

## Out of Africa: molecular phylogenetics and biogeography of *Wolffiella* (Lemnaceae)

REBECCA T. KIMBALL<sup>1\*</sup>, DANIEL J. CRAWFORD<sup>2</sup>, DONALD H. LES<sup>3</sup> and ELIAS LANDOLT<sup>4</sup>

<sup>1</sup>Department of Zoology, University of Florida, PO Box 118525, Gainesville, Florida 32611, USA

<sup>2</sup>Department of Ecology and Evolutionary Biology & The Museum of Natural History and Biodiversity Center, University of Kansas, Lawrence, Kansas 66045, USA

<sup>3</sup>Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269–3043, USA

<sup>4</sup>Geobotanisches Institut ETH, Zürichbergstrasse 38, CH-8044, Zürich, Switzerland

Received 5 August 2002; accepted for publication 29 January 2003

The monophyletic genus *Wolffiella* (Lemnaceae) comprises 10 species divided taxonomically into three sections. Relative to other genera of Lemnaceae, *Wolffiella* has a restricted range, with species distributed in warm temperate to tropical areas of Africa and the Americas, with only one species occurring in both areas. Sequence data from coding (*rbcL* and *matK*) and non-coding (*trnK* and *rpl16* introns) regions of cpDNA were analyzed phylogenetically to resolve relationships within *Wolffiella*, and these results were compared to earlier allozyme and morphological studies. Allozymes, cpDNA and morphology all supported the recognition of three sections. Relationships among species were similar in most respects between the allozyme and cpDNA trees, as well as among the different plastid partitions. In *Wolffiella*, both non-synonymous and synonymous substitutions were greater in *matK* than in *rbcL*, as observed in other taxa. The synonymous substitution rate in *matK* was similar to the substitution rate of the non-coding regions. All partitions, including coding regions, exhibited some homoplasy. Biogeographical reconstructions from a combination of cpDNA partitions indicated that *Wolffiella* originated in Africa with early movement to and radiation in the Americas. The one species found in both Africa and the Americas, *W. welwitschii*, likely originated in the Americas and subsequently dispersed to Africa. Using the SOWH test, the cpDNA data could reject two alternative biogeographical hypotheses suggested from analyses of morphological and allozyme data. The present distribution of *Wolffiella* can be explained by two major dispersal events and this contrasts with the more complex species distributions in other Lemnaceae genera. Limited dispersal in *Wolffiella* relative to other Lemnaceae genera may be due to more recent origins of species, lower dispersibility and poorer colonizing ability. © 2003 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2003, 79, 565–576.

ADDITIONAL KEYWORDS: allozymes – duckweeds – *matK* – *rbcL* – *rpl16*.

### INTRODUCTION

The duckweed family (Lemnaceae) includes the smallest of all flowering plants. These cosmopolitan aquatic monocots are reduced to tiny thalloid fronds that float on or below the surface of the water (Landolt, 1986). Although all species produce flowers, the frequency of flowering differs among the species (Landolt, 1986),

and reproduction in the family is thought to be primarily vegetative. *Wolffiella* is unique biogeographically among genera of Lemnaceae in being restricted to warm temperate, subtropical and tropical areas in the Americas and Africa, with a relatively recent introduction of one species (*W. hyalina*) into India (Landolt, 1986). In contrast, other Lemnaceae genera have one or more species distributed widely in temperate zones (Landolt, 1986). The biogeographical origin of *Wolffiella* is uncertain. Allozyme analyses indicated an African origin (Crawford *et al.*, 1997), although

\*Corresponding author. E-mail: rkimball@zoo.ufl.edu

Landolt (1986) concluded that the origins were likely in South America. In addition, it is not known whether there have been one or multiple dispersal events between the Old and New Worlds.

The family as currently recognized consists of 37 species in five genera: *Landoltia* (one species), *Lemna* (13 species), *Spirodela* (two species), *Wolffia* (11 species) and *Wolffiella* (ten species) (Table 1; Les & Crawford, 1999; Landolt, 2000; Les *et al.*, 2002). The taxonomic history of *Wolffiella*, as reviewed by Landolt (1986), has been intertwined with that of *Wolffia*, the other genus of the subfamily Wolffioideae. Landolt (1986) initially circumscribed *Wolffiella* as comprising

nine species in three sections, with an additional species being described later (Landolt, 1992). Phylogenetic analyses of morphological and anatomical characters (Les, Landolt & Crawford, 1997b) indicated monophyly of the sections in *Wolffiella* but paraphyly of the genus. However, a tree five steps longer could resolve *Wolffiella* and the other genera as being monophyletic. A recent molecular phylogenetic study (Les *et al.*, 2002) provided strong support for both the monophyly of the sections in *Wolffiella* and the monophyly of the genus (amended to include *W. caudata* as recognized by Landolt, 1992).

