CSIRO PUBLISHING
www.publish.csiro.au/journals/asb *Australian Systematic Botany*, **21**, 229–250

Nuclear and plastid DNA sequences reveal complex reticulate patterns in Australian water-lilies (*Nymphaea* subgenus *Anecphya*, Nymphaeaceae)

Cornelia Löhne[A,E,F], *Thomas Borsch*[A,E], *Surrey W. L. Jacobs*[B], *C. Barre Hellquist*[C] *and John H. Wiersema*[D]

[A]Nees-Institut für Biodiversität der Pflanzen, Rheinische Friedrich-Wilhelms-Universität, 53115 Bonn, Germany.
[B]National Herbarium, Royal Botanic Gardens, Sydney, NSW 2000, Australia.
[C]Department of Biology, Massachusetts College of Liberal Arts, North Adams, MA 01247-4100, USA.
[D]Systematic Botany and Mycology Laboratory, Agricultural Research Service, United States Department of Agriculture, Beltsville, MD 20705-2350, USA.
[E]Present address: Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Str. 6-8, 14195, Berlin.
[F]Corresponding author. Email: c.loehne@bgbm.org

Abstract. This study represents the first comprehensive analysis of phylogenetic relationships within the Australian water-lilies, *Nymphaea* subg. *Anecphya*. Our 51-accession dataset covers all 10 species of the subgenus, except the newly described *N. alexii*, and includes information from the nuclear ITS as well as from the chloroplast *trnT–trnF* region. The results show that molecular data are consistent with morphology, because the subdivision of subg. *Anecphya* into two major clades, a large-seeded and a small-seeded group, could be confirmed. Within the large-seeded group, *Nymphaea atrans* and *N. immutabilis* seem to form one clade, whereas samples of *N. gigantea*, *N. georginae*, *N. macrosperma* and *N. carpentariae* form another. Relationships within the small-seeded group, containing all samples of *N. violacea*, *N. elleniae* and *N. hastifolia*, are less clear, since the trees obtained from the chloroplast and the nuclear marker are incongruent. The samples of *N. violacea* do not form a monophyletic group in each of the trees, but—at least in the ITS tree—group with either *N. elleniae* or *N. hastifolia/Ondinea*, respectively. Polymorphisms among ITS paralogues, i.e. substitutions at single nucleotide positions and length polymorphisms, have been observed in some samples of *N. violacea.* This fact as well as the incongruent phylogenetic signal obtained from the chloroplast and the nuclear genomes point to recent hybridisation or introgression in this group. Remarkably, *Ondinea purpurea* is resolved within the small-seeded group by both markers and seems to have a close relationship to *N. hastifolia.* Although incomplete lineage sorting cannot be fully excluded to explain high variability in *N. violacea*, molecular data potentially hint to a case of still imperfect taxonomy.

Introduction

Within the family Nymphaeaceae, the water-lilies (*Nymphaea*) represent the most diverse and most widespread genus, with some of the ~50 species on every continent except Antarctica. The genus *Nymphaea* has been subdivided into the following five subgenera: *Anecphya*, *Brachyceras*, *Hydrocallis*, *Lotos* and *Nymphaea* (Caspary 1891; Conard 1905). Three of these subgenera are widespread, the palaeotropical subg. *Lotos*, the pantropical subg. *Brachyceras* and the northern temperate subg. *Nymphaea*, whereas subg. *Hydrocallis* is restricted to tropical and subtropical America and subg. *Anecphya* to Australasia. Apart from *Nymphaea* subg. *Anecphya*, a few species from other nymphaean subgenera and another genus of Nymphaeaceae grow naturally in Australia. These include *Ondinea purpurea* Hartog, which is endemic to some coastal areas of the Kimberley district in Western Australia, *Nymphaea nouchali* Burm. f. (subg. *Brachyceras*), which grows mostly near coastal areas in Asia and occasionally in tropical Australasia, and *N. pubescens* Willd. (subg. *Lotos*), which is a widespread palaeotropical species that inhabits coastal areas of the Northern Territory and, rarely, northern Queensland. A few other water-lily species have become naturalised in Australia (Jacobs and Porter 2007).

From the Australian water-lilies belonging to subg. *Anecphya* currently 10 species are recognised, and these are listed in Table 1. Most species are restricted to the monsoonal parts of the Australian tropics, with only *N. gigantea* growing south of the Tropic of Capricorn in Queensland and New South Wales (Jacobs and Porter 2007). Some species are widespread in the monsoonal parts of Australia (*N. immutabilis*, *N. macrosperma*, *N. violacea*), whereas others are less widely distributed, e.g. *N. hastifolia* in the higher-rainfall areas of tropical Western Australia and Northern Territory, and *N. atrans* and *N. elleniae* on Cape York Peninsula (northern Queensland). *Nymphaea macrosperma*, *N. violacea* and *N. elleniae* also grow in New Guinea.

 10.1071/SB07010 1030-1887/08/040229

Table 1. The Australasian water-lily subgenus *Nymphaea* subg. *Anecphya*

Species	Number of accessions analysed in this study
Large-seeded group	
N. atrans S.W.L. Jacobs	2
N. carpentariae S.W.L. Jacobs & Hellq.	4
N. georginae S.W.L. Jacobs	1
N. gigantea Hook.	4
N. immutabilis S.W.L. Jacobs	8
N. macrosperma Merr. & L.M. Perry	5
Small-seeded group (= subgen. *Confluentes sensu* Jacobs 2007)	
N. alexii S.W.L. Jacobs & Hellq.	0
N. elleniae S.W.L. Jacobs & Hellq.	3
N. hastifolia Domin	1
N. violacea Lehm.	18

Water-lilies of the subgenus *Anecphya* are characterised by clearly emergent, often very large flowers. Some floral characters are shared with the closely related subgenus *Brachyceras*, such as the incomplete fusion of carpel walls ('Apocarpiae' *sensu* Conard 1905), and the common occurrence of blue petals. In contrast to subg. *Brachyceras*, in which the carpellary appendages are only slightly developed, members of subg. *Anecphya* do not possess them at all and, additionally, often show extremely high numbers of stamens (up to 600 in some members, Jacobs and Porter 2007). Some morphological distinctions can be made among the members of subg. *Anecphya* s.lat.: one group of species is characterised by a distinctive gap between petals and stamens, rather large seeds and toothed leaf margins (subg. *Anecphya* s.str., *N. macrosperma*, *N. gigantea*, *N. immutabilis*, *N. atrans*, *N. carpentariae*, *N. georginae*), whereas another group of species is characterised by petals grading into stamens, relatively small seeds and entire to sinuate leaf margins (*N. violacea*, *N. elleniae*, *N. hastifolia*, *N. alexii*). Jacobs (2007) is transferring the species of the latter group to the new subgenus *Confluentes*. All Australian water-lilies of subg. *Anecphya* are day-blooming and possess erect rhizomes.

The habitats of the above taxa vary from shallow creeks, natural lakes, ponds, lagoons and billabongs (i.e. ephemeral or perennial pools in river flood channels) to man-made dams. Most of the taxa grow in acidic waters with a pH below 6. *N. elleniae* grows in comparatively low-nutrient perennial river or swamp systems on Cape York and Papua New Guinea. Some populations of *N. violacea* also grow in low-nutrient systems in streams with a sandstone or granite catchment. All of the other species in subgenus *Anecphya* s.lat., and including many populations of *N. violacea*, mostly grow in floodplain billabongs or waterholes, or deeper holes of low-energy parts of small rivers or streams. Many of these habitats dry on an annual or few-year cycle, with the plants perennating by either the tuberous rhizomes or seeds or, most commonly, a mixture of both.

The subgenus *Anecphya* has a complex taxonomic history. Conard (1905) was the first to formally establish subg. *Anecphya*, referring to two earlier publications by Caspary (1865−1866, 1891). Caspary, however, referred to *Anecphya* as a subsection and included it in a different section in each publication. Although many of Caspary's subsections are now regarded as subgenera, some are still treated as equivalent to subsections. Both Caspary and Conard included only one species in the group, *N. gigantea.* Caspary seems to have missed or disregarded the publication of *N. violacea* by Lehmann (1853) but Conard reduced it to a variety of *N. gigantea*, overlooking two distinctive features in his description that would have excluded it both from the species and the subgenus (*sensu stricto*, Jacobs 2007), i.e. the lack of a space between the petals and the stamens and the small seeds. Conard's (1905) misinterpretation of *N. violacea* caused confusion until Jacobs (1992) was able to sort it out. In the meanwhile, all the new and currently recognised species described from Australia (*N. hastifolia* Domin (1929), *N. macrosperma* Merrill & Perry (1942)) were simply assumed to also belong to subg. *Anecphya.* The species described by Jacobs (1992) and Jacobs and Hellquist (2006) were described without comment as to subgeneric classification.

A close relationship of subgenera *Anecphya* and *Brachyceras* has been demonstrated in recent phylogenetic analyses on the basis of molecular data (Borsch *et al*. 2007; Löhne *et al*. 2007). However, whereas *Anecphya* clearly appears as a well supported clade in both studies, there is evidence from the chloroplast *trnT−trnF* region and a dense taxon sampling that *Brachyceras* might be paraphyletic with respect to *Anecphya* (Borsch *et al*. 2007). Another, more striking result of these two recent studies is the lack of statistical support for the monophyly of the genus *Nymphaea*. In fact, a comprehensive analysis of Nymphaeales including 12 species of *Nymphaea* and multiple regions of the genome depicted the South American genus *Victoria* and the Asian genus *Euryale* as sister to a *Hydrocallis−Lotos* clade, but with rather low support (Löhne *et al*. 2007). Furthermore, combined evidence from the chloroplast genome strongly indicated that the Australian endemic *Ondinea purpurea* is derived from within the Australian water-lilies *Nymphaea* subg. *Anecphya* (Borsch *et al*. 2007; Löhne *et al*. 2007). However, plastid gene relationships need to be evaluated for their conclusiveness in light of the organismic evolutionary history in *Nymphaea* because evidence for present or possible past hybridisation exists throughout the genus (Les *et al*. 2004; Woods *et al*. 2005).

The present study is the first examination of the phylogenetic relationships in *Nymphaea* subg. *Anecphya*. In order to get a full picture and to unravel possible reticulate events at the species level, the analysis is based on both the nuclear internal transcribed spacer region (ITS) and the chloroplast *trnT−trnF* region. Both markers have been widely used for phylogenetic studies at low taxonomic levels (Álvarez and Wendel 2003; Shaw *et al*. 2005, and references therein), and have been effective in previous studies on the genus *Nymphaea* (Borsch 2000; Woods *et al*. 2005; Borsch *et al*. 2007). The present study comprises the first extensive taxon sampling with 46 individuals of different populations, representing all but one (*N. alexii*) of the currently described species and subspecies and the entire range of distribution of subg. *Anecphya* in Australia. The large sample base became available through the extensive collections of S. Jacobs and B. Hellquist during field studies in 1997 and 2002.

Ondinea purpurea is also included in the taxon sampling in order to verify the previous results from the chloroplast genome.

Representatives of *N.* subg. *Brachyceras* are chosen as outgroup taxa. On the basis of the described project design, the following questions are addressed: are the observed two major groups in *N.* subg. *Anecphya* supported by molecular data; what are the relationships within these groups; is there any evidence for hybridisation or introgression among the *Anecphya* taxa; can the previously inferred close affinity of *Ondinea purpurea* to the Australian water-lilies be reproduced with nuclear and chloroplast sequences and an extensive taxon sampling; and is there any specific taxon in *N.* subg. *Anecphya* to which *Ondinea* might be more closely related?

Materials and methods

Taxon sampling and plant material

The dataset used in the present study comprises 46 representatives of *Nymphaea* subg. *Anecphya* plus *Ondinea purpurea* and 4 members of *Nymphaea* subg. *Brachyceras* as outgroup taxa. The majority of plant material was collected by Jacobs, Hellquist and Wiersema between 1997 and 2002, during visits to the Northern Territory, Western Australia and Queensland. Detailed information on all specimens, including field localities, collectors and vouchers are given in Table 2. Coordinates of the field localities from all sampled specimens were digitised by using the geographic information software ArcView 3.2. and localities were plotted on a map of Australia. If coordinates were not given by the collectors they were approximated on the basis of the name of the locality (marked by '~' in Table 2). Most of the sequence data were generated during the course of this study. A few of the sequences were taken from the analysis of Borsch *et al*. (2007).

DNA isolation, amplification and sequencing

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

ITS

The nuclear marker was amplified in a single fragment with the standard primers ITS4 and ITS5 (White *et al*. 1990). This region spans the internal transcribed spacer 1 (ITS1) between 18S and 5.8S rDNA, the 5.8S rDNA itself, and the internal transcribed spacer 2 (ITS2) between 5.8S and 26S rDNA (Fig. 1). PCR was conducted on a T3 Thermocycler (Biometra, Göttingen, Germany) with standard PCR conditions, purified by using a QiaQuick gel extraction kit (QIAGEN Inc., Valencia/CA, USA) and sequenced directly by using an CEQ^{TM} DTCS Quick Start Kit (Beckman Coulter, Fullerton/CA, USA) on a CEQ^{TM} 8000 sequencer. Amplification primers were also used for sequencing both strands, but in some cases these primers were not sufficient to read through the complete fragment because of length polymorphisms in the original DNA (see results). Additional internal primers (NY-5.8F, NY-5.8R, and ITS1-R) were designed to overcome these problems (Fig. 1, Table 3).

trnT–trnF

The chloroplast marker was amplified in two fragments, with primers rps4-5R and trnL110R for the 5′-fragment and primers C and F for the 3′-fragment (see Table 3 for primer sequences and authors). Amplification, purification of products and sequencing followed the same procedures as described for ITS. In addition to the amplification primers, the internal sequencing primers D and E were necessary in order to read through a complex microsatellite region (poly-T and poly-TA in the P8 stem loop of the *trnL* intron, Fig. 1, Table 3).

