic regions separately, (2) all spacers together (trnT-trnL, trnL-trnF, petB-petD, trnK-psbA), (3) all introns (trnL, trnK, petD, rpl16), (4) all group II introns (trnK, petD, rpl16), (5) coding regions (only matK), (6) noncoding regions (all introns and spacers), (7) all indel characters, and (8) a combined dataset including all nucleotide and indel characters.

Maximum parsimony

Parsimony analyses (MP) were conducted with PAUP* v.4.0b10 (Swofford, 2002) employing heuristic searches with 1000 random addition replicates and TBR branch swapping. The limit of trees saved was set to 10 000 for small matrices. Branch support was estimated through 10 000 jackknife (JK) replicates (simple addition, keeping one tree per replicate, deleting 36.788% of characters in each replicate). In addition, Bremer support (BS) was calculated using PAUP* with the help of PRAF v.1.21 (ten random addition replicates per constraint tree, parsimony ratchet not employed; Müller, 2004). The consistency index (CI; Kluge & Farris, 1969) and the rescaled consistency index (RC; Farris, 1989) were calculated to assess levels of homoplasy. For a quantitative assignment of support or disagreement to each of the data partitions, the partitioned Bremer (PBS) support was estimated using the program TREEROT v.2 (Sorenson, 1999). Additionally, the phylogenetic structure R (Quandt, Müller & Hutten, 2003; Müller, Borsch & Hilu, 2006) of each data partition was calculated in order to evaluate the phylogenetic utility of the respective markers. As a measure of average support per node R can range from 0 to 1, i.e. $R = 1$ if all nodes gain maximum support and $R = 0$ if not a single node is supported by more than 50% of jackknife or bootstrap replicates.

Bayesian inference

Besides parsimony analyses, Bayesian inference (BI) of phylogeny for the combined dataset was conducted using MRBAYES v.3.1 (Ronquist & Huelsenbeck, 2003). Following the Akaike information criterion, MODELTEST v.3.06 (Posada & Crandall, 1998) assigned the GTR + G model of molecular evolution to all nucleotide partitions, except the petB-petD spacer and the trnT-trnL spacer, to which the GTR model was assigned. First, an analysis was conducted with all nucleotide characters. Then, all characters obtained from indel coding were appended to the matrix, and the binary (restriction site) model implemented in MRBAYES was applied to this partition. Both analyses were performed for 1 000 000 generations applying the default settings (MCMCMC, four runs with four chains each, heating temperature 0.2, saving one tree every 100 generations). In all runs, the probabilities had converged to a stable value after 15 000 generations. Thus, a consensus was calculated from a total of 39 400 trees sampled after the burn-in.

ROOTING OF THE TREES

The Nymphaeales clade is most probably the second branch of the angiosperm tree, diverging after Amborella, and followed by Austrobaileya (Parkinson, Adams & Palmer, 1999; Mathews & Donoghue, 2000; Soltis et al., 2000; Borsch et al., 2003; Hilu et al., 2003; Borsch et al., 2005; Qiu et al., 2005). Alternative hypotheses on relationships of extant angiosperms assume an Amborella-Nymphaeales clade as first branch (Barkman et al., 2000; Zanis et al., 2002; Stefanovic, Rice & Palmer, 2004; Leebens-Mack et al., 2005). Rather than using only Amborella (which actually might possess a large number of derived states) as a single outgroup taxon, representatives of Austrobaileya were also included to better reflect possible variation in character states present in the next relatives of Nymphaeales. More distant outgroups such as representatives of Magnoliids or gymnosperms were not included as this could have increased the chance for effects of long branch attraction (Graham, Olmstead & Barrett, 2002). However, a comparison of alternative rootings for the Nymphaeales did not show significant influence on hypotheses inferred for the ingroup (Borsch et al., 2007).
Evaluating alternative topologies

In order to compare the likelihood of the topology inferred from our dataset with alternative phylogenetic hypotheses Kishino-Hasegawa (KH) tests (Kishino & Hasegawa, 1989) were performed with the combined matrix (all substitutions) using the *Lscores* option in PAUP* and GTR + G model settings obtained with MODELTEST v.3.06 (Posada & Crandall, 1998). The evaluated alternative hypotheses, as illustrated in Figure 1, refer to the position of *Nuphar* as either sister to the remaining Nymphaeales (A) or sister to Cabombaceae (B; see Löhne & Borsch, 2005), and to the monophyly of the genus *Nymphaea* with *Victoria-Euryale* as a sister clade (C; Borsch et al., in press).

Analyses with reduced taxon sets

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

**RESULTS**

**SEQUENCE VARIABILITY OBSERVED IN THE DATASETS**

An average of 6630 nucleotides (nt) were sequenced per taxon to obtain the multilocus chloroplast dataset for Nymphaeales. About 660 nt per taxon were not analysed because they were short fragments of sequence adjacent to the genomic regions under study or from short tRNA exons. An average of 527 nt per taxon fell within mutational hotspot regions and was therefore excluded, too. The final matrix containing all data partitions (spacers, introns and *matK*) comprised 6597 characters plus a presence–absence matrix of 369 indels, providing total numbers of 2224 variable and 1356 parsimony informative characters. A total of 19 hotspots was determined, occurring in all genomic regions except the *matK* gene and the *petD* intron. A small hotspot was found in the *petB-petD* spacer (1–12 nt). The other spacers contained two to four hotspots (three in *trnK-psbA* with 147–257 nt in total; four in *trnT-trnL* with 46–245 nt; two in *trnL-trnF* with 29–84 nt). Four hotspots were also excluded from the *trnK* intron (74–135 nt in total) and from the *rpl16* intron (36–412 nt in total), respectively. The high length variability of these hotspots is caused mainly by long sequence stretches that are only present in the outgroup, especially in *Amborella* and *Ambroaileyae*.

**Figure 1.** Alternative hypotheses for relationships within Nymphaeales, tested against the optimal tree obtained from the present dataset in Kishino-Hasegawa tests. A, B, topologies testing different positions of *Nuphar*. C, tree referring to the monophyly of the genus *Nymphaea* with respect to *Victoria-Euryale*. 
(data not shown). Sequences of Nymphaeales taxa are generally shorter. An exception is the single hotspot in the trnL intron (7–140 nt), which is located in the P8 domain and contains a highly length-variable AT-microsatellite only present in the ingroup.