In this study, we examine phylogenetic relation-

**Table 1.** Accessions used in analyses

Species	<i>rbcL</i>	<i>matK, trnK</i>	<i>rpl16</i>
<i>Wolffiella</i>			
Section <i>Wolffiella</i>			
<i>W. caudata</i> Landolt	9158 (Bolivia)	9173 (Bolivia)	9158* 9214 (Bolivia)
<i>W. denticulata</i> (Hegelm.) Hegelm.	8221 (S. Africa)	8221	8221
<i>W. gladiata</i> (Hegelm.) Hegelm.	8261 (USA)	8261	7173 (USA) 8768 (USA)*
<i>W. lingulata</i> (Hegelm.) Hegelm.	7289 (Brazil)	7289	7289* 7655 (Mexico)
<i>W. neotropica</i> Landolt	8848 (Brazil)	8848	7290 (Brazil) 8848*
<i>W. oblonga</i> (Phil.) Hegelm.	8984 (Columbia)	8984	7997 (Brazil)* 8072 (USA) 8393 (USA) 8984
<i>W. welwitschii</i> (Hegelm.) Monod	7468 (Columbia)	7468	7468* 9096 (Zimbabwe)
Section <i>Stipitatae</i>			
<i>W. hyalina</i> (Del.) Monod	8640 (Tanzania)	8640	7376 (Egypt) 8640*
<i>W. repanda</i> (Hegelm.) Monod	9122 (Zimbabwe)	9122	9054 (Zimbabwe) 9062 (Zimbabwe)* 9104 (Botswana) 9107 (Botswana) 9122
Section <i>Rotundae</i>			
<i>W. rotunda</i> Landolt	9121 (Zimbabwe)	9121	9072 (Zimbabwe)* 9121
<i>Wolffia</i>			
<i>W. australiana</i> (Benth.) Hartog & Plas	7733 (Australia)	7733	7631 (Australia)
<i>W. borealis</i> (Englem.) Landolt	9123 (USA)	9123	9123
<i>W. brasiliensis</i> Wedd.	8743 (Argentina)	8743	9134 (Brazil)
<i>W. microscopica</i> (Griff.) Kurz	8359 (India)	8359	8359

Numbers are those of E.L. with vouchers in ZT. Geographic locality is indicated in parentheses for the first listing for each accession. \*Accession used in combined analyses (and in Les *et al.*, 2002).

ships in *Wolffiella* using cpDNA sequence data from all species. For most species, multiple, geographically divergent accessions were included. The phylogenetic trees of *Wolffiella* obtained from different cpDNA partitions were evaluated for congruence. Aspects of molecular evolution were evaluated for each data partition in order to better understand and to interpret instances of incongruence. A well-supported phylogeny for the genus is presented and is used to reconstruct the biogeographical history of *Wolffiella*. Alternative biogeographical hypotheses from earlier allozyme and morphological studies are then examined to better clarify the evolution of this unusual Lemnaceae genus.

## MATERIAL AND METHODS

### DNA AMPLIFICATION AND SEQUENCING

Four regions of the chloroplast genome were used, including two protein coding loci (*rbcL*, *matK*) and two intron regions (*rpl16*, *trnK* – combining both the 5′- and the 3′ regions). Previously published data were supplemented by sequencing additional accessions for the *rpl16* intron. PCR amplification of the *rpl16* intron was performed using primers F71 (Jordan, Courtney & Neigel, 1996) and R622 (Les *et al.*, 2002). PCR reactions were carried out using standard protocols, and products were cleaned using either QIAquick PCR purification columns (Qiagen, Inc., Valencia, CA) or by precipitation using an equal volume of PEG/NaCl (20%/2.5 M). Sequencing of the *rpl16* intron was conducted using the amplification primers. Cycle sequencing reactions ( $\frac{1}{4}$  or  $\frac{1}{2}$  volumes) were performed using the BigDye Terminator kit (PE Applied Biosystems, Foster City, CA) and by following the standard protocol provided for the ABI Prism 310 automated sequencer (PE Applied Biosystems, Foster City, CA). Sequence chromatographs were edited manually and assembled into double-stranded contigs.