Sequence alignment and phylogenetic analysis

Both genomic regions were aligned manually with BioEdit version 7.0.5 (Hall 1999), following the rules outlined in Löhne and Borsch (2005). Mutational hotspots (Borsch *et al*. 2003 for definition), and all parts of exons were excluded from analysis. Indels were coded automatically in a '01'-matrix with SeqState version 1.25 (Müller 2005), applying the 'simple indel coding' strategy after Simmons and Ochoterena (2000). Both, the ITS and the *trnT–trnF* dataset, were analysed separately including either only substitutions or substitutions + indels. Combined analyses of the nuclear and the chloroplast datasets were not conducted because of apparent incongruence.

Maximum parsimony

Maximum parsimony (MP) ratchet analyses (Nixon 1999) were conducted in PAUP* version 4.0b10 (Swofford 2002) with the help of PRAP version 1.21 (Müller 2004). Twenty random addition cycles of 200 ratchet iterations were applied, with 25% of the positions reweighted. The four representatives of *Nymphaea* subg. *Brachyceras* were chosen as outgroup to root the trees. Node support was estimated through jackknifing (JK) 10 000 replicates (simple addition, keeping 1 tree per replicate, deleting 36.8% of characters in each replicate). Additionally, Bremer support (BrS) was calculated with PAUP* and PRAP (10 random addition replicates per constraint tree, parsimony ratchet employed).

Bayesian inference

Modelltest version 3.06 (Posada and Crandall 1998) was used to determine the best models of molecular evolution for ITS and *trnT–trnF*. In both cases, the GTR + G model was selected according to the Akaike information criterion. Bayesian analyses of the datasets (substitutions + indels) were performed by using MrBayes (Ronquist and Huelsenbeck 2003), with the binary (restriction site) model applied to the indel partition. Analysis was performed for 1 000 000 generations (settings: MCMCMC, 4 runs with 4 chains each, saving 1 tree every 100 generations). In the case of the ITS, 44-accession dataset (see below) tree likelihoods had converged to a stable value after 10 000 generations. Thus, the 'burnin' was set to 100 and 39 603 trees were sampled for calculating the consensus tree and the posterior probabilities (PP) of nodes. In the four runs of the *trnT–trnF*, 51-accession analysis probabilities had converged after 20 000 generation or earlier; therefore, the burnin was set to 200 and 39 204 trees were sampled.

Table 2. Plant material of 51 samples used in the present study on *Nymphaea* subg. *Anecphya*

Information on collectors (BH = B. Hellquist, JW = J. Wiersema, KB = K. Brennan, KW = K. Woods, AN = A. Novelo R., SJ = S. W. L. Jacobs, TB = T. Borsch, VW = V. Wilde, AL = Andre Leu), location of voucher specimens and geographic information on field origin are included. Coordinates were approximated on the basis of the name of the locality (marked with '~'), if not given by the collectors. Minutes are given in decimal values. Coordinates have not been recorded for outgroup taxa (marked with *). Most of the DNA sequence data were generated for the present study. Only some *trnT–trnF* sequences were taken from previous studies by to authors (Borsch *et al.* 2007; Löhne *et al.* 2007, GenBank numbers starting with 'AM')

Species (DNA sample number)	Field origin (region, locality; latitude, longitude)	Voucher information (collector, number, herbarium)	ITS GenBank number	*trnT–trnF* GenBank number
Nymphaea atrans S.W.L.Jacobs (NY102)	Queensland; Low Lake, Lakefield Nat. Park; 14°38,352′S, 143°54,381′E	*SJ, BH & JW 8212*; NASC, NSW, BRI	FJ026554	AM422055
N. atrans S.W.L.Jacobs (NY432)	Queensland; Low Lake, Lakefield Nat. Park; 14°38,352′S, 143°54,381′E	*BH & AL 16766*; NASC, NSW,BRI	FJ026555	FJ026518
N. carpentariae S.W.L.Jacobs & Hellq. (NY396)	Queensland; NE of Normanton; 17°31,712′S, 141°82,696′E	*SJ & BH 8757*; NASC, NSW, BRI	FJ026556	FJ026519
N. carpentariae S.W.L.Jacobs & Hellq. (NY398)	Queensland; 73 km W of Normanton on Burketown Rd; 18°7,362′S, 140°32,249′E	*SJ & BH 8768*; NASC, NSW, BRI	FJ026557	FJ026520
N. carpentariae S.W.L.Jacobs & Hellq. (NY399)	Queensland; Burketown; 17°44,774′S, 139°32,888′E	*SJ & BH 8770*; NASC, NSW, BRI	FJ026558	FJ026521
N. carpentariae S.W.L.Jacobs & Hellq. (NY434)	Queensland; Cattle Dam, S of Greenvale; 19°0,258′S, 145°1,302′E	*BH & AL 16774*; NASC, NSW, BRI	FJ026559	FJ026522
N. elleniae S.W.L.Jacobs (NY103)	Queensland; Bamaga; 10°53,604′S, 142°23,237′E	*SJ, BH & JW 8224*; NASC, NSW, BRI	FJ026560	AM422056
N. elleniae S.W.L.Jacobs (NY137)	Queensland; Jardine River; 11°9,046′S, 142°21,338′E	*SJ, BH & JW 8227*; NASC, NSW, BRI	FJ026561	AM422057
N. elleniae S.W.L.Jacobs (NY381)	Queensland; Jardine River; 11°9,046′S, 142°21,338′E	*BH & AL 16757*; NASC, NSW, BRI	FJ026562	AM489714
N. georginae S.W.L.Jacobs & Hellq. (NY425)	Northern Territory; James River; 20°1,517′S, 137°29,55′E	*SJ & BH 8868*; NASC, NSW, PERTH	FJ026563	FJ026523
N. gigantea Hook. (NY126)	Queensland; Condamine River; 26°59,08′S, 150°6,69′E	SJ & BH 8357; NASC, NSW, BRI	FJ026564	AM422059
N. gigantea Hook. (NY426)	Queensland; Chinamans Lagoon, 1 km S Miles; 26°40,114′S, 150°11,167′E	*SJ & BH 8870*; NASC, NSW, BRI	FJ026565	FJ026524
N. gigantea Hook. (NY435)	Queensland; Cattle Creek, S of Ingham 18°44,765′S, 146°8,492′E	*BH & AL 16772*; NASC, NSW, BRI	FJ026566	FJ026525
N. cf. *gigantea* Hook. (NY395)	Queensland; Reid River; 19°45,739′S, 146°50,067′E	*SJ & BH 8752*; NASC, NSW, BRI	FJ026567	FJ026526
N. hastifolia Domin (NY134)	Northern Territory; Darwin 12°47′S, 130°92′E	*JW & BH*; no voucher	FJ026568	AM422060
N. immutabilis S.W.L.Jacobs (NY121)	Queensland; Cabbage Tree Ck; 15°18,19′S, 144°36,97′E	*BH, JW & KB*; no voucher	FJ026569	AM422061
N. immutabilis S.W.L.Jacobs (NY136)	Queensland; Mt Molloy–Mareeba Rd ~16°50′S, 145°20′E	*BH, JW & KB*; no voucher	FJ026570	AM422062
N. immutabilis S.W.L.Jacobs (NY462)	Queensland; Ninds Ck, Innisfail; 17°34,263′S, 146°2,298′E	*BH & AL 16770*; NASC, NSW, BRI	FJ026571	FJ026527
N. immutabilis S.W.L.Jacobs (NY503)	Queensland; Waluma Swamp, Mt Garnet; 17°42,229′S, 145°7,749′E	*BH & AL 16775*; NASC, NSW, BRI	FJ026572	FJ026528
N. immutabilis S.W.L.Jacobs (NY383)	Queensland; Langi Lagoon, Rokeby; 13°27,052′S, 142°42,004′E	*BH & AL 16760*; NASC, NSW, BRI	FJ026573	FJ026529
N. immutabilis S.W.L.Jacobs (NY427)	Queensland; Ross River, Townsvill; 19°21,21′S, 146°43,94′E	*BH & AL 16773*; NASC, NSW, BRI	FJ026574	FJ026530
N. cf. *immutabilis* S.W.L.Jacobs (NY450)	Queensland; Low Lake, Lakefield NP; 14°38,352′S, 143°54,381′E	*BH & AL 16767*; NASC, NSW, BRI	FJ026575	FJ026531
N. immutabilis ssp. *kimberleyensis* S.W.L.Jacobs (NY380)	Western Australia; Brooking Springs; 18°7,248′S, 125°36,121′E	*SJ & BH 8813*; NASC, NSW, PERTH, K, B	FJ026576	FJ026532
N. macrosperma Merr. & L.M.Perry (NY373)	Northern Territory; Yellow Water, Kakadu NP; ~12°54′S, 132°31′E	*SJ & BH 8802*; NASC, NSW, DNA	FJ026577	FJ026533
N. macrosperma Merr. & L.M.Perry (NY391)	Northern Territory; Wildman River, Kakadu NP; 12°34,25′S, 132°13,118′E	*SJ & BH 8796*; NASC, NSW, DNA, B, G	FJ026578	FJ026534

Table 2. (*continued*)

Species (DNA sample number)	Field origin (region, locality; latitude, longitude)	Voucher information (collector, number, herbarium)	ITS GenBank number	*trnT–trnF* GenBank number
N. macrosperma Merr. & L.M.Perry (NY127)	Northern Territory; Island Billabong, W of road to Ubirr along Magela Ck, N of Jabiru, Kakadu NP; ~12°45′S, 132°30′E	*BH, JW & KB 16181*; NASC, NSW, DNA	FJ026579	AM422063
N. cf. *macrosperma* Merr. & L.M.Perry (NY418)	Northern Territory; W of Roper Bar; 14°42,273′S, 132°27,653′E	*SJ & BH 8864*; NASC, NSW, DNA	FJ026580	FJ026535
N. cf. *macrosperma* Merr. & L.M.Perry (NY433)	Western Australia; Perrys Lagoon; 15°32,969′S, 128°15,592′E	*SJ, BH & KWi 10143*; NASC, NSW, PERTH	FJ026581	FJ026536
N. violacea Lehm. (NY110)	Queensland; ~22 km W of Batavia Downs–Weipa road; 12°42,62′S, 142°30,82′E	*SJ, BH & JW 8230*; NASC, NSW, BRI	FJ026582	AM422064
N. violacea Lehm. (NY131)	Queensland; Mt Molloy; 16°41,188′S, 145°19,572′E	*BH & AL 16589*; NASC, NSW, BRI	FJ026583	FJ026537
N. violacea Lehm. (NY382)	Queensland; Langi Lagoon, Rokeby; 13°27,052′S, 142°42,004′E	*BH & AL 16759*; NASC, NSW, BRI	FJ026584	FJ026538
N. violacea Lehm. (NY436)	Queensland; Palmer River Roadhouse; 16°6,443′S, 144°46,63′E	*BH & AL 16754*; NASC, NSW, BRI	FJ026585	FJ026539
N. violacea Lehm. (NY448)	Queensland; Bobs Lagoon, Rokeby; 13°27,305′S, 142°43,603′E	*BH & AL 16761*; NASC, NSW, BRI	FJ026586	FJ026540
N. violacea Lehm. (NY501)	Queensland; S of Coen; 14°0,288′S, 143°11,436′E	*BH & AL 16755*; NASC, NSW, BRI	FJ026587	FJ026541
N. violacea Lehm. (NY502)	Queensland; Sweetwater Lake, Lakefield; 14°39,813′S, 143°50,315′E	*BH & AL 16765*; NASC, NSW, BRI	FJ026588	FJ026542
N. violacea Lehm. (NY504)	Queensland; Wonga Beach; 16°19,774′S, 145°24,545′E	*BH & AL 16779*; NASC, NSW, BRI	FJ026589	FJ026543
N. violacea Lehm. (NY372)	Northern Territory; Kangaroo Ck; 16°49,841′S, 137°9,546′E	*SJ & BH 8779*; NASC, NSW, DNA	FJ026590	FJ026544
N. violacea Lehm. (NY374)	Northern Territory; E of Timber Ck; 15°44,014′S, 130°32,573′E	*SJ & BH 8862*; NASC, NSW, DNA	FJ026591	FJ026545
N. violacea Lehm. (NY405)	Northern Territory; Jabiru Lake; 12°40,429′S, 132°50,534′E	*SJ & BH 8792*; NASC, NSW, DNA	FJ026592	FJ026546
N. violacea Lehm. (NY407)	Northern Territory; Jim Jim Billabong, Kakadu; 12°56,538′S, 132°32,076′E	*SJ & BH 8799*; NASC, NSW, DNA	FJ026593	FJ026547
N. violacea Lehm. (NY409)	Northern Territory; Fogg Dam; 12°36,631′S, 131°17,989′E	*SJ & BH 8803*; NASC, NSW, DNA	FJ026594	FJ026548
N. violacea Lehm. (NY410)	Northern Territory; Little Roper River; 14°45,583′S, 132°37,166′E	*SJ & BH 8863*; NASC, NSW, DNA	FJ026595	FJ026549
N. violacea Lehm. (NY420)	Northern Territory; Drysdale Station; ~15°34′S, 126°13′E	*SJ & BH 8845*; NASC, NSW, PERTH	FJ026596	FJ026550
N. violacea Lehm. (NY413)	Western Australia; Apex Ck; 17°6,437′S, 125°10,923′E	*SJ & BH 8834*; NASC, NSW, PERTH	FJ026597	FJ026551
N. violacea Lehm. (NY419)	Western Australia; March Fly Glen; 17°9 784′S, 125°18, 62′E	*SJ & BH 8835*; NASC, NSW, PERTH	FJ026598	FJ026552
N. violacea Lehm. (NY424)	Western Australia; Hidden Valley, Kununurra; 15°46,334′S, 128°45,051′E	*SJ & BH 8860*; NASC, NSW, PERTH	FJ026599	FJ026553
Ondinea purpurea Hartog (NY377)	Western Australia; ~6 km N of Kalumburu, Pago Rd; 14°15,363′S, 126°37,216′E	*SJ & BH 8853*; NASC, NSW, PERTH	FJ0265600	AM422023
* *N. elegans* Hook. (NY006)	USA, Florida, Collier County	*TB & VW 3084*; FR	FJ0265601	AM422047
* *N. elegans* Hook. (NY370)	USA, Texas	*TB & KW 3424*; BONN, VPI	FJ0265602	–
* *N. heudelotii* Planch. (NY066)	Bonn Botanic Gardens, Acc. No. 14244 [Rwanda]	*E. Fischer s.n.*; BONN	FJ0265603	AM422052
* *N. micrantha* Guill. & Perr. (NY007)	Bonn Botanic Gardens, Acc. No. 5830 [Zimbabwe]	*Koehnen s.n.*; BONN	–	AM422051
* *N. ampla* (Salisb.) DC. (NY100)	Mexico, Veracruz	*AN et al. 1295*; MEXU	FJ026604	AM422044