The rpl16 intron is the region with highest length variability (mean sequence length = 825.7 nt, SD = 84.6; Table 2). The matK gene, on the other hand, is most conserved in sequence length as is evident by a low SD of 10.6 although matK has twice the length of the rpl16 intron (1517.8 nt; Table 2). The number of indels coded in the matK gene (N = 19) is low compared to its length (1.3 indels per 100 nt; see last column in Table 2). More indels per character were coded in the introns (6.5 indels per 100 nt) and, especially, in the spacers (11.0 indels per 100 nt; Table 2). The GC-content is higher in the introns (38.6%) than in the spacers (36.3%) and the matK gene (36.1%). The GC-content of the petB-petD spacer (30.0%) is rather low compared to other partitions of this dataset (Table 2, column 5). Sequence divergence, i.e. average pairwise sequence distance (Table 2, column 6), is lowest in the petD region (spacer and intron), highest in the rpl16 intron and the trnT-trnL and trnL-trnF spacers. The petD region also provides the lowest amounts of variable and informative characters relative to the total number of characters (20.5% and 8.4%, respectively; Table 2, columns 7 and 8). The highest percentages of parsimony informative characters in the present dataset can be found in the matK gene (31.1% variable, 20.6% informative).

**Table 2.** Comparison of sequence statistics between all chloroplast genomic regions analysed individually (Part A) and in different combinations (Part B) for the 29 taxon dataset of Nymphaeales

<table>
<thead>
<tr>
<th>Data partition</th>
<th>Char.</th>
<th>Length range</th>
<th>Mean length</th>
<th>% divergence</th>
<th>% GC</th>
<th>% Var.</th>
<th>% Inform.</th>
<th>No. indels</th>
<th>% Inf. indels</th>
<th>Indel freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Individual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>petB-petD spacer</td>
<td>273</td>
<td>164–224</td>
<td>196.8 (13.2)</td>
<td>4.5 (0.0–15.4)</td>
<td>30.0</td>
<td>20.5</td>
<td>8.4</td>
<td>27</td>
<td>63.0</td>
<td>13.7</td>
</tr>
<tr>
<td>petD intron</td>
<td>817</td>
<td>621–733</td>
<td>652.7 (33.3)</td>
<td>4.6 (0.0–14.0)</td>
<td>39.5</td>
<td>22.4</td>
<td>12.2</td>
<td>42</td>
<td>54.8</td>
<td>6.4</td>
</tr>
<tr>
<td>rpl16 intron</td>
<td>858</td>
<td>763–1109</td>
<td>825.7 (84.6)</td>
<td>8.0 (0.0–21.3)</td>
<td>37.8</td>
<td>30.3</td>
<td>20.0</td>
<td>49</td>
<td>59.2</td>
<td>5.9</td>
</tr>
<tr>
<td>trnK intron</td>
<td>1130</td>
<td>957–1068</td>
<td>1007.9 (21.8)</td>
<td>6.6 (0.0–19.1)</td>
<td>38.1</td>
<td>28.2</td>
<td>18.1</td>
<td>66</td>
<td>57.6</td>
<td>6.5</td>
</tr>
<tr>
<td>matK</td>
<td>1590</td>
<td>1499–1542</td>
<td>1517.8 (10.6)</td>
<td>6.5 (0.0–19.0)</td>
<td>36.1</td>
<td>31.3</td>
<td>20.6</td>
<td>19</td>
<td>57.9</td>
<td>1.3</td>
</tr>
<tr>
<td>trnK-psbA spacer</td>
<td>170</td>
<td>287–396</td>
<td>316.6 (17.0)</td>
<td>6.6 (0.0–24.0)</td>
<td>39.4</td>
<td>30.6</td>
<td>15.9</td>
<td>10</td>
<td>50.0</td>
<td>3.2</td>
</tr>
<tr>
<td>trnT-trnL spacer</td>
<td>600</td>
<td>349–684</td>
<td>483.8 (54.4)</td>
<td>8.8 (0.0–25.2)</td>
<td>38.8</td>
<td>31.5</td>
<td>20.3</td>
<td>59</td>
<td>49.2</td>
<td>12.2</td>
</tr>
<tr>
<td>trnL intron</td>
<td>581</td>
<td>474–606</td>
<td>537.9 (32.8)</td>
<td>5.6 (0.0–14.3)</td>
<td>39.5</td>
<td>23.1</td>
<td>13.8</td>
<td>41</td>
<td>53.7</td>
<td>7.6</td>
</tr>
<tr>
<td>trnL-trnF spacer</td>
<td>579</td>
<td>244–441</td>
<td>385.2 (33.0)</td>
<td>9.5 (0.0–28.2)</td>
<td>35.1</td>
<td>28.0</td>
<td>17.3</td>
<td>56</td>
<td>46.4</td>
<td>14.5</td>
</tr>
<tr>
<td>(B) Combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introns</td>
<td>3386</td>
<td>2916–3384</td>
<td>3024.1 (98.4)</td>
<td>6.3 (0.0–17.6)</td>
<td>38.6</td>
<td>26.5</td>
<td>16.5</td>
<td>198</td>
<td>56.6</td>
<td>6.5</td>
</tr>
<tr>
<td>Spacers</td>
<td>1621</td>
<td>1225–1569</td>
<td>1382.4 (52.5)</td>
<td>7.9 (0.0–23.3)</td>
<td>36.3</td>
<td>28.3</td>
<td>16.8</td>
<td>152</td>
<td>50.7</td>
<td>11.0</td>
</tr>
<tr>
<td>Noncoding</td>
<td>5007</td>
<td>4268–4852</td>
<td>4406.5 (121.1)</td>
<td>6.7 (0.0–19.1)</td>
<td>37.9</td>
<td>27.1</td>
<td>16.6</td>
<td>350</td>
<td>54.0</td>
<td>7.9</td>
</tr>
<tr>
<td>All substitutions</td>
<td>6597</td>
<td>5767–6356</td>
<td>5924.4 (119.7)</td>
<td>6.7 (0.0–19.1)</td>
<td>37.4</td>
<td>28.1</td>
<td>17.5</td>
<td>369</td>
<td>54.2</td>
<td>6.3</td>
</tr>
<tr>
<td>Total</td>
<td>6966</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>31.9</td>
<td>19.5</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Char., number of characters in the alignment matrix (excluding hotspots); Length range, actual sequence length in nucleotides (minimal and maximal observed value, including hotspots); Mean length, mean of all observed sequence lengths (standard deviation in brackets); % divergence, pairwise sequence distance in percent (uncorrected p distance, overall mean, lowest and highest scores in brackets); % GC, GC content in percent; % Var., percentage of variable characters; % Inform., percentage of parsimony informative characters; No. indels, number of length mutations that were coded by SeqState; % Inf. indels, percentage of parsimony informative indels; Indel freq., indel frequency (number of indels per 100 nucleotides, calculated from the mean sequence length of each partition).