Sequences were aligned initially in Clustal W (Thompson, Higgins & Gibson, 1994), then manually optimized by visual inspection. Alignments were trimmed to exclude highly variable regions (e.g. near exon 1 of *rpl16*) where positional homology was difficult to establish. GenBank accession numbers for all molecular data, including our previously published (Les *et al.*, 2002) and new sequences are: AY034200-AY034209 (*rbcL*), AY034316-AY034325 (*trnK* 3′ intron), AY034355-AY034364 (*trnK* 5′ intron), AY034200-AY034209 (*matK*), AY034277-AY034286 and AY131184-AY131197 (*rpl16* intron).

We used DNA sequence data from each species of *Wolffiella*; four species in the genus *Wolffia* (*W. australiana*, *W. borealis*, *W. brasiliensis* and *W. microscopica*;

Table 1) were included as outgroups. For the *rpl16* intron, two or more independent accessions were sequenced for each species except *Wolffiella denticulata*; when possible, additional accessions were sampled from geographically distinct regions (Table 1).

### PHYLOGENETIC ANALYSES

Phylogenetic analyses were performed using default values in PAUP\* 4.0b8 (Swofford, 1999) unless noted otherwise. To obtain the most parsimonious (MP) tree using equally weighted parsimony, a heuristic search was performed with 100 random sequence additions and tree bisection–reconnection (TBR) branch swapping. The reliability of specific taxon groupings under parsimony was examined using 1000 bootstrap replicates with ten random sequence additions per replicate. For parsimony analyses, data were analyzed both by treating indels as missing data and by including indels as characters following the ‘simple indel coding’ method of Simmons & Ochoterena (2000).

To determine the appropriate evolutionary models for maximum likelihood (ML) analyses we used the hierarchical likelihood ratio test as implemented in MODELTEST 3.04 (Posada & Crandall, 1998). Parameters used in ML analyses were those recommended by MODELTEST. To compare parameters across data partitions, the transition/transversion ratio and the shape parameter ( $\alpha$ ) of a gamma distribution were also estimated for all four partitions using HKY85 +  $\Gamma$  (Hasegawa–Kishino–Yano model with gamma distributed rates) and the MP topology obtained from the combined dataset.

To determine whether the different plastid partitions were concordant, we performed the partition homogeneity test (incongruence length difference test (ILD); Farris *et al.*, 1995). Tests were performed using only the informative sites, with 1000 replicates and 10 random sequence additions per replicate.

To estimate the degree of sequence divergence among the four partitions (*rpl16*, *trnK*, *matK* and *rbcL*), the sums of the branch lengths from the ML trees for each partition were obtained. In addition, *p*-distances were estimated for the two non-coding partitions. Non-synonymous and synonymous *p*-distances were estimated using the method of Nei & Gojobori (1986), as implemented in MEGA 1.02 (Kumar, Tamura & Nei, 1993). Corrected non-synonymous and synonymous distances were estimated using the method of Yang & Nielsen (2000), as implemented in PAML 3.12 (Yang, 2002).

### TESTS OF BIOGEOGRAPHICAL HYPOTHESES

We used the SOWH test (Swofford *et al.*, 1996; Goldman, Anderson & Rodrigo, 2000) to examine

specific biogeographical hypotheses for *Wolffiella*. This approach compared a test statistic,  $2\delta$  (two times the difference in likelihood values), for the ML tree estimated from cpDNA data with that of an alternative topology. The two alternative topologies (representing alternative biogeographical hypotheses) were those indicated by analyses of either morphology (Landolt, 1986, 1992) or allozymes (Crawford *et al.*, 1997). To determine whether an alternative topology was less likely statistically than was the cpDNA ML topology, a null distribution of the test statistic was generated using 500 simulated data sets. For each topology tested, we simulated data sets based upon the complete cpDNA data set using Seq-Gen 1.1 (Rambaut & Grassly, 1997). Parameters and branch length information were estimated using the alternative ML topology being tested in that specific SOWH test. For each of the 500 simulated data sets, a heuristic search was used to find the ML tree (as performed on the raw data above). Parameter estimates for each simulated data set were estimated using ML and the topology of the alternative tree being considered, as recommended by Goldman *et al.* (2000). The  $2\delta$  test statistic generated from the simulated data sets was used to establish the null distribution for this statistic. We rejected the null hypothesis (that the ML and the alternative topology did not differ significantly) if fewer than 5% of the simulated data sets had  $2\delta$  values greater than the observed  $2\delta$  value.