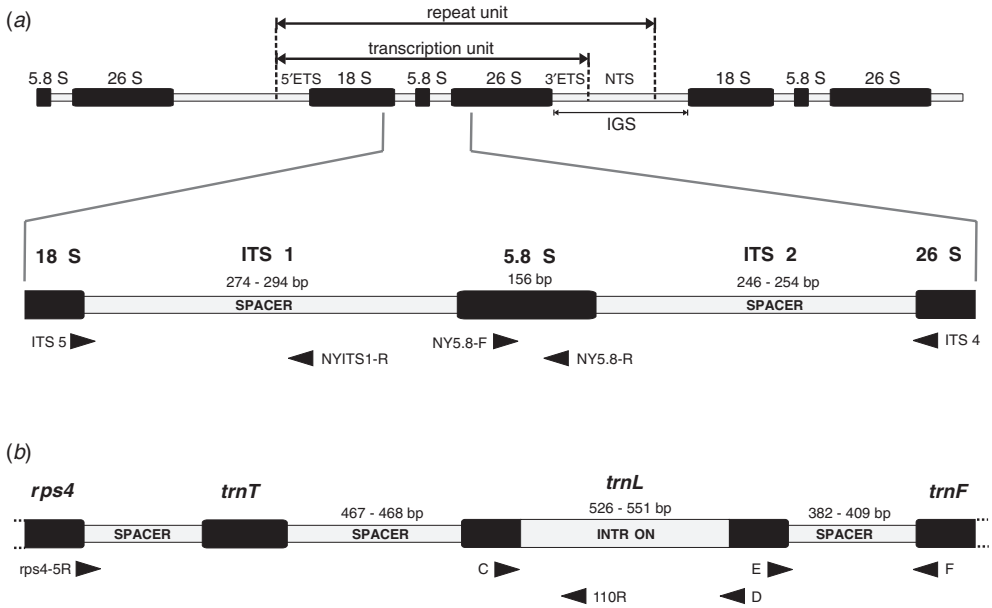

Fig. 1. Structure of the genomic markers used for phylogenetic analysis of *Nymphaea* subg. *Anecphya* and position of primers. (*a*) The ITS (internal transcribed spacer) region is part of a larger repeat unit of the nuclear genome comprising the ribosomal 18S, 5.8S and 26S DNA, the intergenic spacer (IGS) between 26S and 18S, and the internal transcribed spacers between 18S and 5.8S (ITS1) and between 5.8S and 26S (ITS2). Parts of the IGS are also transcribed (ETS = external transcribed spacer, NTS = non-transcribed spacer). (*b*) The *trnT*–*trnF* region of the chloroplast genome. Coding stretches are marked by black boxes; non-coding regions are grey. Arrows mark the position and reading direction of primers used in this study. Length ranges observed in the present data set are given above the boxes.

Table 3. Primers used in the present study for amplification (A), sequencing (S) or both (A,S)
Additionally, the reading direction, i.e. forward (F) or reverse (R), as well as nucleotide sequences and authors of the primers are given. See also Fig. 1 for position of primers

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

Results

Sequence variability in ITS

The ITS dataset comprised a total of 695 characters of which 40% were obtained from ITS1, 38% from ITS2 and 22% from the 5.8S rDNA. The length variability among the sequences was relatively low. Several length mutations of 1–3 nucleotides occurred in ITS1 and ITS2, which were coded as 35 indel characters in total. The 5.8S rDNA had a conserved length of 156 nucleotides in all sequences. Sequence divergence was highest in ITS1 (8.0%), lower in ITS2 (5.9%) and extremely low in 5.8S (0.6%, Table 4). In most of the samples a poly-A stretch up to 14 nucleotides in ITS1 (alignment pos. 172–188) caused problems in sequencing because of slippage. Thus, internal primers annealing to the respective opposite strands had to be used to complete the sequences. This microsatellite region was excluded from phylogenetic analysis, because such microsatellites normally show a high rate of insertions and deletions leading to uncertain homology assessment and spurious signal in reconstructing phylogenetic relationships of more distant species.

Table 4. Characteristics of ITS and *trnT–trnF* sequences in the *Nymphaea* subg. *Anecphya* dataset (including 4 outgroup taxa of *Nymphaea* subg. *Brachyceras*)

'Mean length' = mean of all observed sequence lengths (s.d. in parentheses); 'Length range' = actual sequence length in nucleotides (minimal and maximal observed value, including hotspots), 'No. of char.' = number of characters in the alignment matrices (excluding hotspots), '%Div.' = percentage pairwise sequence distance (uncorrected p distance, overall mean, s.e. in parentheses); '%GC' = GC content in percent; 'Var.', number of variable characters (percentage in parentheses); 'Inf.' = number of parsimony informative characters (percentage in brackets); 'Indels' = number of length mutations that were coded by SeqState (number of informative indels in parentheses)

Genomic region	Mean length (s.d.)	Length range	Hot spots	No. of char.	%Div. (s.e.)	%GC (s.d.)	Var. (%)	Inf. (%)	Indels (inf.)
ITS (total)	684 (3.4)	681–700	1	695	5.5% (0.5)	49.3 (0.8)	175 (25%)	135 (19%)	35 (23)
ITS 1	280 (3.2)	274–294	1	277	8.0% (1.0)	47.3 (0.9)	92 (33%)	76 (27%)	18 (12)
5.8 S	156 (0.1)	156	0	157	0.6% (0.3)	55.1 (0.2)	11 (7%)	3 (2%)	0 (0)
ITS 2	249 (2.0)	246–254	0	261	5.9% (0.7)	47.9 (1.7)	72 (28%)	56 (21%)	17 (11)
trnT–trnF (total)	1400 (12.7)	1377–1414	2	1464	0.6% (0.1)	35.4 (0.3)	65 (4%)	35 (2%)	14 (7)
trnT–trnL	468 (0.1)	467–468	0	468	0.8% (0.2)	37.0 (0.6)	22 (5%)	14 (3%)	1 (0)
trnL intron	539 (8.9)	526–551	1	568	0.5% (0.1)	35.2 (0.7)	27 (5%)	13 (2%)	6 (4)
trnL–trnF	393 (10.9)	382–409	1	428	0.5% (0.2)	34.0 (0.5)	16 (4%)	8 (2%)	7 (3)

Polymorphic sites in ITS

In several of the examined specimens of *Nymphaea* subg. *Anecphya* there were polymorphic sites, hinting at divergent ITS paralogues. In 29 sequences additive signals, i.e. double peaks in the pherograms (Fig. 2), were present. In 22 of these sequences, polymorphisms occurred at parsimony informative sites and 14 sequences contained more than one polymorphic site (these taxa are listed in Table 5). Polymorphic nucleotide sites were coded with IUPAC ambiguity codes in the sequence alignment (Fig. 2*b*). However, polymorphisms among the paralogues were not restricted to single nucleotide positions. Additionally, length differences were present. Length polymorphism was due to different repeat numbers of an otherwise conserved 'AG'-motif in ITS2 (alignment pos. 503–509). In most individuals, four AG-repeats are present whereas five samples of *N. violacea* (NY131, NY372, NY405, NY436, NY504) possess only three, leading to a 2-bp gap in the alignment. There are some other individuals of *N. violacea* (NY382, NY409, NY424, NY501, NY502; Fig. 3) in which length polymorphisms at the respective position could be observed. Apparently, this is caused by different ITS paralogues possessing either three or four AG-repeats.

Further length polymorphisms were observed in ITS1 of the *N. violacea* samples NY424, NY501, and NY502 (with 2 length-polymorphic sites in the latter). In the sequence of *N. immutabilis* ssp. *kimberleyensis* (NY380), length polymorphisms even occurred in both ITS1 and ITS2. In general, all samples in which length polymorphisms were observed also showed a high frequency of polymorphic nucleotide sites. In the case of sample NY427 (*N. immutabilis*), no length polymorphisms occurred, but the pherograms showed double peaks for 12 parsimony informative sites. To test the effect of polymorphic sites on phylogenetic inference, analyses were run with either the complete taxon sampling (51 accessions) or with datasets reduced by eliminating those sequences with frequent polymorphisms (44 accessions, after excluding NY380, NY382, NY409, NY424, NY427, NY501, and NY502). Whereas the seven above-mentioned individuals exhibited a high number of polymorphisms, there were other individuals with very few (2–4) polymorphic sites. These individuals belonged to *N. gigantea* (NY126, NY395, NY425, NY426, NY435), *N.* cf. *macrosperma* (NY418) and *N. violacea* (NY110). Their inclusion in the matrix also had no significant effect on the reconstructed trees as was elucidated by a matrix further

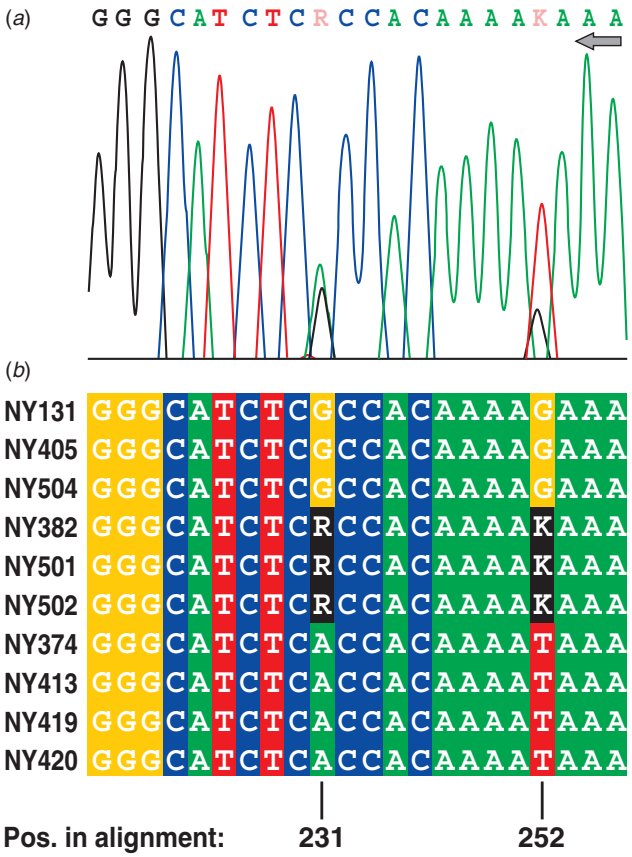

Fig. 2. Polymorphic nucleotide sites in ITS. (*a*) Section of the pherogram of the NY5.8-R primer in accession NY382 (*Nymphaea violacea*). The arrow indicates the reading direction of the primer. Two sites with double peaks are observable. (*b*) Section of the ITS sequence alignment showing the same positions as in (*a*). The respective nucleotide positions (231, 252) are marked with ambiguity codes (R = A or G, K = G or T) for those sequences in which double peaks occur.

Table 5. Taxa with polymorphic nucleotide sites and length polymorphisms in ITS

Only those samples with two ore more polymorphic nucleotides at parsimony informative sites are listed (see text for further explanations). A question mark indicates that no polymorphisms have been observed, but the respective sequence was incomplete because of sequencing problems related to length polymorphisms. The first set of seven samples was excluded from the dataset for the 44-accession analyses

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

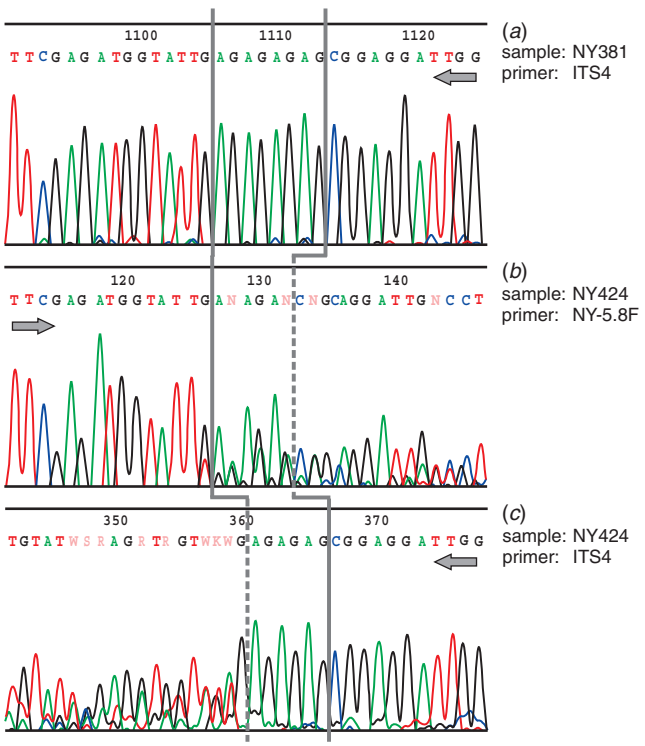

Fig. 3. Length polymorphism observed in ITS2 sequences of several samples. The respective section (an 'AG'-repeat) is marked by vertical lines. Arrows indicate the reading directions of the primers. (*a*) Section of the pherogram of the ITS4 primer in accession NY381 (*N. elleniae*). This is a 'normal' sequence with 4 'AG' units. (*b*, *c*) Pherogram sections of the forward primer NY-5.8F and the reverse primer ITS4 in accession NY424 (*N. violacea*). Breakdown of the signal and overlapping peaks after the respective 'AG'-repeat indicate the presence of templates differing in the length of the AG-repeat.

reduced to 36 accessions that excluded all before-mentioned samples (trees not shown).