respectively. Consensus trees of individual datasets differed in degree of resolution, ranging from 8 nodes resolved by the petB-petD and trnK-psbA spacers to 24 nodes resolved by the trnK intron (Table 3, column 7). In parallel, $R$ is lowest in the petB-petD and trnK-psbA spacers ($R = 0.18$) and highest in the trnK intron ($R = 0.59$; Table 3, column 6). Partitioned Bremer support (PBS) for selected nodes and summed PBS values for all data partitions are presented in Table 4. Six nodes (nodes 5, 8, 14, 22, plus 20 and 23 which are not shown in the table) receive PBS values below zero from one of the data partitions (or two in the case of node 14), indicating phylogenetic signal that contradicts the total evidence tree. However, because values do not drop below $-1$ in these few cases, the overall conflict is rather small.

**TREES OBTAINED FROM BAYESIAN INFERENCE**

The combined dataset comprising all markers was used for Bayesian inference. Inclusion or exclusion of information from length mutations (369 indel characters) had no effect on tree topology and only minor effects in posterior probability values of some nodes. Figure 3 shows the Bayesian consensus of the calculation including indel information, with branch lengths estimated from 39 400 trees. Posterior probabilities (PP) of clades are given above branches. Only

---

**Figure 2.** Single most parsimonious tree of Nymphaeales obtained from a combined analysis of all markers (substitutions and indel matrix). Jackknife values are given above branches, BS below. Nodes are numbered consecutively (in circles).
### Table 3. Results of maximum parsimony analyses and phylogenetic structure (R) of different data partitions

<table>
<thead>
<tr>
<th>Data partition</th>
<th>Trees</th>
<th>TL</th>
<th>CI</th>
<th>RC</th>
<th>R</th>
<th>No. of nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>(A) Individual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>petB-petD spacer</td>
<td>42</td>
<td>69</td>
<td>0.94</td>
<td>0.89</td>
<td>0.18</td>
<td>8</td>
</tr>
<tr>
<td>petD intron</td>
<td>7</td>
<td>231</td>
<td>0.88</td>
<td>0.81</td>
<td>0.40</td>
<td>13</td>
</tr>
<tr>
<td>rpl16 intron</td>
<td>2</td>
<td>387</td>
<td>0.81</td>
<td>0.71</td>
<td>0.58</td>
<td>17</td>
</tr>
<tr>
<td>trnK intron</td>
<td>5</td>
<td>464</td>
<td>0.84</td>
<td>0.73</td>
<td>0.59</td>
<td>24</td>
</tr>
<tr>
<td>matK</td>
<td>15</td>
<td>687</td>
<td>0.84</td>
<td>0.75</td>
<td>0.58</td>
<td>18</td>
</tr>
<tr>
<td>trnT-trnL spacer</td>
<td>40</td>
<td>250</td>
<td>0.87</td>
<td>0.80</td>
<td>0.53</td>
<td>18</td>
</tr>
<tr>
<td>trnL intron</td>
<td>26</td>
<td>167</td>
<td>0.89</td>
<td>0.88</td>
<td>0.39</td>
<td>13</td>
</tr>
<tr>
<td>trnL-trnF spacer</td>
<td>540</td>
<td>218</td>
<td>0.85</td>
<td>0.76</td>
<td>0.42</td>
<td>16</td>
</tr>
<tr>
<td>(B) Combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introns</td>
<td>2</td>
<td>1255</td>
<td>0.84</td>
<td>0.75</td>
<td>0.82</td>
<td>25</td>
</tr>
<tr>
<td>Spacers</td>
<td>99</td>
<td>618</td>
<td>0.87</td>
<td>0.78</td>
<td>0.65</td>
<td>20</td>
</tr>
<tr>
<td>Noncoding</td>
<td>1</td>
<td>1874</td>
<td>0.85</td>
<td>0.76</td>
<td>0.88</td>
<td>26</td>
</tr>
<tr>
<td>All substitutions</td>
<td>1</td>
<td>2562</td>
<td>0.85</td>
<td>0.76</td>
<td>0.90</td>
<td>26</td>
</tr>
<tr>
<td>Indels</td>
<td>12</td>
<td>414</td>
<td>0.89</td>
<td>0.83</td>
<td>0.54</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>2979</td>
<td>0.85</td>
<td>0.77</td>
<td>0.90</td>
<td>26</td>
</tr>
</tbody>
</table>

*3/4 node numbers refer to nodes in Figure 1.

Trees, number of trees saved; TL, length of shortest trees; CI, consistency index; RC, rescaled consistency index; R, phylogenetic structure; No. of nodes, number of nodes present in the strict consensus; n.p., node not present in the respective tree;
three nodes (5, 14 and 19) gain different PP values depending on the inclusion or exclusion of indels (indicated by the numbers in brackets in Fig. 3), but their probabilities are rather low in both trees.