## RESULTS

### MOLECULAR EVOLUTION

The variability of the partitions differed, with the two non-coding partitions having similar levels of variability and being only slightly more variable than *matK* (Table 2). By summing the ML branch lengths across

the most likely tree for each partition (Table 2), which gives a divergence rate corrected for homoplasy, the *trnK* and *rpl16* introns appeared to be diverging most rapidly, with *rbcL* diverging much more slowly.

To compare the rate of divergence at sites that should not be under selection, we compared *p*-distances among the synonymous sites in *matK* and *rbcL* with the non-coding sequences of *rpl16* and *trnK*. Synonymous sites in *matK* have evolved the most rapidly, accumulating substitutions at 2.0, 1.6 and 1.5 times the rate of *rbcL*, *trnK* and *rpl16*, respectively. Non-synonymous substitutions have also accumulated much more rapidly in *matK* than in *rbcL*, occurring about 4.8 times faster than the rate in *rbcL*. The partitions also differed in the ratio of transitions to transversions and the shape parameter of a gamma distribution (Table 2), indicating different patterns of molecular evolution. In general, *matK*, *trnK* and *rpl16* were the most similar, while *rbcL* was generally quite different.

### PHYLOGENETIC RELATIONSHIPS

Analyses including indels as characters provided similar topologies for each partition (data not shown), with little or no improvement in resolution and consistency compared with analyses treating gaps as missing data (Table 3). Therefore, we elected to show our results with gaps treated as missing data.

The ILD test indicated that the four cpDNA partitions were not significantly incongruent ( $P = 0.11$ ). Some of the observed incongruence was distributed among the outgroup taxa, and their removal increased congruence of the different plastid partitions substantially ( $P = 0.40$ ). Because significant incongruence was not observed, we analyzed the data partitions separately and in combination.

The phylogenetic tree estimated from all four parti-

**Table 2.** Parameters of different molecular data partitions

Parameter	<i>rbcL</i>	<i>matK</i>	<i>trnK</i>	<i>rpl16</i>
No. of sites	1348	1548	1017	460
% variable sites	5.8	15.9	17.1	17.6
% informative sites	2.9	6.7	8.4	8.7
Sum ML branches	0.08707	0.21412	0.26820	0.29567
Best model*	HKY85 + $\Gamma$ + I	F81 + $\Gamma$	F81 + $\Gamma$	F81 + $\Gamma$
Shape parameter ( $\alpha$ )†	0.01	0.34	0.23	0.21
ti/tv†	0.94	0.59	0.56	0.45
No. of gap characters	–	4	49	20

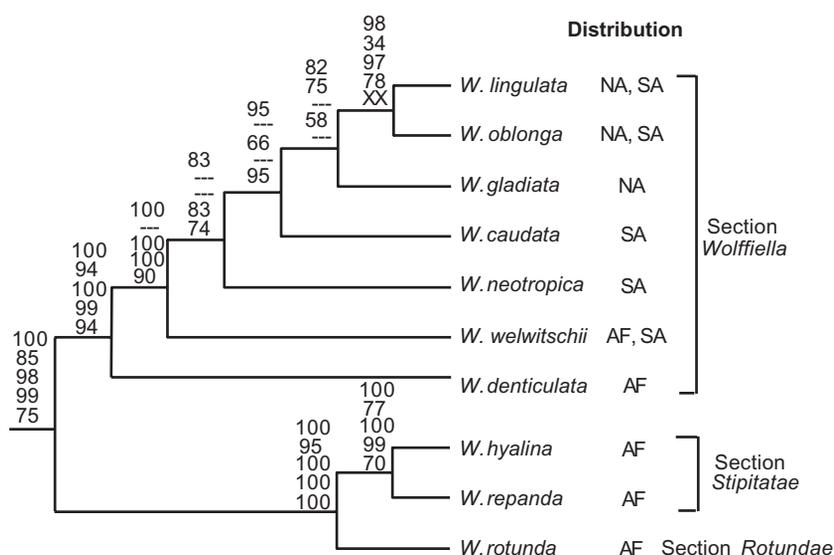
\*Determined using MODELTEST 3.04.  $\Gamma$  = incorporates site-to-site rate heterogeneity using a gamma distribution, I = incorporates invariant sites.