Sequence variability of the trnT–trnF *region*

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

Trees obtained from nuclear ITS sequences

Figure 5 shows the strict consensus of 40 trees yielded from maximum parsimony analysis (MP) of the ITS 44-accession dataset. Bayesian analysis (Fig. 6) revealed basically the same topology. There is only a single hardly supported node in each of the trees that is not present in the other tree (marked with an asterisk in both figures). The topology obtained from the 51-accession dataset is less resolved and also much less supported (Table 6). The inclusion or exclusion of indel characters had no effect on the

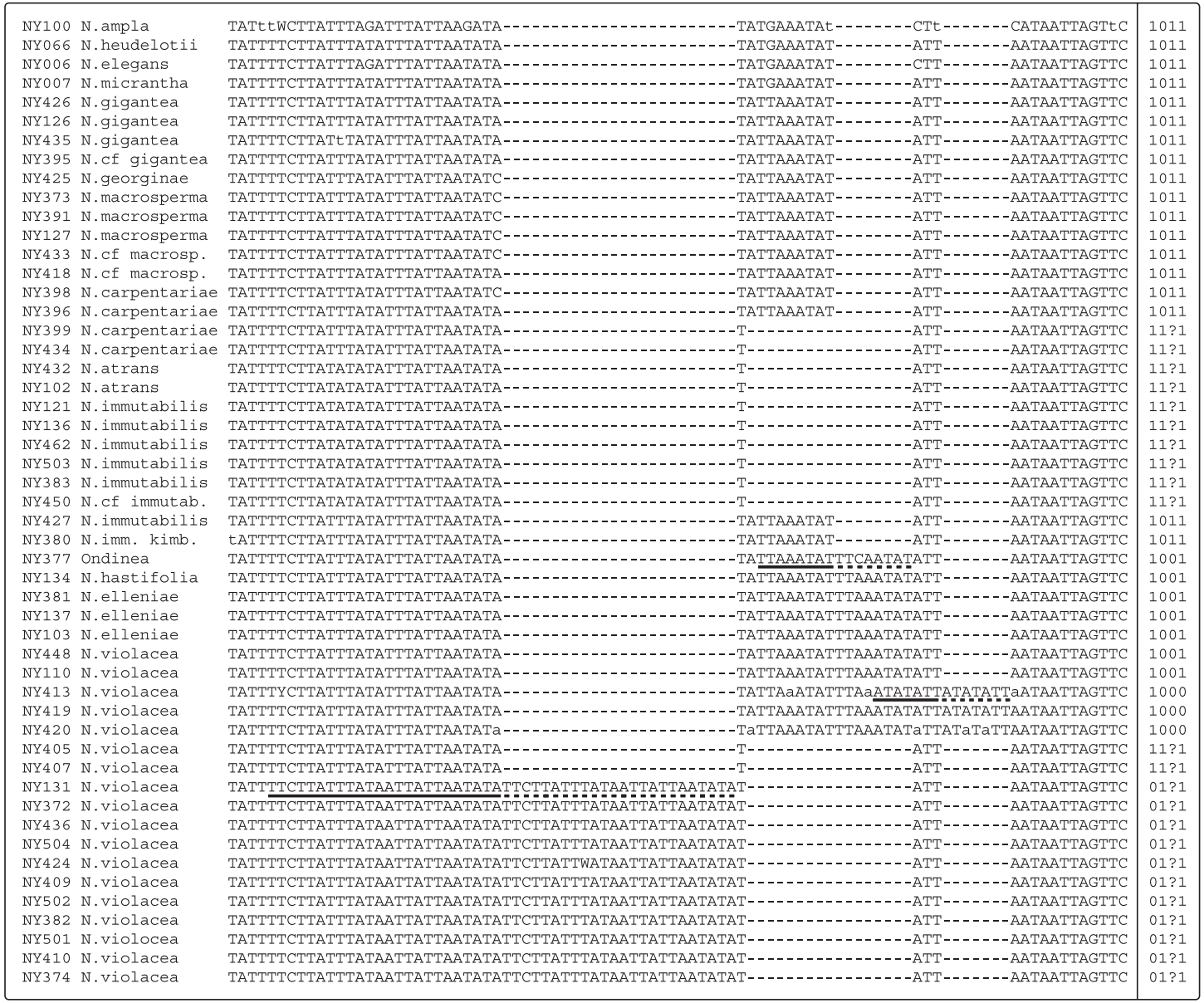

Fig. 4. Section of the *trnT–trnF* matrix, showing a highly variable part of the P8 stem-loop in the *trnL* intron. This section yields four of the seven parsimony informative indel characters of the whole *trnT–trnF* data set. Three of these indels are simple sequence repeats (template and repeat indicated by full and dotted lines, respectively), the other might be a deletion. In the right column of the figure, the binary codes for these length mutations are illustrated (1 = gap present, 0 = no gap).

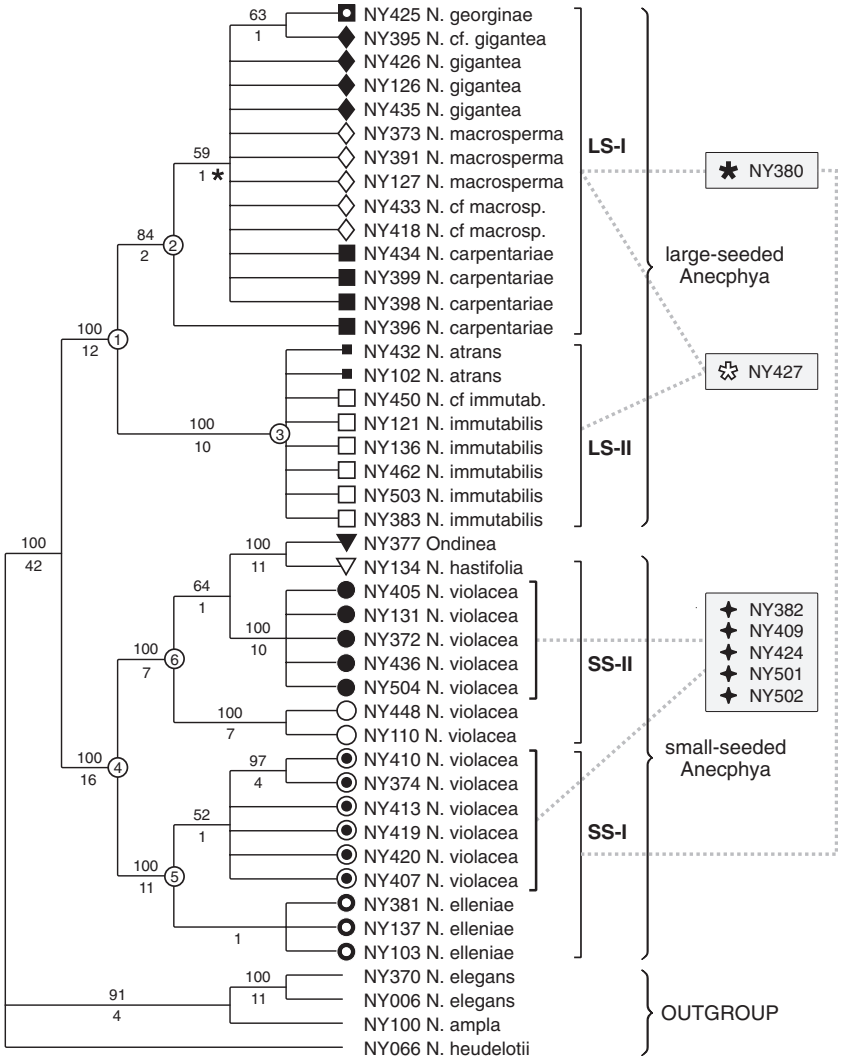

Fig. 5. Strict consensus of 40 shortest trees obtained from maximum parsimony analysis of ITS sequence data from 44 accessions (substitutions + indels). Jackknife values (if higher than 50) are given above the branches. Symbols correspond to the ones used in the *trnT–trnF* tree (Fig. 6) and in the distribution maps (Figs 10 and 11). Nodes relevant for discussion are numbered (in circles, see also Table 6). The asterisk marks the node that is not present in the Bayesian tree. The individuals in grey boxes have not been included in the 44-accession analysis because their sequences contain several nucleotide and length polymorphisms, indicating hybrid origin. Dotted lines show their affinities to subgroups of *Nymphaea* subg. *Anecphya*. See also Figs 2 and 3 and explanations in the text.

inferred topologies but generally the support values are higher if indels are included (Table 6).

Both MP and Bayesian analyses of the ITS dataset reveal a major subdivision of *Nymphaea* subg. *Anecphya* into two clades: one clade (Node 1 in Figs 5 and 6) comprises *N. georginae*, *N. gigantea*, *N. macrosperma*, *N. carpentariae*, *N. atrans* and *N. immutabilis*, and a second clade (Node 4 in Figs 5 and 6) comprises *N. violacea*, *N. elleniae*, *N. hastifolia* as well as *Ondinea purpurea*. These two clades, which are separated by relatively long branches (Fig. 5), correspond to a grouping suggested by morphological characters such as seed size. Therefore, these two groups are in the following referred to as the 'large-seeded' *Anecphya* clade (e.g. *N. macrosperma*, *N. gigantea*) and the 'small-seeded' *Anecphya* clade (e.g. *N. violacea* = subg. *Confluentes sensu* Jacobs, 2007). Among the 'large-seeded' (LS) species, samples further group into two clades, one comprising *N. macrosperma*, *N. gigantea*, *N. georginae* and *N. carpentariae* ('LS-I', node 2 in Figs 5 and 6), and another comprising *N. atrans* and *N. immutabilis* ('LS-II', node 3). The LS-II clade gains maximum support in all trees (irrespective of taxon sampling differences). In contrast, support for the LS-I clade is rather low in both the 51-accession analysis (JK = 52, PP = 0.88, BrS = 1) and the 44-accession analysis (JK = 80, PP = 0.72, BrS = 2, Table 6 and Fig. 5). The LS-I and LS-II groups are not further resolved internally, and ITS sequences do not allow recognition of species within these clades. There is only weak indication that two individuals of *N.* cf. *gigantea* (NY395 and NY425) might be more closely related (JK = 63, PP = 0.95, BrS = 1) to each other than to other

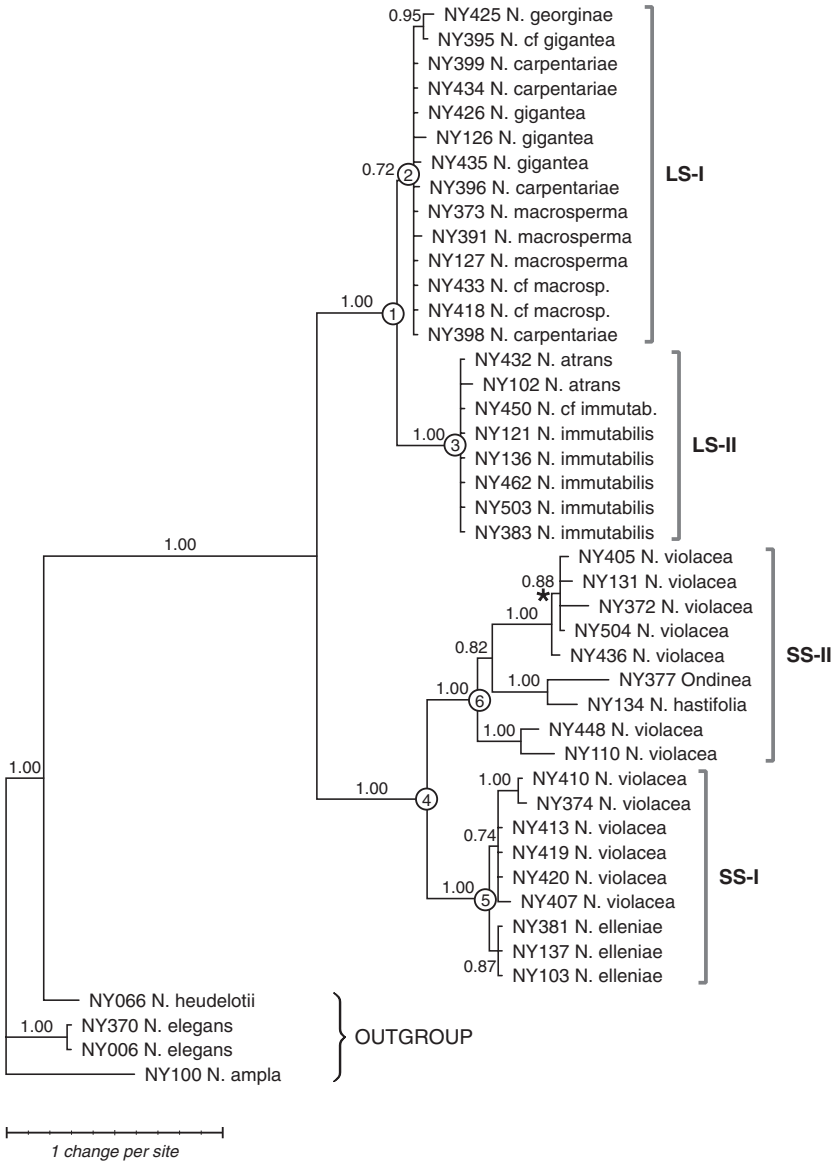

Fig. 6. Phylogram of relationships in *Nymphaea* subg. *Anecphya* inferred from ITS. The tree is a 50% majority rule consensus of 39 603 trees obtained from four runs of Bayesian inference of the 44-accession data set (substitutions + indels). Numbers above branches indicate posterior probabilities of the respective nodes. Nodes relevant for discussion are numbered (in circles, see also Table 6). The asterisk marks the node that is not present in the maximum parsimony tree (see Fig. 4). Apart from this node, both trees are identical. Note that the branch of Ondinea is longer than the branch of its sister taxon, pointing to autapomorphies in *Ondinea*.

plants. Similar to the large-seeded clade, there are also two well supported groups within the small-seeded (SS) lineage of *Anecphya*. Subclade SS-I (node 5, JK = 100, PP = 1.0, BrS = 11) comprises all samples of *N. elleniae* and six of the *N. violacea* samples. Subclade SS-II (node 6, JK = 100, PP = 1.0, BrS = 7) comprises all other *N. violacea* samples as well as *N. hastifolia* and *Ondinea purpurea*. The sister group relationship of the latter two accessions gains high statistical support (JK = 100, PP = 1.0, BrS = 11). Two samples of *N. violacea* (NY110, NY448) appear as a well supported clade, separated from the other *N. violacea* samples in the SS-II group.