**Figure 3.** Phylogram of the combined dataset of Nymphaeales including all markers (substitutions and indels). The tree is a 50% majority rule consensus of 39 400 trees obtained from four runs of Bayesian analysis implementing the GTR + G model. Branch lengths reflect changes per site. Posterior probabilities are given above branches.

**PHYLOGENY OF NYMPHAEALES**

Partition homogeneity tests, the results of individual and combined analyses of the markers, as well as the PBS analyses, show that there is effectively no conflict
well-supported clade (JK consisting of the following.  

In both total evidence trees the Cabombaceae, consisting of Cabomba and Brasenia, are revealed as a well-supported clade (JK = 100, BS = 82, PP = 1.0). Nymphaeaceae, consisting of Nuphar, Barclaya, Nymphaea, Ondinea, Euryale and Victoria, also emerge as monophyletic with high jackknife (JK = 99) but rather low Bremer support (BS = 8) and medium posterior probability (PP = 0.96). Nuphar is depicted as sister to the remaining Nymphaeaceae, followed by Barclaya, Victoria and Euryale form a well-supported lineage. However, the genus Nymphaea appears as paraphyletic with respect to both Ondinea and the Victoria–Euryale clade. Ondinea is placed within Nymphaea subgen. Anechpahy with maximum support in both MP and Bayesian analysis. Nymphaea subgenera Anechpaha and Brachyceras form a well-supported clade (JK = 100, BS = 31, PP = 1.0), whereas N. subgen. Hydrocallis is grouped with N. subgen. Lotos (JK = 100, BS = 8, PP = 1.0). Victoria and Euryale emerge as a sister group to the Hydrocallis–Lotos clade with average jackknife (JK = 88), but low Bremer support (BS = 1) and low posterior probability (PP = 0.52). The temperate subgenus Nymphaea appears sister (JK = 90, BS = 2, PP = 0.91) to all other subgenera as well as Victoria–Euryale and Ondinea. Whereas the position of Ondinea nested within the Anechpahy–Brachyceras clade is supported by all individual partitions except the trnK-psbA spacer and the trnL intron (node 15, Table 3, Fig. 2; note that the P8 stem-loop of the trnL intron will resolve Ondinea but this is excluded as hotspot here), the position of Euryale–Victoria nested within the genus Nymphaea is revealed only by the rpl16 intron (node 14) and the trnK intron (node 19).

In order to evaluate alternative hypotheses (illustrated in Fig. 1) on phylogenetic relationships in Nymphaeales, Kishino-Hasegawa tests were performed. The total evidence tree obtained from the present dataset (Figs 2, 3) was significantly favoured over the alternatives (P = 0.000).

**Effects of Taxon Sampling**

In each of the MP analyses of the five taxon subsets a single most parsimonious tree was obtained. Selecting different species of Nymphaea to represent the genus had significant effects on the trees inferred for Nymphaeales. The trees differ in topology, and in particular the position of Ondinea was dependent upon the respective species of Nymphaea included (Fig. 4A–E). If N. elleniae (subgen. Anechpaya) or N. micrantha (subgen. Brachyceras) were chosen, Ondinea grouped with them with maximum support (JK = 100, Fig. 4A,B). If N. amazonum (subgen. Hydrocallis) or N. lotus (subgen. Lotos) were chosen as representatives of Nymphaea (Fig. 4C,D), Ondinea diverged below the split between Victoria–Euryale and Nymphaea, with high support (JK = 96) in the first case and medium support (JK = 87) in the latter. Nymphaea odorata (subgen. Nymphaea), on the other hand, emerges sister to a clade consisting of Ondinea and Victoria–Euryale (JK = 74, Fig. 4E) when sampled alone.

**Discussion**

**Structure of the Data and Reliability of the Trees**

This study represents the most extensive molecular dataset for Nymphaeales compiled to date. Due to a...
relatively high percentage of variable and informative characters, that is comparable to other datasets based on noncoding chloroplast DNA (e.g. Renner & Chandersbali, 2000; Borsch et al., 2003; Löhne & Borsch, 2005; Neinhuis et al., 2005), a total of 1156 parsimony informative alignment positions plus 200 informative characters from coded length mutations was obtained. Also, similar to previous analyses based on noncoding genomic regions, the occurrence of extremely variable nucleotide stretches, which impair a clear alignment,

Figure 4. Result of phylogenetic analyses with taxon sampling reduced from 29 to 8 taxa, thus reflecting the number of taxa sampled by Les et al. (1999). Five independent parsimony analyses were run to test whether members of different subgenera of *Nymphaea* yield different trees (A, subgenus *Anecphya*; B, subgenus *Brachyceras*; C, subgenus *Hydrocallis*; D, subgenus *Lotos*; E, subgenus *Nymphaea*). *Cabomba* and *Brasenia* were used as outgroup taxa. Jackknife values are given above branches, BS below. Branch lengths reflect number of changes.
was confined to several, clearly demarcated mutational hotspots comprising less than 10% of all nucleotides (Borsch et al., 2005).

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

The fact that none of the data partitions alone provided enough information to fully resolve the relationships among the sampled taxa, but that combining markers resulted in a single most parsimonious tree and strongly increased the number of well-supported clades, demonstrates the necessity of a certain amount of informative characters to clarify phylogenetic relationships among genera of Nymphaeales. The striking advance in resolution and support relative to the trnT-trnF tree of Borsch et al. (2007) underscores the importance of identifying and selecting highly performing molecular markers. Our study shows that combining information from genomic regions that are particularly suitable for this taxonomic spectrum, i.e. noncoding and fast-evolving DNA (see below), is the most powerful strategy to obtain a completely resolved phylogeny of Nymphaeales. The inclusion of further fast-evolving markers might therefore be helpful to verify those relationships that do not gain sufficient support by the present dataset.