Closer examination of sequences and pherograms of those individuals with a high frequency of polymorphic sites (not included in the 44-accession dataset) showed that paralogues present in these individuals are a mixture of sequences otherwise occurring in two different subclades of subg. *Anecphya*. In the case of the five samples of *N. violacea* (NY382, NY409, NY424, NY501 and NY502) the polymorphic sites are composed of character states present in the SS-I and the SS-II group. The *N. immutabilis* sample with polymorphic sites (NY427) shares character states among the LS-I and the LS-II groups, whereas ITS from *N. immutabilis* ssp. *kimberleyensis* (NY380) shows sites characteristic of the LS-I and the SS-I clades.

Table 6. Results of phylogenetic analyses of the ITS dataset by using maximum parsimony (MP) and Bayesian (BI) approaches
Both MP and BI have been conducted for the complete dataset (51 accessions) as well as for the dataset reduced to 44 accessions. The following samples were excluded for the 44-accession dataset: NY382, NY409, NY424, NY501, NY502 (all identified as *N. violacea*) and NY427 (identified as *N. immutabilis*) as well as NY380 (*N. immutabilis* ssp. *kimberleyensis*). See text for explanations. For MP, analyses of substitutions alone (first line) and substitutions and indels combined (second line, italics) were run separately. For BI, only the complete datasets (substitutions + indels) were analysed

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

Trees obtained from the plastid marker trnT–trnF

Bayesian and MP analyses of the *trnT–trnF* sequences were conducted in parallel to ITS, using the complete taxon sampling (51 sequences) and datasets reduced to 44 accessions, respectively. In contrast to ITS, taxon sampling had no effects on the inferred topologies and only minor effects on branch support (Table 7). Thus, only the full evidence trees (51 accessions, substitutions + indels) are discussed in the following, but see Table 7 for statistics of the other reduced-dataset trees. Parsimony and Bayesian analyses yielded congruent topologies (Fig. 7). There is only one node resolved in each of the respective trees that is absent from the other (marked with an asterisk).

The plastid marker did not resolve the large-seeded and small-seeded clades. Instead, four major clades appear in a polytomy. Two of these clades correspond to the large-seeded groups LS-I and LS-II as inferred from ITS. The other two clades contain only small-seeded species, but the composition of the two clades is different from the SS-I and SS-II groups found with ITS. Resolution within the LS-I and LS-II clades is poor, although many of the individuals, even those assigned to the same species, have autapomorphic mutations. This is also evident from the Bayesian phylogram (Fig. 8). In LS-I, most samples of *N. macrosperma* (NY121, NY 373, NY391, NY433) form a clade together with one individual identified as *N. carpentariae* (NY398) and one as *N. georginae* (NY425; JK = 84, PP = 1.0, BrS = 2).

The symbols used in the MP trees of ITS (Fig. 5) and *trnT–trnF* (Fig. 7) clearly illustrate that the nuclear and the chloroplast genomic partitions provide incongruent signal regarding the relationships among the small-seeded species of *Anecphya*. Representatives of the SS-I and SS-II groups (as inferred from ITS) are completely intermingled in the *trnT–trnF* tree (Fig. 7). Chloroplast sequences of *Nymphaea elleniae* (NY103, NY137, NY381), contrary to the ITS tree, group with two samples of *N. violacea* (NY448, NY110). Three samples designated as *N. violacea* (NY413, NY419, NY420) appear as a well supported clade (JK = 95, PP = 1.0, BrS = 3) in the chloroplast tree, which was not resolved with ITS. In the other small-seeded clade, the two *N. violacea* samples NY405 and NY407 are inferred to be sisters, although these samples emerge in different subclades of the ITS tree (NY405: SS-II, NY407: SS-I). Another well supported *trnT–trnF* subclade comprises all remaining SS-II representatives (NY131, NY372, NY436, NY504) as well as two members of the ITS SS-I clade (NY374, NY410) and the five samples of *N. violacea* that were not included in the 44-accession analyses because they were considered as hybrids (NY382, NY409, NY424, NY501, NY502). The other two samples that were not included in the 44-accession analyses, NY380 (*N. immutabilis* ssp. *kimberleyensis*) and NY427 (*N. immutabilis*), appear in the *trnT–trnF* trees in the well supported LS-I clade.

Table 7. Results of phylogenetic analyses of the *trnT–trnF* dataset with maximum parsimony (MP) and Bayesian (BI) approaches
Both MP and BI were conducted for the complete dataset (51 accessions) as well as for the dataset reduced to 44 accessions. The following samples were excluded for the 44-accession dataset: NY382, NY409, NY424, NY501, NY502 (all *N. violacea*) and NY427 (*N. immutabilis*) as well as NY380 (*N. immutabilis* ssp. *kimberleyensis*). In contrast to ITS, the inclusion of sequences with polymorphic sites had no effect on topology and node support. See text for explanations. For MP, analyses of only substitutions (first line) and substitutions + indels (second line, italics) were run separately. For BI, only the complete datasets (substitutions + indels) were analysed

Tree statistics and nodes analysed	MP (51 accessions)	MP (44 accessions)	Bayesian (51 accessions)	Bayesian (44 accessions)
Tree statistics				
No. of characters	1464	1464	*1478*	*1478*
	1478	*1478*		
No. of trees	23	22	39204	39404
	45	*44*		
Tree length	68	66	–	–
	84	*82*		
CI	0.971	0.970	–	–
	0.952	*0.951*		
RC	0.960	0.957	–	–
	0.937	*0.933*		
Jackknife values, Bremer support (in parentheses) for MP analyses, and posterior probabilities for Bayesian analyses Note: nodes 1, 5 and 6 (as inferred from ITS) are not present in any of the *trnT–trnF* trees				
Node 2 (LS-I group, but including NY380, NY427)	87 (2)	87 (3)	1.0	1.0
	94 (3)	*94 (3)*		
Node 3 (LS-II group)	95 (3)	95 (2)	1.0	1.0
	95 (3)	*95 (3)*		
Node 4 (small-seeded *Anecphya* clade)	63 (1)	63 (1)	n.p.	0.62
	n.p.	*n.p.*		
Node 7 (only present in *trnT–trnF*)	63 (1)	64 (1)	1.0	1.0
	79 (1)	*77 (1)*		
Node 8 (only present in *trnT–trnF*)	n.p.	n.p.	0.98	0.93
	75 (1)	74 (1)		

Discussion

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

ITS provides more than three times the number of informative characters than *trnT–trnF* (158 in ITS *v*. 42 in *trnT–trnF*), although the ITS matrix comprises fewer than half the number of characters in the *trnT–trnF* matrix (Table 4). This difference in variability and information content is also reflected in the low mean sequence divergence in the *trnT–trnF* dataset (0.6% compared with 5.5% in ITS, Table 4). Sequence divergence is higher in the *trnT–trnL* spacer than in the *trnL* intron or the *trnL–trnF* spacer, in line with the general patterns in this region (Borsch *et al*. 2003; Shaw *et al*. 2005).

Similar to other Nymphaeales, there is an extended terminal part of the P8 stem-loop region of the *trnL* intron as was shown in secondary structure analyses (Borsch *et al*. 2003; Quandt *et al*. 2004). Long AT-rich sequence parts are thought to have evolved through a stepwise mutation process leading to an independent growth of this region in different land plant groups (Quandt *et al*. 2004). In *Nymphaea*, Borsch *et al*. (2007) recently showed that subgg. *Brachyceras* and *Anecphya* and the genus *Ondinea* have highly similar P8 stem-loop regions. Therefore, the *trnL* sequences could be aligned completely throughout this dataset, allowing the use of the numerous mutations in this AT-rich part. This part entails length mutations of up to 24 nucleotides, even among different samples of *N. violacea* or *N. immutabilis* (Fig. 4). Polytomies in the *trnT–trnF* tree seem to reflect insufficient phylogenetic signal rather than being the result of homoplasy, as evident from high CI and RC values (Tables 6 and 7) in the MP trees.

In the ITS region, 25% of all characters are variable and 19% are parsimony informative. When comparing the different partitions in ITS, it becomes obvious that ITS1 (8% mean sequence divergence, 27% informative characters) is more variable than ITS2 (5.9% mean sequence divergence, 21% informative characters), which is in line with what Woods *et al*. (2005) found in hardy water-lilies. Moreover, it is congruent to the general patterns of variability observed in the ITS region in angiosperms (Baldwin *et al*. 1995; Fuertes Aguilar and Nieto Feliner 2003).

In parallel to the distribution of variable and informative characters, the frequency of polymorphic nucleotide sites in ITS1 is somewhat higher than in ITS2 (20 *v.* 17 polymorphic nucleotides at parsimony informative sites). Only a single polymorphism was observed in the 5.8S rDNA, occurring in the *N. immutabilis* ssp. *kimberleyensis* sample (NY380). Polymorphic sites are extremely valuable for identifying putative hybrids in a given dataset, especially if the two bases involved in a polymorphic site were also found separately in other accessions of the dataset (Sang *et al*. 1995; Fuertes Aguilar *et al*.

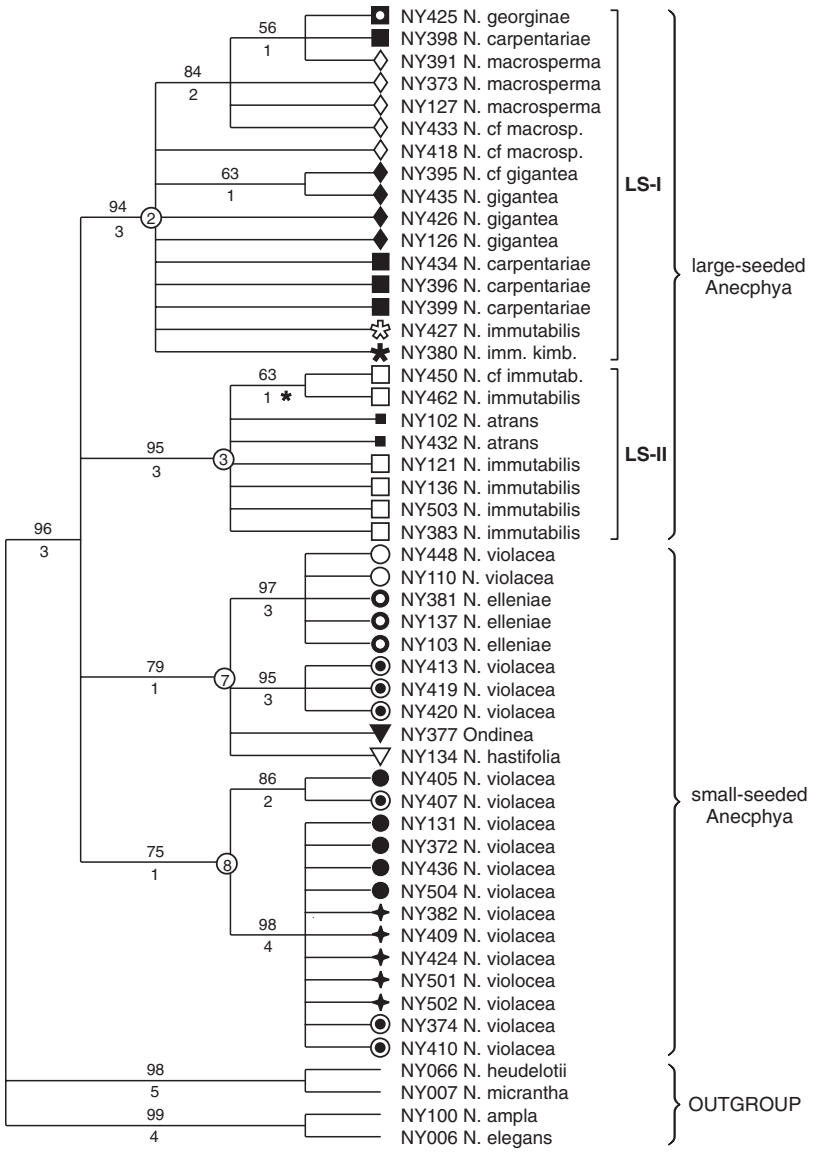

Fig. 7. Strict consensus of 44 shortest trees obtained from maximum parsimony analysis of *trnT–trnF* data from the complete data set (51 accessions, substitutions+indels). Jackknife values are given above the branches. Symbols correspond to the ones used in the ITS tree (Fig. 5) and in the distribution maps (Figs 10 and 11). Nodes relevant for discussion are numbered (in circles, see also Table 7). The asterisk marks the node that is not present in the Bayesian tree (Fig. 7).

1999; Andreasen and Baldwin 2003; Fuertes Aguilar and Nieto Feliner 2003).