Microstructural mutations
The indel matrix compiled from microstructural mutations in the different genomic regions provides phylogenetic information that is congruent with the topology obtained from substitutions. Clades with high statistical support from substitutions and with long branches are characterized by high numbers of apomorphic indels, such as the Cabombaceae (7 synapomorphic indel characters) and the genus Cabomba (25 indels), or the clade consisting of Nymphaea subgenera Brachyceras and Anechya, and Ondinea (7 indels). Our indel matrix is one of the largest that has been compiled so far (see Simmons et al., 2001). The fact that most clades in Nymphaeales are resolved even from indels alone (the only seven nodes not resolved are 8, 14, 17, 18, 19, 23, 26), demonstrates the high value of length mutations in phylogenetic analyses. We agree with Simmons & Ochoterena (2000) and other authors (e.g. Giribert & Wheeler, 1999; Kelchner, 2000; Müller, 2006), that not coding length mutations is equivalent to discarding data. In this study the benefit from including indels is once again shown.

Branch lengths
Conspicuous differences in branch lengths can be observed among internal and terminal branches of the Bayesian phylogram (Fig. 3). Such differences might be caused by changes in diversification rates or by shifts in the rates of molecular evolution within a lineage. A closer look at Figure 3 reveals that all internal nodes with rather short branches do not gain maximum support in Parsimony or Bayesian analyses. This is true for node 8 (uniting all members of Nymphaeaceae), node 14 (uniting the Anechya-Brachyceras and the Hydrocallis–Lotos–Victoria–Euryale clades), node 16 (N. subgen. Brachyceras), node 19 (uniting Hydrocallis–Lotos with Victoria–Euryale), and node 26 (N. amazonum sister to N. novogranatensis). Short branch lengths in both the Bayesian (Fig. 3) and the MP tree (branch length not shown), in combination with low BS values (Fig. 2), indicate the presence of only few character states that changed on the branch to the respective reconstructed ancestor under a given most parsimonious or most probable scenario. Such low numbers of changes on a branch may have resulted from rather rapid diversification and, as a consequence, the short time that was available to accumulate mutations (Wendel & Doyle, 1998; Bremer et al., 1999; Cronn et al., 2002). Besides this, as Löhne & Borsch (2005) showed for the petD intron, short branches might be strongly influenced by homoplastic characters. Long branches subtending terminal taxa, on the other hand, often hint at accelerated rates of molecular evolution. The most striking examples in the present dataset are the long branches leading to Cabomba, and to Barclaya, respectively. Cabomba is sister to Brasenia, which is subtended by a much shorter branch. Thus, the rate of molecular evolution in Cabomba relative to Brasenia must be higher. Rate differences are therefore present within Nymphae-
ales. Barclaya, however, is a rather isolated lineage within Nymphaeaceae and is represented by only one sequence in our dataset. Nevertheless, it is remarkable that Cabomba and Barclaya represent two extremes of life forms in the Nymphaeales. Whereas Cabomba is a completely submersed plant with strongly dissected leaves, one species of Barclaya (B. rotundifolia Hotta) is semiterrestrial. This may indicate that strong shifts in morphological features or life form are related to accelerated molecular evolution (Müller et al., 2004).

**Phylogeny and evolution of Nymphaeales**

The results of our study confirm several previous hypotheses on phylogenetic relationships in the order Nymphaeales; owing to comprehensive taxon and character sampling they also provide new insights into the evolutionary history of this group. Our results corroborate the monophyly of Cabombaceae, which has been convincingly stated before based on morphological, anatomical and molecular characters (Williamson & Schneider, 1993). Within Nymphaeaceae, the study confirms the position of Barclaya as sister to a well-supported clade consisting of Nymphaea, Ondinea, Victoria and Euryale (similar to Les et al., 1999). Furthermore, it provides additional evidence for the sister relationship of Victoria and Euryale, which had been proposed as early as 1891 by Caspary (based on seed morphology and the presence of spines), and was supported by phylogenetic analyses of sequence data a century later (Les et al., 1991; Les et al., 1999; Borsch et al., 2007). Nuphar is inferred as first-branching in a clade comprising all members of Nymphaeaceae, similar to previous studies based on anatomy, morphology and molecules (Ito, 1987; Les et al., 1999). However, despite the high amount of characters sampled in this study the monophyly of the Nymphaeaceae is still not convincingly supported (see below). More strikingly and in contrast to all previous phylogenetic studies and classifications, this study infers the genus Nymphaea to be paraphyletic with respect to the Victoria–Euryale clade and to Ondinea. Because of their relevance for understanding the evolution of the water-lily family, we want to address these findings in closer detail.

**Nuphar and the monophyly of Nymphaeaceae**

Nymphaeaceae are revealed as a monophyletic group in the present analysis with Nuphar branching first in this clade (node 8). Alternative topologies, i.e. Nuphar basal in Nymphaeales (Fig. 1A) or Nuphar sister to Cabombaceae (Fig. 1B), were rejected by Kishino-Hasegawa tests. However, this inference is based on only a few characters, as evident from the low Bremer support (Fig. 2) as well as the short branch leading to Nymphaeaceae (Figure 3). A reason for the scarcity of informative characters at the base of Nymphaeaceae could be a rapid, early diversification into the three major lineages. Based on the observation that support values in our dataset increased remarkably after combining different partitions, we can expect that the monophyly of Nymphaeaceae will be corroborated by future studies including more characters. However, there are several morphological and anatomical features of Nuphar, such as its stout creeping rhizomes, a superior syncarpous gynoecium with discontinuous stigmatic rays, echinate anasulcate pollen, and emergent fruits having smooth exarillate seeds with distinctive dehiscence, which demonstrate the distinctness of this genus from other Nymphaeaceae and from Cabombaceae. Such morphological observations also led to the idea of Nupharaceae as a separate family (suggested by Kerner von Marilaun, 1891; Nakai, 1943; Takhtajan, 1997). The earliest fossils (from 52 Ma) that can be unambiguously assigned to any of the extant genera of Nymphaeaceae belong to Nuphar (Chen, Manchester & Chen, 2004). It will therefore be particularly interesting to understand which of the phenotypic features in Nuphar represent a plesiomorphic condition for the Nymphaeales crown group, and which are derived.

Relationships among the three sampled species of Nuphar agree with the hypothesis advanced by Padgett, Les & Crow (1999) in their comprehensive analysis of this genus. Nuphar lutea and N. japonica, as representatives of the Old World Nuphar sect. Nuphar, form a well-supported clade that is sister to N. advena, the only representative of New World Nuphar sect. Astylus in this study.

**Paraphyly of the genus Nymphaea**

Perhaps the most striking result of our analysis is the inferred paraphyly of the genus Nymphaea with respect to Victoria and Euryale. Because previous studies did not consider subgenera of Nymphaea in their analyses (e.g. Ito, 1987; Moseley, Schneider & Williamson, 1993; Les et al., 1999), or found only very weak evidence for the monophyly of Nymphaea (Borsch, 2000; Borsch et al., 2007), this analysis of almost 6000 base pairs of rapidly evolving and non-coding DNA per species yields further insights into this topic. Here, Victoria and Euryale are inferred to be sister to a clade comprising Nymphaea subgenera Hydrocallis and Lotos. This node gains considerable jackknife support (JK = 88), but Bremer support and posterior probabilities are rather low (BS = 1, PP = 0.51). The same applies for node 14, which is the next deeper node and separates the temperate subgenus Nymphaea from the rest of the genus including Ondinea, Euryale and Victoria (JK = 90, BS = 2, PP = 0.91). Although an alternative topology (Fig. 1C),
constraining the genus *Nymphaea* to be monophyletic with respect to *Victoria–Euryale* (but not with respect to *Ondinea*!), is only three steps longer, our total evidence tree was significantly favoured over these alternatives as shown by Kishino-Hasegawa tests.

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

**Ondinea: the apetalous water-lily**

Although doubt may exist on the inferred position of *Victoria* and *Euryale*, our molecular dataset adds further compelling evidence to the hypothesis of a close affinity between *Ondinea* and the Australian water-lilies (*Nymphaea* subgen. *Anecphya*), which was first uncovered by Borsch et al. (2007) using data from the *trnT-trnF* region. The clade comprising *N. elleniae*, *N. macrosperma* and *Ondinea purpurea* receives high support from all statistical tests (JK = 100, BS = 6, PP = 1.0). In fact, our results strongly indicate an origin of *Ondinea* from within the *Anecphya* clade, because *Ondinea* is depicted as sister to the small-seeded *Nymphaea elleniae*. The significance of these results for understanding the evolution of Nymphaeales becomes evident if we consider the striking morphological differences between *Ondinea* and *Nymphaea*. Obviously, dramatic morphological changes involving many parts of the plant have occurred during the evolution of *Ondinea*. A possible mechanism that could explain such drastic shifts is neoteny. Schmucker (1932) observed comparable phenomena, i.e. dwarving and reduction of floral organ number, in *Nymphaea micrantha* clones grown from leaf offshoots. Schneider, Tucker & Williamson (2003) also discuss neoteny in Nymphaeales, but with regard to the reduced flowers in Cabombaceae.

However, despite its distinct morphological features *Ondinea* has been considered to be closely related to *Nymphaea* by other scientists before (Den Hartog, 1970; Williamson & Moseley, 1989; Williamson, Schneider & Malins, 1989). In fact, den Hartog (1970) stated in his description of *Ondinea purpurea*, that ‘it is closer to *Nymphaea* than to any of the other genera within the Nymphaeaceae. In general the *Ondinea* flower can be regarded as an apetalous *Nymphaea* flower.’ The petalous forms of *Ondinea*, which have been described as *Ondinea purpurea* ssp. *petaloidea* by Kenneally & Schneider (1985), provide further evidence for the close relation to *Nymphaea* as they also show the typical morphological gradation from sepals through petals to stamens.

Les et al. (1999) observed in their analysis ‘a weak tendency for *Nymphaea* and *Ondinea* to resolve as a separate clade’, which was induced mainly by molecular data (18S and *matK*). Because Les et al. (1999) sampled only one species of *Nymphaea* subgen. *Nymphaea* (*N. odorata*) they could not detect the close affinity of *Ondinea* to subgenus *Anecphya*, which underscores the importance of judicious taxon sampling (see below). Considering the fact that our tree is based solely on the chloroplast genome, the possibility of a hybrid origin of *Ondinea* remains plausible, with the maternal parent being a water-lily and the paternal being an unknown, presumably extinct member of Nymphaeales. Thus, a cross check with nuclear markers will be necessary to track the real phylogenetic history of *Ondinea*. However, ongoing studies in Australian water-lilies using nuclear markers (present authors’ unpubl. data) currently point to a similar topology.

**Other relationships among and within the subgenera of Nymphaea**

Our results strongly corroborate a close relationship of the *Nymphaea* subgenera *Hydrocallis* and *Lotos*. This affinity has been suggested earlier (e.g. Wiersema, 1987), because it is indicated by several morphological synapomorphies, such as pollen morphology, presence of highly developed carpellary appendages (see below). Considering the fact that our tree is underscores the importance of judicious taxon sampling (see below). Considering the fact that our tree is based solely on the chloroplast genome, the possibility of a hybrid origin of *Ondinea* remains plausible, with the maternal parent being a water-lily and the paternal being an unknown, presumably extinct member of Nymphaeales. Thus, a cross check with nuclear markers will be necessary to track the real phylogenetic history of *Ondinea*. However, ongoing studies in Australian water-lilies using nuclear markers (present authors’ unpubl. data) currently point to a similar topology.

This affinity has been suggested earlier (e.g. Wiersema, 1987), because it is indicated by several morphological synapomorphies, such as pollen morphology, presence of highly developed carpellary appendages, a medial position of anthers on the stamens and the nocturnal flowering pattern. The *Hydrocallis–Lotos* clade is well supported also from subsets of our data, e.g. the *matK* gene (see Table 3), as well as the *trnT-trnF* dataset of Borsch et al. (2007). With this analysis we could also confirm previous findings of Borsch et al. (2007) that the south-east African species *Nymphaea petersiana* does not belong to subgenus *Brachyceras*, as previously assumed, but appears to be closely related to *Nymphaea lotus* (JK = 100, BS = 6, PP = 1.0).