Polymorphic sites and incongruent gene trees provide evidence for reticulate evolution

In contrast to the chloroplast genome, the nuclear genome is inherited biparentally. Thus, a successful fertilisation is followed by recombination of the parental alleles. Where sequences of a given locus are divergent among parents, polymorphisms will be observed if the hybridisation event happened rather recently. In ITS, these differences among paralogues are levelled out over time through concerted evolution (Fuertes Aguilar *et al*. 1999). Concerted evolution occurs when all repeat copies of the nuclear rDNA evolve in unison owing to unequal crossing-over during meiosis and conversion (Arnheim 1983; Hillis *et al*. 1991; Elder and Turner 1995). However, there is recent evidence that the tempo and degree of concerted evolution strongly depends on factors specific to the organisms, such as generation time (annual *v.* perennial) and mode of reproduction (sexual *v.* vegetative, e.g. Suh *et al*. 1993; Baldwin *et al*. 1995; Buckler *et al*. 1997). In some cases, divergent repeat types may persist, especially if hybridisation is coupled with an allopolyploidisation event or if the hybrids reproduce only vegetatively (e.g. Campbell *et al*. 1997; Zhang and Sang 1999). As an alternative explanation for the observed polymorphisms in ITS one might think of random

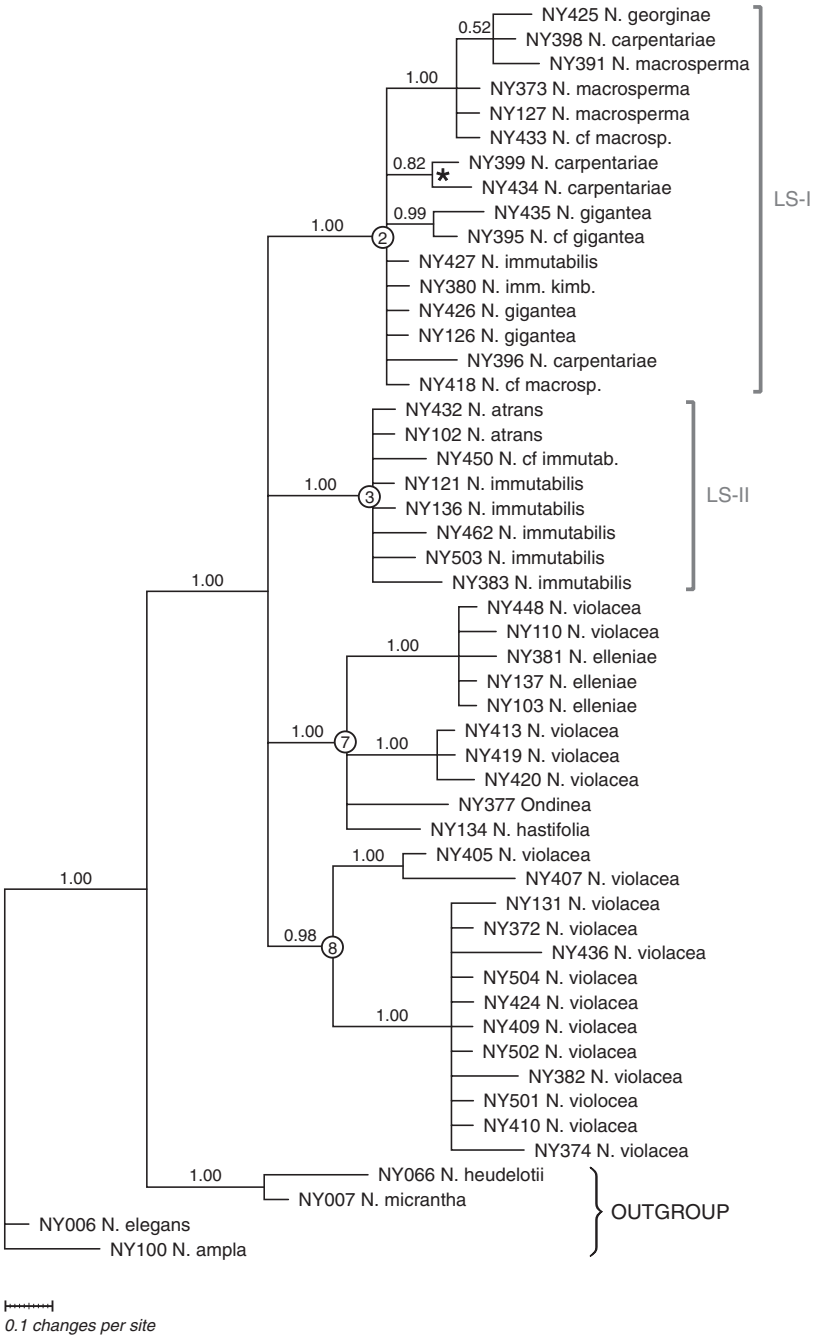

Fig. 8. Phylogram of relationships in *Nymphaea* subg. *Anecphya* inferred from the chloroplast marker *trnT–trnF*. The tree is a 50% majority rule consensus of 39 204 trees obtained from four runs of Bayesian analysis of the 51-accession data set (substitutions + indels). Nodes relevant for discussion are numbered (in circles, see also Table 6). Numbers above branches indicate posterior probabilities of the respective nodes. The asterisk marks the node that is not present in the maximum parsimony tree (Fig. 7). Apart from this node both trees are identical. Note that there is a distinct branch leading to *Ondinea* caused by a number of autapomorphies.

mutations in some of the ITS copies of the same genome that have not been levelled out yet (Baldwin *et al*. 1995; Buckler *et al*. 1997). In the *Anecphya* dataset this alternative explanation seems unlikely because most of the nucleotide polymorphisms and all of the length polymorphisms occur at parsimony informative sites, often being present in several individuals, rather than being distributed at random. Both of the character states occurring at a given polymorphic site often also occur alone in one or more other individuals sequenced and thus suggest hypothetical parents. In the *Anecphya* ITS dataset, 14 samples (Table 5) show between 2 and 12 polymorphisms at parsimony informative sites. All polymorphisms in a given

sample are consistent in suggesting parentage from the same respective species pairs.

The process of hybridisation followed by concerted evolution results in either intermediate sequences, in which characters from both parents have been fixed in all paralogues, or in only one parental sequence dominating the rDNA repeats within a genome. Both versions can have strong, possibly misleading, effects on phylogenetic inference if putative hybrids are included in the dataset (Álvarez and Wendel 2003). Fuertes Aguilar and Nieto Feliner (2003) attributed the large polytomies in their trees of species relationships in the genus *Armeria* to disruptive effects of reticulation and concerted evolution. In the present dataset, the inclusion of putative hybrids also had adverse effects on the trees inferred from ITS. As obvious from Fig. 9 in comparison to Fig. 5, the resolution of the tree is reduced and, more remarkably, the Jackknife support is reduced significantly for the clades LS-I (node 2), SS-I (node 5) and SS-II (node 6). Thus, the exclusion of putative hybrid sequences from phylogeny reconstruction seems justified and is in line with the general approach in other studies (e.g. Sang *et al*. 1995; Fuertes Aguilar and Nieto Feliner 2003; Koontz *et al*. 2004; Marhold *et al*. 2004). If included, all putative hybrids—except NY380 (*N. immutabilis* ssp. *kimberleyensis*)—are depicted in one of the assumed parental clades, which points to ongoing concerted evolution towards one of the parental genomes in the respective taxa (Sang *et al*. 1995). On the other hand, this might also be an indication of introgression of another genome and subsequent backcrossing(s) with one parent (Fuertes Aguilar and Nieto Feliner 2003). Evidence from other sources besides the sequences will be necessary to decide whether ancient hybridisation or introgression is the dominant underlying process. Such evidence could come from chromosome numbers and ploidy levels, geographical distribution, morphology or unlinked gene sequences (Andreasen and Baldwin 2003; Vriesendorp and Bakker 2005).

Phylogenetic relationships and modes of speciation in Nymphaea *subg.* Anecphya

Phylogenetic analysis of the ITS dataset clearly reveals a major subdivision of *Nymphaea* subg. *Anecphya* into a large-seeded (LS) and a small-seeded (SS) group of species (Fig. 5). As mentioned earlier, the LS species share several morphological synapomorphies, such as rather large seeds, toothed leaf margins and a distinctive gap between petals and stamens, whereas the SS group is characterised by relatively small seeds, entire leaf margins and petals grading into stamens (Jacobs 1992, 1994). Thus, molecular data are consistent with morphology regarding the two major clades in *Nymphaea* subg. *Anecphya*.

Despite the fact that both ITS and *trnT–trnF* exhibit phylogenetic structure within *Anecphya*, not all species recognised with morphological characters can be unambiguously differentiated by these two molecular markers. This is particularly obvious in the large-seeded lineage, where *N. atrans* and *N. immutabilis* can be distinguished neither by ITS nor by *trnT–trnF* sequences, and *N. carpentariae*, *N. gigantea* and *N. macrosperma* cannot be distinguished by ITS sequences. A likely explanation is that the evolution of morphological traits in the context of species diversification was relatively fast in relation to sequence divergence. Further sampling of the genomic compartments could show whether low degrees of molecular differences between these species can be detected in an extended sequence dataset. Topological differences in gene trees are mainly restricted to inferred relationships among the small-seeded members of *Anecphya* (Fig. 9). This group comprises the morphologically very variable *N. violacea*, but also the morphologically distinct and geographically restricted species *N. elleniae*, *N. hastifolia* and *Ondinea purpurea*. Like in the large-seeded clade, hybrid ITS sequences also indicate reticulate evolution among small-seeded *Anecphya* (Fig. 5, taxa boxed in grey) but introgressive hybridisation may not fully explain the observed patterns of variability. Incongruence of plastid and nuclear trees could also be caused by incomplete lineage sorting.

Large-seeded Anecphya

Both the nuclear and the chloroplast data revealed a close affinity of *N. immutabilis* and *N. atrans*, which are resolved in a well supported clade in all trees (Node 3 in Figs 5–8, see also Tables 6, 7). Hybrids between these two species with intermediate character states and reduced fertility have been observed in areas where both species grow sympatrically (Jacobs 1992). No indication for hybrids between *N. immutabilis* and *N. atrans* could be observed in the sampled sequences, because there were no parsimony informative sites that would distinguish between the two species. However, one of the samples of *N. immutabilis* (NY427), from the east coast of Queensland (Fig. 10), was designated as a putative hybrid between the LS-II (*N. immutabilis*, *N. atrans*) and the LS-I group (*N. gigantea*, *N. macrosperma*, *N. carpentariae*) on the basis of frequent nucleotide polymorphisms. The maternal parent of NY427 seems to be a member of the LS-I group since the accession NY427 is clearly depicted in this clade in the chloroplast tree (Fig. 7). Also, the ITS sequences show some affinities to sequences of the LS-I clade, pointing to ongoing concerted evolution towards the nuclear genome of the maternal (LS-I) parent or to continuing gene flow through backcrossing with the paternal (LS-II) parent. Considering the distribution of species in both LS groups, NY427 presumably is a hybrid between *N. immutabilis* and *N. gigantea*. Jacobs (1992) reported frequent intergradation between the two species in the areas of overlapping distribution along the Queensland coast. In fact, the population from which the NY427 sample was taken differs from 'normal' *N. immutabilis* in the shape of the leaf margin and the flower colour ('much bluer throughout', C. B. Hellquist, pers. obs.). Another sample of the large-seeded members of *Anecphya* that stands out, through the occurrence of both additive nucleotide polymorphisms and length polymorphisms, is *N. immutabilis* ssp. *kimberleyensis* (NY380) from Western Australia (map in Fig. 10). This subspecies is known only from a single population in a seasonally water-filled lagoon in the Kimberley region, and differs from *N. immutabilis* ssp. *immutabilis* in having considerably fewer stamens (up to 200 *v.* up to 400), glabrous seeds, anthers with a purple gland at the base and blue petals slightly fading with age (Jacobs 1992). The fact that only immature seeds could be found in the field hints on reduced

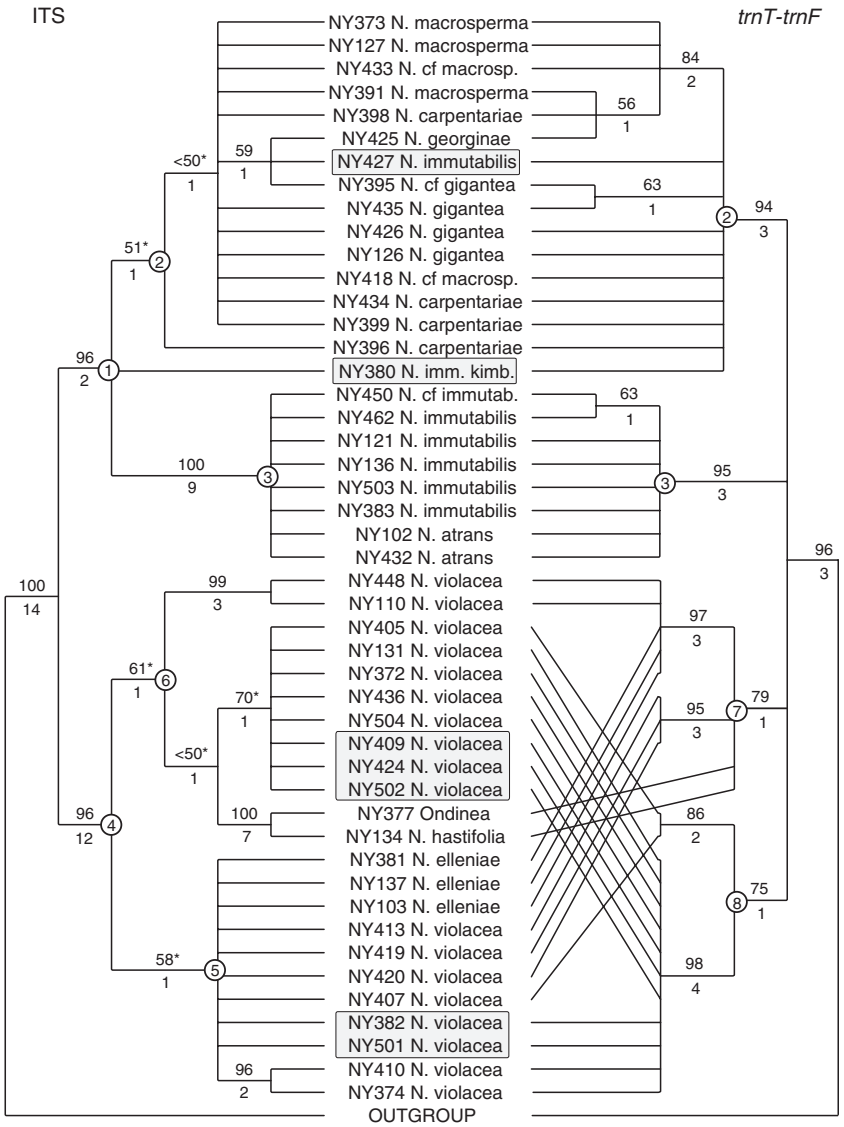

Fig. 9. Incongruence of the ITS tree (left) and the *trnT–trnF* tree (right) for *Nymphaea* subg. *Anecphya*. The figure shows the respective strict consensus trees of maximum parsimony analyses of the 51-accession data sets. Jackknife support for nodes is given above the branches. Note that in ITS support values marked with (*) are significantly higher if potential hybrid sequences (accessions highlighted by grey boxes) are excluded (see Fig. 5 for the 44-accession ITS tree). See text for discussion. Node numbers (in circles) correspond to those in other figures and in Tables 6 and 7.