Within subgenus *Hydrocallis, Nymphaea oxypetala* is sister to a clade consisting of *N. amazonum*, *N. jamesoniana* and *N. novogranatensis*. Among the subgenus *Hydrocallis* species sampled *N. oxypetala* is
exceptional owing to its submersed habit, an unusual floral morphology, presence of spherical staminal sclereids, and its polyploid nature (6n = 84; see Wiersema, 1987). A possible affinity of *N. novogranatensis* (2n = 28) to a 2n = 18 group of species including *N. amazonum* has been proposed by Wiersema (1987) based on similarities in morphology, phytochemistry and floral biology. However, because the *N. amazonum–novogranatensis* clade does not gain high statistical support in our analysis (JK = 63, BS = 1, PP = 0.96), any conclusions will have to await substantiation by further data. The analysis of relationships within *Hydrocallis* based on trnT-trnF sequences by Borsch et al. (2007) suffered from too few synapomorphies. However, based on the four species included here, adding a number of further noncoding genomic regions for the remaining species in *Hydrocallis* seems promising in elucidating relationships in the subgenus.

Another highly supported relationship within *Nymphaea* is the clade uniting the subgenera *Anecphya* and *Brachyceras*. Several morphological and anatomical characters coincide with this grouping. The most prominent is the incomplete carpellary fusion in both subgenera, which was first observed by Caspary (1865–66, 1891) and Conard (1905) leading them to group these two subgenera together in the so-called ‘Leptopleura’ (Caspary, 1865–66, 1891) or ‘Apocarpiae’ (Conard, 1905). However, the previously established hypothesis of the paraphyly of subgenus *Brachyceras* with respect to subgenus *Anecphya* (Borsch et al., 2007) could not be confirmed in our analysis. The two sampled representatives of subgenus *Brachyceras*, *N. micrantha* and *N. gracilis*, emerge as a monophylum although it is one of the least-supported clades in the total evidence trees (JK = 78, BS = 1, PP = 0.98).

Similar to the results of Borsch et al. (2007) the subgenus *Nymphaea* is revealed as sister to all other subgenera of *Nymphaeae*, but statistical support is not high enough to dispel any doubt about this position (JK = 90, BS = 2, PP = 0.91). However, a first-branching position of the hardy, northern-hemisphere waterlilies would shed new light on the evolution of the genus *Nymphaea*, in which all other members inhabit tropical to subtropical regions.

**TAXON SAMPLING**

Our own previous studies on molecular phylogenetics in the Nymphaeaceae (Borsch, 2000; Borsch et al., 2007) showed that the hypothesis of a monophyletic genus *Nymphaea* did not gain much statistical support, so we included a more representative taxon sampling of all *Nymphaea* subgenera for the present analysis. By combining multiple chloroplast regions we obtained well-supported hypotheses on intergeneric relationships in Nymphaeales. The paraphyletic nature of the genus *Nymphaea* was not revealed before simply because other studies failed to sample across the subgenera of *Nymphaeae* (e.g. Les et al., 1999) or did not include a sufficient amount of characters (Borsch, 2000; Borsch et al., 2007). Figure 4 points out the relevance of judicious taxon sampling: including only one representative of each genus of Nymphaeales leads to different well-supported topologies, depending on which subgenus of *Nymphaea* was sampled. The correct position of *Ondinea* (according to the total evidence tree) could be inferred only by including *N. elleniae* (subgen. *Anecphya*) or *N. micrantha* (subgen. *Brachyceras*). This effect is also illustrated by indel no. 192 in Figure 5 (occurring in the *matK* gene): if subgenera *Anecphya* and *Brachyceras* had not been sampled, this indel would appear autapomorphic for *Ondinea*, as was the case in the dataset of Les et al. (1999; but indels have not been analysed there). Thus, our study exemplifies the protocol that increasing knowledge on the diversity and phylogenetic relationships within a group of taxa can improve study design and thereby also the scientific output of follow-up analyses.

**MORPHOLOGICAL CHARACTERS IN NYMPHAEALES**

The morphological and anatomical characteristics of Nymphaeaceae and Cabombaceae have been studied intensively since the 1950s (Li, 1955; Moseley, 1958, 1961; Williamson & Moseley, 1989; Schneider et al., 1995; Schneider et al., 2003). A detailed treatment and evaluation of those characters in this study would be out of scope. However, our new insights into the evolution of Nymphaeales reveal the need for a critical reconsideration of morphological and anatomical characters with special regard to variability among the subgenera of *Nymphaea*. Up to now, variability in *Nymphaea* has not been considered in phylogenetic analyses of Nymphaeaceae (Les et al., 1999) or Nymphaeae (Li, 1955; Ito, 1987; Moseley et al., 1993). The inferred close relationship of *Nymphaea* to *Victoria–Euryale* but not to *Ondinea* in the study of Les et al. (1999), was supported only by floral vasculature characters (vascular supply from the receptacular plexus, source and structure of the petal trace, characters 26–28 in their matrix). Although Moseley (1961) reports some variability of floral vasculature within *Nymphaea*, complete information on states of anatomical characters in all five subgenera are mostly missing. A critical re-examination of anatomy and morphology might help to substantiate our new hypotheses on the evolution of Nymphaeales. Further possibly informative characters, that are variable in Nymphaeaceae and within *Nymphaea*, include: structure and number of petiolar and peduncular air canals (Wiersema, 1987).
derived nature of states within Nymphaeales. For example, several character states that are currently interpreted as being autapomorphic for the Nymphaeaceae including *Nuphar*, such as the presence of laticifers, a compound ovary with laminar placentation, numerous seeds with a distinct apical cap, and numerous petals and stamens, could in fact be plesiomorphic in Nymphaeales. If this were the case, such characters would not contradict a position of *Nuphar* as basal in Nymphaeales.
Nymphaeales or sister to Cabombaceae. However, it might be a challenging task to compare morphological and anatomical traits of Nymphaeales with those of outgroup taxa, because this plant group is characterized by unique features, which have resulted mainly from their early separation from the rest of angiosperms and drastic phenotypic changes in the course of their adaptation to the aquatic habit.