fertility in this population. The sample NY380 is somehow exceptional in the present dataset because it apparently is the outcome of a hybridisation event between large-seeded (LS-I) and small-seeded (SS-I) *Anecphya* clades. At the polymorphic sites in ITS character states from both the LS-I group and the SS-I group are present. The chloroplast marker clearly depicts NY380 within the LS-I clade (*N. gigantea*, *N. georginae*, *N. macrosperma*, *N. carpentariae*, see Fig. 7), suggesting one of these species as the maternal parent. This is somehow puzzling since from phenology, *N. immutabilis* from the LS-II group would have been expected to be the maternal parent of NY380 (*N. immutabilis* ssp. *kimberleyensis*!). From the species in the SS-I clade (*N. violacea* p.p., *N. elleniae*), only *N. violacea* occurs in Western Australia (see Fig. 11); therefore, it can be considered as the paternal parent of this taxon. However, a careful re-examination of morphological characters and a more detailed analysis with additional molecular markers would be necessary to unravel the history and affinities of *N. immutabilis* ssp. *kimberleyensis*.

Small-seeded Anecphya

Neither in trees obtained from ITS (Figs 5, 6) nor in those from *trnT–trnF* (Figs 7, 8) do the individuals sampled from *N. violacea* form a monophyletic group. Instead, *N. violacea* samples appear in both subclades, SS-I and SS-II, together with either *N. elleniae* (SS-I) or *N. hastifolia* and *Ondinea purpurea* (SS-II). Apparently, at least two different nuclear genomes occur in *N. violacea*, and in

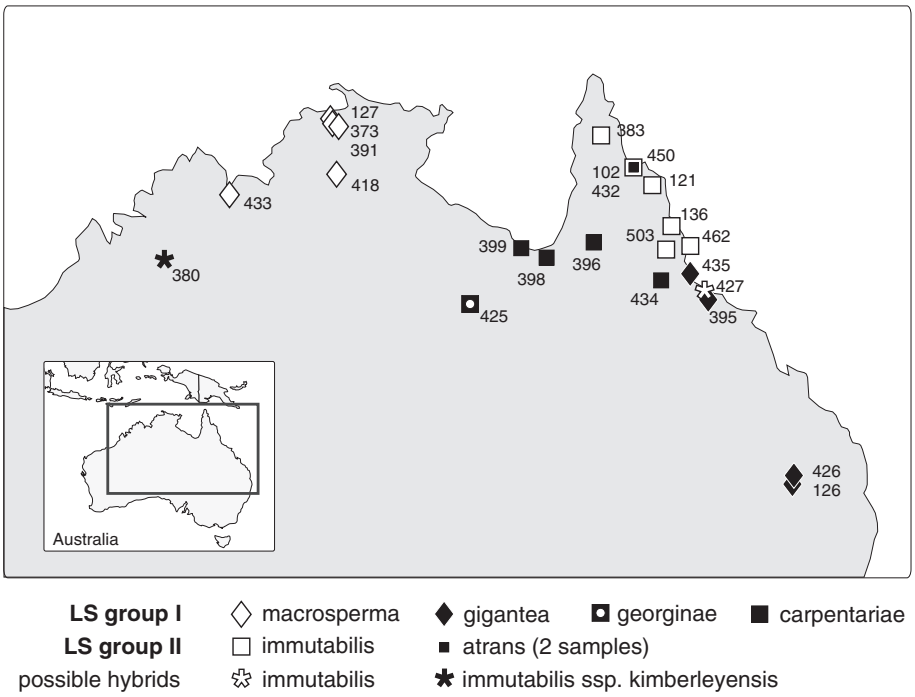

Fig. 10. Field localities of *Nymphaea* subg. *Anecphya* samples of the 'large-seeded group'. Numbers in the figure are the DNA sample numbers given in Table 2. Symbols correspond to those used in the maximum parsimony trees inferred from ITS and *trnT–trnF* (Figs 5 and 7).

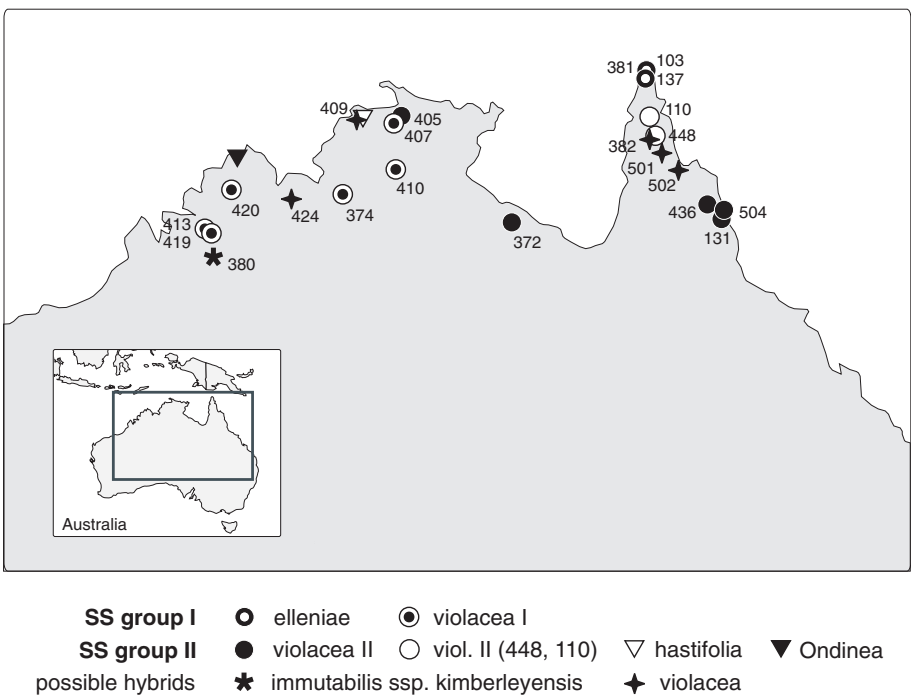

Fig. 11. Field localities of *Nymphaea* subg. *Anecphya* samples of the 'small-seeded group'. Numbers in the figure are the DNA sample numbers given in Table 2. Symbols correspond to those used in the maximum parsimony trees (Figs 5 and 7).

five samples there is evidence for admixture of these different types (marked as possible *violacea*-hybrids with a star in Figs 5, 7 and 11). However, a general conclusion on the relationships within the small-seeded group is not possible, because of the complex pattern observed (see Fig. 9). It appears impossible to decide to which subclade (SS-I or SS-II) the maternal and paternal ancestors of these hybrids belong. If the five putative hybrids are excluded from consideration, the discordant patterns revealed by the nuclear and the plastid marker might still be explained by incomplete lineage sorting (thereby eliminating the necessity for hybridisation or introgression as underlying processes). In that sense the obvious presence of distinct chloroplast haplotypes in different individuals of *N. violacea* could point to an ancient polymorphism. Under that scenario, directed selective speciation in some restricted geographical areas (Fig. 11) could have led to *N. violacea* becoming paraphyletic to *N. elleniae*, *N. hastifolia* and *Ondinea*. If such progenitor-derivative speciation happened rather recently, the time following the speciation event was not sufficient to allow for sorting and extinction of lineages in the ancestral species (Rieseberg and Broulliet 1994), thus retaining a variable paraphyletic *N. violacea*. Species non-monophyly and incomplete lineage sorting have been increasingly detected in flowering plant genera when large numbers of individuals from across species' ranges were sampled with molecular markers (Comes and Abbott 2001; Syring *et al*. 2007). On the basis of network algorithms and lineages-through-time plots, Jakob and Blattner (2006) showed that ancient chloroplast haplotypes survived for a long time within individual *Hordeum* species, potentially even affecting the reconstruction of deep nodes within the genus when incompletely sampled. Since reticulation and incomplete lineage sorting are hard to distinguish, an assessment of the spectrum of haplotypes and their distribution at the population level would allow to construct a more complete chloroplast genealogy and to evaluate the status of the respective populations. Infraspecific molecular data in *Nymphaea* so far are scarce except for *N. odorata*, for which only two *trnL–trnF* types were found in North America (Woods *et al*. 2005). Provided that chloroplast mutational rates are not accelerated in *Anecphya*, the observed variability in *N. violacea* in the present study thus appears rather high. In light of this, the high morphological variability in *N. violacea* in combination with high sequence variability could potentially be attributed to imperfect taxonomy. A critical re-evaluation of phenotypic characters and ploidy levels thus has to be carried out in the small-seeded *Anecphya*. On the other hand, further sequence data from the same individuals are needed to confirm topological differences of plastid and nuclear gene trees as statistical confidence for many nodes is low, in particular for the ITS tree.

However, in view of the biogeographic data some interesting phenomena become obvious. The two *N. violacea* samples, NY448 and NY110, both from nearby populations at the Cape York Peninsula (see map in Fig. 11), apparently share the chloroplast genome with *N. elleniae* (Fig. 7), but emerge distantly from *N. elleniae* in the ITS tree (*N. elleniae* in SS-I, NY448 and NY110 in SS-II; see Fig. 5). A close relationship of these two samples to *N. elleniae* possibly originated from gene flow in one or the other direction, a plausible interpretation in view of the morphological distinctiveness and scent similar to *N. elleniae* apparent especially in NY448 (C. B. Hellquist, unpubl. data). A similar situation can be found in the two *N. violacea* samples NY405 and NY407. Although they possess different nuclear genomes (NY405 in SS-II, NY407 in SS-I; see Fig. 5), they share several apomorphies in the chloroplast DNA (Fig. 7). The close proximity of the respective populations (Fig. 11) suggests ongoing introgression or hybridisation.

Those *N. violacea* samples in which polymorphic sites in ITS were observed (NY382, NY409, NY424, NY501, NY502) all have the same chloroplast haplotype and appear in the *trnT–trnF* tree (Fig. 7) in a clade together with six other *N. violacea* samples (4 from SS-II, 2 from SS-I). Given their scattered geographic distribution and given the noticeable phenological differences between them, they are probably the result of independent events of hybridisation or introgression. However, despite the observed polymorphism in the nuclear rDNA, the term 'hybrid' might not be appropriate for these taxa because the possible parents belong to different lineages of the same species and 'hybridisation' is commonly used to describe interbreeding of members of different species (Vriesendorp and Bakker 2005). Another interesting group of samples of *N. violacea* are the three samples from the Kimberley region (NY413, NY419 and NY420). Morphologically they are quite distinct from the rest of *N. violacea*, as evident from the presence of large stipular sheaths, larger and more pubescent seeds, and striking white flowers with blue- or magenta-tipped petals (C. B. Hellquist, unpubl. data). These samples are depicted within the SS-I group from ITS (Fig. 5), but are resolved as a monophylum in the *trnT–trnF* tree (Fig. 7), supported even by a synapomorphic indel (see Fig. 4, last indel).

The position of Ondinea purpurea

Previous analyses of Nymphaeales that used a large set of cpDNA markers (Löhne *et al*. 2007), and a detailed examination of relationships within the genus *Nymphaea* on the basis of a dense taxon sampling and *trnT–trnF* sequences (Borsch *et al*. 2007), revealed a close affinity of the *Ondinea* plastid sequences to those of the Australian water-lilies *Nymphaea* subg. *Anecphya* and, within this subgenus, to the samples of the small-seeded group. The present study confirms the previous findings for the maternally inherited plastid marker and adds further evidence in support of this hypothesis from nuclear DNA. The ITS tree (see Fig. 5) provides high support for the close affinity of *Ondinea* and the small-seeded *Anecphya* clade. The sequence of *Ondinea* shares several synapomorphies with that of *N. hastifolia* (NY134), which is a member of the SS-II clade. Nevertheless, the exact position of the *N. hastifolia–Ondinea* lineage among the individuals of the variable *N. violacea* is not clarified either with nuclear ITS or with plastid *trnT–trnF* sequences. For this reason, no hybridisation or introgression is evident that might have played a role in the evolution of *Ondinea*. Rather, the uniqueness of its *trnT–trnF* haplotype among the many different *trnT–trnF* sequences of small-seeded *Anecphya* now known seems indicative of its novel origin in a progenitor-derivative speciation process. Further analysis of *N. hastifolia* at the population level will be needed to confirm this. In view of their geographic distribution, a close affinity between *Ondinea* and *N. hastifolia*, however, seems reasonable. Both species grow in the Kimberley region of Western Australia, and both grow in

ephemeral habitats, *N. hastifolia* in temporarily water-filled swamps on a laterite plateau, and *Ondinea* grows in ephemeral creek systems. The presence of glabrous seeds in *Ondinea* might add further evidence for a close affinity to the small-seeded species of *Anecphya*.