PHYLOGENETIC SIGNAL OF DATA PARTITIONS

The increased application of molecular data in plant phylogenetics has led to an enormous amount of sequence datasets. Because more and more molecular data are becoming available, a debate on whether information from different genomic regions should be analysed in combination or individually to test phylogenetic hypotheses is going on (e.g. Bull et al., 1993; Cunningham, 1997; Castoe, Doan & Parkinson, 2004). Whereas the focus of this debate is on the treatment of heterogeneous data partitions, the combined analysis of homogeneous data is generally accepted (Bull et al., 1993; Chippindale & Wiens, 1994; De Queiros, Donoghue & Kim, 1996). There is no apparent conflict among the data partitions in the present analysis, as revealed by partition homogeneity tests as well as partitioned Bremer support. Furthermore, homoplasy is generally low across all partitions. Therefore, the assessment of phylogenetic utility can be confined to the evaluation of sequence divergence, number of informative characters, frequency of length mutations and hotspots, as well as the phylogenetic structure R (Quandt et al., 2003; Müller et al., 2006) inherent in each data partition.

The highest phylogenetic structure R, i.e. the highest average support per node, was observed in the rpl16 intron, the trnK intron and the matK gene. These three markers are also characterized by the highest percentages and absolute numbers of informative characters in our dataset. In general, the introns and the matK gene provide more information than spacers at this taxonomic level. The spacers are more variable (both in nucleotide substitution as well as in sequence length) but the percentage of informative nucleotide and indel characters is similar or smaller than in introns. This coincides with a lower overall phylogenetic structure (R = 0.65 in spacers vs. R = 0.82 in introns; Table 3).

However, a general conclusion on the phylogenetic utility of spacers relative to introns cannot be drawn here, because there can be considerable difference among differently evolving spacers. Transcribed spacers, such as the petB-petD and the trnL-trnF spacers in our analysis, are much more conserved than nontranscribed spacers. Differences in information content and phylogenetic utility exist also among the introns investigated in our analysis. The introns in trnL and petD, which have been shown to provide valuable information in a broader taxonomic context (i.e. the basal angiosperms; Borsch et al., 2003; Löhne & Borsch, 2005), perform below average in Nymphaeales. Compared to other group II introns currently used in phylogenetic analyses, the intron in petD seems to be the most conserved of the chloroplast genome single copy regions. It may thus provide a signal that is particularly useful for reconstructing relationships of more distant angiosperms (e.g. among families). However, the present results on phylogenetic utility might be influenced by lineage specific effects and generalizations will have to await studies in different angiosperm clades. In fact, relative rates of molecular evolution seem to be rather low in Nymphaeales compared to other angiosperms (K. Müller & A. Worberg, unpubl. data).

The introns in trnK and rpl16, which are among the best performing markers in our analysis, seem to be generally more suitable than orders that represent rapid radiations because they have been applied successfully in previous comparable analyses (e.g. Renner & Chanderbali, 2000, for rpl16 in Laurales; Müller & Borsch, 2005a, for trnK in Lamiales). The study of Shaw et al. (2005) also revealed the rpl16 intron as one of the more variable introns and showed the trnL intron to be the most conserved among the four introns studied (rps16, trnG, trnL, and rpl16). In contrast to our results, Shaw et al. (2005) designated the trnK-matK region as well as trnT-trnF as less phylogenetically informative than most of the 21 regions included in their analysis, based on the relative amount of potentially informative characters (i.e. nucleotide substitutions, indels and inversions) provided by each marker. However, Shaw et al. (2005) were searching for suitable markers for analyses at infrageneric levels, which is a completely different goal. The special patterns of molecular evolution in the matK gene, that cause the high quantity of information and high quality of characters compared to other coding regions of the chloroplast region (Müller et al., 2006), are the reason for the broad utility of this marker. The matK gene has been shown to be more informative than any other single marker at higher taxonomic levels (Hilu et al., 2003; Müller et al., 2006), and again in this phylogenetic analysis of Nymphaeales it was one of the most effective markers.

Our study adds another piece of evidence to the general phylogenetic utility of noncoding and fast-evolving regions at higher taxonomic levels. Extreme sequence variability and homoplasy is generally confined to mutational hot spots, and therefore those markers can easily be employed in investigations including a broad taxonomic spectrum, often providing better resolution and support than rather slowly
evolving genes. The increasing number of datasets implementing noncoding and fast-evolving sequences at higher taxonomic levels (e.g. Hilu et al., 2003; Quandt et al., 2004; Borsch et al., 2005; Löhne & Borsch, 2005; Qiu et al., 2005; see Borsch et al., 2005 for review) confirm those markers as promising tools in molecular phylogenetics.

ACKNOWLEDGEMENTS

This paper is part of the PhD dissertation of C.L. that was carried out in the framework of the Graduiertenkolleg ‘Evolution und Biodiversität in Raum und Zeit’ funded by the Deutsche Forschungsgemeinschaft (DFG). We thank W. Barthlott, T. Litt, and K. P. Sauer for continuous support during this program. Financial support from the International Water Lily and Water Gardening Society for field work to T.B. is greatly appreciated. T.B. acknowledges support through a Heisenberg Scholarship of the Deutsche Forschungsgemeinschaft. Furthermore, we are indebted to: C. Barre Hellquist and Surrey W. L. Jacobs for providing material of Australian water-lilies and gemeinschaft. Furthermore, we are indebted to: C. Ondinea Chawanje for samples of N. petersiana; Alejandro Novelo and Nur Ritter for prior assistance in Chrissie Chawanje for samples of N. petersiana; Alejandro Novelo and Nur Ritter for prior assistance in obtaining material of neotropical Nymphaeae; and the Botanic Gardens of Bonn University, especially Dr Wolfram Lobin and Michael Neumann, for curating the living collection of water-lilies. We thank Pamela Sch, 2005; Qiu et al., 2005; see Borsch et al., 2005 for

REFERENCES


The following supplementary material is available for this article:

**Appendix S1.** Primers used in this study.  
**Appendix S2.** PCR and cycle sequencing conditions.  
**Appendix S3.** Results of partition homogeneity tests (P-values of pairwise comparisons).  
**Appendix S4.** Results of the Kishino-Hasegawa tests.

This material is available as part of the online article from:  
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.