The inferred close relationship of *Ondinea* and the small-seeded species of *Anecphya*, especially *Nymphaea hastifolia*, raises a lot of new, interesting questions on the evolutionary history of these taxa. A close relationship of *Ondinea* to the genus *Nymphaea* was proposed long ago on the basis of several shared characters from morphology, anatomy and ecology, e.g. floral anatomy and venation patterns (Williamson and Moseley 1989; Schneider *et al*. 1995), floral organs grading from petals into stamens (Den Hartog 1970; Kenneally and Schneider 1983), the morphology of the gynoecial cup and stigmatic papillae (Schneider 1983; Schneider and Williamson 1993), pollen morphology (Müller 1970), fruit and seed anatomy (Schneider and Ford 1978), as well as similar pollination and seed dispersal syndromes (Schneider 1983). However, so far no one has assumed that *Ondinea* might be more closely related to a specific taxon within the genus *Nymphaea*. Instead, close relationships of *Ondinea* and *Barclaya* have been proposed alternatively (Kenneally and Schneider 1983; Schneider 1983; Jacobs and Porter 2007). The dramatic phenotypic shifts that occurred with the evolution of *Ondinea* (e.g. loss of petals, modification of leaves), and that obviously led to recognition as its own genus, may in fact be caused by mutations in one or a few genes. Being positively selected for, such traits might have been fixed during the adaptation to a new environment, thereby furthering rapid speciation. Similar phenomena have also been observed in other plant genera such as the monotypic *Heterogaura* that was shown to be derived from *Clarkia* (Onagraceae) by molecular data (Sytsma and Gottlieb 1986).

Conclusions and suggestions for further studies

The present phylogenetic analysis of the Australian water-lilies in *Nymphaea* subg. *Anecphya* s.lat. is the first comprehensive study to include a broad taxon sampling of the subgenus and evidence from the nuclear and the chloroplast genome. Molecular data confirm the subdivision of subg. *Anecphya* into two major clades, i.e. subgenus *Anecphya* Conard s.str. and subgenus *Confluentes* (*sensu* Jacobs 2007), which corresponds to groupings based on several morphological characters such as seed size. Among the large-seeded group (= subg. *Anecphya* s.str.), there is one subclade comprising *N. gigantea*, *N. georginae*, *N. macrosperma* and *N. carpentariae* (LS-I) and another subclade containing *N. atrans* and *N. immutabilis* (LS-II). Relationships within the small-seeded group (= subg. *Confluentes sensu* Jacobs 2007) are less clear, since the trees obtained from the chloroplast marker *trnT–trnF* and from the nuclear marker ITS are strongly incongruent. The observed pattern of polymorphic sites and homogenous ITS copies further hints at past and current reticulate evolution.

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

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

References

Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29**, 417–434. doi: 10.1016/S1055-7903(03)00208-2

Andreasen K, Baldwin BG (2003) Nuclear ribosomal DNA sequence polymorphism and hybridization in checker mallows (*Sidalcea*, Malvaceae). *Molecular Phylogenetics and Evolution* **29**, 563–581. doi: 10.1016/S1055-7903(03)00136-2

Arnheim N (1983) Concerted evolution in multigene families. In 'Evolution of genes and proteins'. (Eds M Nei, R Koehn) pp. 38–61. (Sinauer: Sunderland, MA)

Bailey CD, Carr TG, Harris SA, Hughes CE (2003) Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution* **29**, 435–455. doi: 10.1016/j.ympev.2003.08.021

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* **82**, 247–277. doi: 10.2307/2399880

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

Borsch T, Hilu KW, Quandt D, Wilde V, Neinhuis C, Barthlott W (2003) Non-coding plastid *trnT–trnF* sequences reveal a well resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* **16**, 558–576. doi: 10.1046/j.1420-9101.2003.00577.x

Borsch T, Hilu KW, Wiersema JH, Löhne C, Barthlott W, Wilde V (2007) Phylogeny of *Nymphaea* (Nymphaeaceae): evidence from substitutions and microstructural changes of the chloroplast *trnT–F* region. *International Journal of Plant Sciences* **168**, 639–671. doi: 10.1086/513476

Buckler ES, Ippolito A, Holtsford TP (1997) The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* **145**, 821–832.

Campbell CS, Wojciechowski MF, Baldwin BG, Alice LA, Donoghue MJ (1997) Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). *Molecular Biology and Evolution* **14**, 81–90.

Caspary R (1865–1866) Nymphaeaceae. *Annales Musei Botanici Lugduno-Batavi* **2**, 241–253.

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

Comes HP, Abbott RJ (2001) Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (Asteraceae). *Evolution* **55**, 1943–1962.

Conard HS (1905) 'The waterlilies. A monograph of the genus *Nymphaea*.' (Lord Baltimore Press: Baltimore, MD)

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

Domin K (1929) 'Beiträge zur Flora und Pflanzengeographie Australiens.' (Schweizerbart: Stuttgart)

Elder JF Jr, Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *The Quarterly Review of Biology* **70**, 297–320. doi: 10.1086/419073

Fuertes Aguilar J, Nieto Feliner G (2003) Additive polymorphisms and reticulation in an ITS phylogeny of thrifts (*Armeria*, Plumbaginaceae). *Molecular Phylogenetics and Evolution* **28**, 430–447. doi: 10.1016/S1055-7903(02)00301-9

Fuertes Aguilar J, Rosselló JA, Nieto Feliner G (1999) nrDNA concerted evolution in natural and artificial hybrids of *Armeria* (Plumbaginaceae). *Molecular Ecology* **8**, 1341–1346. doi: 10.1046/j.1365-294X.1999.00690.x

Gupta PP (1978) Cytogenetics of aquatic ornamentals. II. Cytology of Nymphaeas. *Cytologia* **43**, 477–484.

Gupta PP (1980) Cytogenetics of aquatic ornamentals. VI. Evolutionary trends and relationships in the genus *Nymphaea*. *Cytologia* **45**, 307–314.

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

Hillis DM, Moritz C, Porter CA, Baker RJ (1991) Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* **251**, 308–310. doi: 10.1126/science.1987647

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

Jacobs SWL (1994) Further notes on *Nymphaea* (Nymphaeaceae) in Australasia. *Telopea* **5**, 703–706.

Jacobs SWL (2007) *Nymphaea* subg. *Confluentes*. In 'Flora of Australia. Vol. 2'. (Ed. AJG Wilson) p. 458. (CSIRO and Australian Biological Resources Study: Melbourne)

Jacobs SWL, Hellquist CB (2006) Three new species of *Nymphaea* (Nymphaeaceae) in Australia. *Telopea* **11**, 155–160.

Jacobs SWL, Porter CI (2007) Nymphaeaceae. In 'Flora of Australia. Vol. 2'. (Ed. AJG Wilson) pp. 259–275. (CSIRO and Australian Biological Resources Study: Melbourne)

Jakob SS, Blattner FR (2006) A chloroplast genealogy of *Hordeum* (Poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Molecular Biology and Evolution* **23**, 1602–1612. doi: 10.1093/molbev/msl018

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

Koontz JA, Soltis PS, Soltis DE (2004) Using phylogeny reconstruction to test hypotheses of hybrid origin in *Delphinium* section *Diedropetala* (Ranunculaceae). *Systematic Botany* **29**, 345–357. doi: 10.1600/036364404774195539

Langlet O, Söderberg E (1927) Über die Chromosomenzahlen einiger Nymphaeaceen. *Acta Horti Bergiani* **9**, 85–104.

Lehmann (1853) Die Gattung *Nymphaea*. *Otto's Hamburger Garten- und Blumen-Zeitung* **9**, 218.

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

Löhne C, Borsch T (2005) Molecular evolution and phylogenetic utility of the *petD* group II intron: a case study in basal angiosperms. *Molecular Biology and Evolution* **22**, 317–332. doi: 10.1093/molbev/msi019

Löhne C, Borsch T, Wiersema JH (2007) Phylogenetic analysis of Nymphaeales using fast-evolving and non-coding chloroplast markers. *Botanical Journal of the Linnean Society* **154**, 141–163. doi: 10.1111/j.1095-8339.2007.00659.x

Marhold K, Lihova J, Perny M, Bleeker W (2004) Comparative ITS and AFLP analysis of diploid *Cardamine* (Brassicaceae) taxa from closely related polyploid complexes. *Annals of Botany* **93**, 507–520. doi: 10.1093/aob/mch073

Merrill ED, Perry LM (1942) Aquatic plants of British New Guinea. *Journal of the Arnold Arboretum. Arnold Arboretum* **23**, 388–389.

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

Müller K (2004) PRAP—computation of Bremer support for large data sets. *Molecular Phylogenetics and Evolution* **31**, 780–782. doi: 10.1016/j.ympev.2003.12.006

Müller K (2005) SeqState: primer design and sequence statistics for phylogenetic DNA datasets. *Applied Bioinformatics* **4**, 65–69. doi: 10.2165/00822942-200504010-00008

Nixon KC (1999) The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* **15**, 407–414. doi: 10.1111/j.1096-0031.1999.tb00277.x

Okada H, Tamura M (1981) Karyomorphological study of the Nymphaeales. *Journal of Japanese Botany* **56**, 367–374.

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818. doi: 10.1093/bioinformatics/14.9.817

Quandt D, Müller K, Stech M, Frahm J-P, Frey W, Hilu KW, Borsch T (2004) Molecular evolution of the chloroplast *trnL–F* region in land plants. In 'Molecular systematics of bryophytes'. (Eds B Goffinet, V Hollowell, R Magill) pp. 13–37. (Missouri Botanical Garden Press: St Louis, MI)

Rieseberg LH, Broulliet L (1994) Are many plant species paraphyletic? *Taxon* **43**, 21–32. doi: 10.2307/1223457

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574. doi: 10.1093/bioinformatics/btg180

Sang T, Crawford DJ, Stuessy TF (1995) Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. *Proceedings of the National Academy of Sciences, USA* **92**, 6813–6817. doi: 10.1073/pnas.92.15.6813

Sauquet H, Doyle JA, Scharaschkin T, Borsch T, Hilu KW, Chatrou LW, Le Thomas A (2003) Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: implications for character evolution. *Botanical Journal of the Linnean Society* **142**, 125–186. doi: 10.1046/j.1095-8339.2003.00171.x

Schneider EL (1983) Gross morphology and floral biology of *Ondinea purpurea* den Hartog. *Australian Journal of Botany* **31**, 371–382. doi: 10.1071/BT9830371

Schneider EL, Ford EG (1978) Morphological studies of the Nymphaeaceae. X. The seed of *Ondinea purpurea* den Hartog. *Bulletin of the Torrey Botanical Club* **105**, 192–200. doi: 10.2307/2484114

Schneider EL, Williamson PS (1993) Nymphaeaceae. In 'The families and genera of vascular plants II'. (Eds K Kubitzki, JG Rohwer, V Bittrich) pp. 486–493. (Springer: Berlin)

Schneider EL, Carlquist S, Beamer K, Kohn A (1995) Vessels in Nymphaeaceae: *Nuphar*, *Nymphaea*, and *Ondinea*. *International Journal of Plant Sciences* **156**, 857–862. doi: 10.1086/297310

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92**, 142–166. doi: 10.3732/ajb.92.1.142

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* **49**, 369–381. doi: 10.1080/10635159950173889

Small RL, Cronn RC, Wendel JF (2004) L.A.S. Johnson Review No. 2. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* **17**, 145–170. doi: 10.1071/SB03015

Suh Y, Thien LB, Reeve HE, Zimmer EA (1993) Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. *American Journal of Botany* **80**, 1042–1055. doi: 10.2307/2445752

Swofford DL (2002) 'PAUP* Phylogenetic analysis using parsimony (*and other methods) v. 4.0b10.' (Sinauer Associates: Sunderland, MA)

Syring J, Farrell K, Businský R, Cronn R, Liston A (2007) Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. *Systematic Biology* **56**, 163–181. doi: 10.1080/10635150701258787

Sytsma KJ, Gottlieb LD (1986) Chloroplast DNA evidence for the origin of the genus *Heterogaura* from a species of *Clarkia* (Onagraceae). *Proceedings of the National Academy of Sciences, USA* **83**, 5554–5557. doi: 10.1073/pnas.83.15.5554

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of 3 noncoding regions of chloroplast DNA. *Plant Molecular Biology* **17**, 1105–1109. doi: 10.1007/BF00037152

Vriesendorp B, Bakker FT (2005) Reconstructing patterns of reticulate evolution in angiosperms: what can we do? *Taxon* **54**, 593–604.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In 'PCR protocols: a guide to methods and applications'. (Eds DGM Innis, J Sninsky, T White) pp. 315–322. (Academic Press: San Diego, CA)

Wiersema JH (1987) 'A monograph of *Nymphaea* subgenus *Hydrocallis* (Nymphaeaceae).' (The American Society of Plant Taxonomists: Ann Arbor, MI)

Williamson PS, Moseley MF (1989) Morphological studies of the Nymphaeaceae sensu lato. XVII. Floral anatomy of *Ondinea purpurea* subspecies *purpurea* (Nymphaeaceae). *American Journal of Botany* **76**, 1779–1794. doi: 10.2307/2444477

Woods K, Hilu KW, Borsch T, Wiersema JH (2005) Pattern of variation and systematics of *Nymphaea odorata*: II. Sequence information from ITS and *trnL–trnF*. *Systematic Botany* **30**, 481–493. doi: 10.1600/0363644054782143

Zhang D, Sang T (1999) Physical mapping of ribosomal RNA genes in peonies (*Paeonia*, Paeoniaceae) by fluorescent in situ hybridization: implications for phylogeny and concerted evolution. *American Journal of Botany* **86**, 735–740. doi: 10.2307/2656583

Manuscript received 5 March 2007, accepted 29 July 2008