†Estimated using HKY85 +  $\Gamma$  and the topology of the most parsimonious tree from the combined analysis. ti/tv = transition/transversion ratio.

**Table 3.** Results of analysis on different molecular data partitions

Parameter	<i>rbcL</i>	<i>matK</i>	<i>trnK</i>	<i>rpl16</i>
Gaps as missing data				
No. of MP trees	5	3	3	1
CI, excluding uninformative	0.65	0.72	0.75	0.77
With gap matrix				
No. of MP trees	–	3	1	4
CI, excluding uninformative	–	0.72	0.70	0.76
With rapidly evolving sites removed				
No. of MP trees	1	3	6	1
CI, excluding uninformative	0.75	0.74	0.80	0.78

MP = most parsimonious.

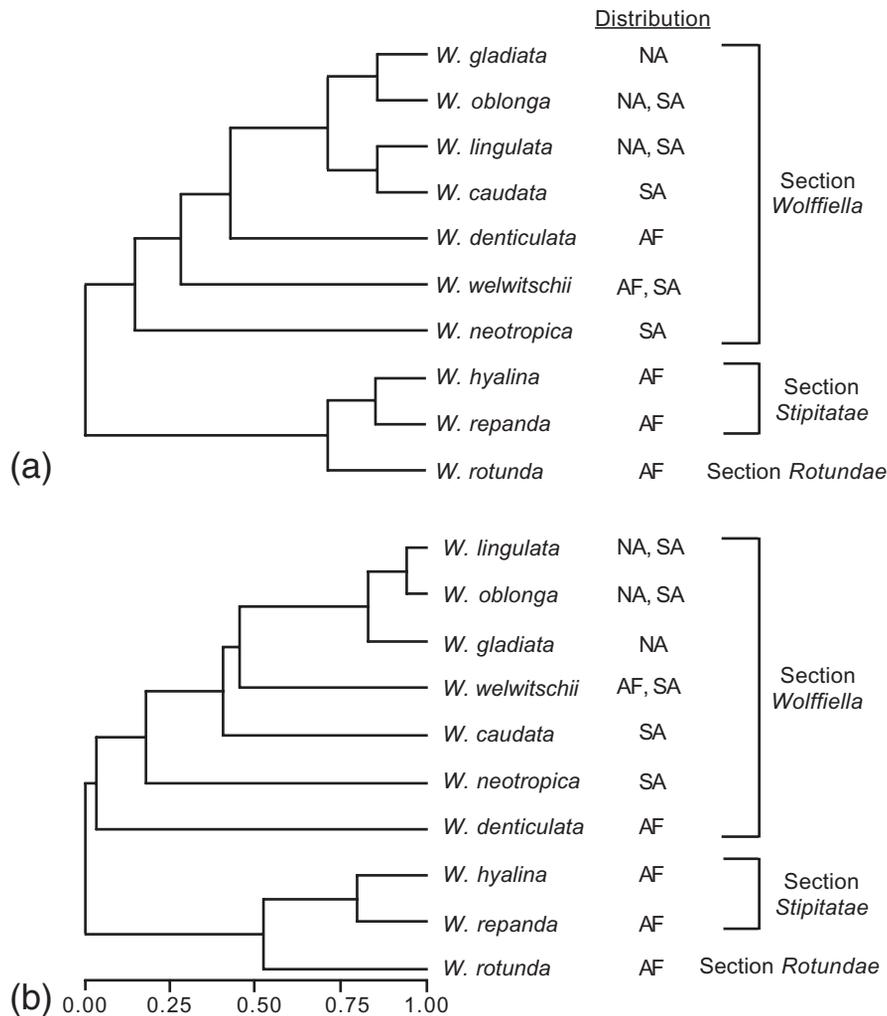


**Figure 1.** Most parsimonious tree using the combined sequence partitions, rooted with *Wolffia* (outgroup taxa not shown). Values at nodes represent percent of 1000 parsimony bootstrap replicates. Top value is from analysis of the combined sequence partitions, followed by analysis of *rbcL*, *matK*, *trnK* intron and *rpl16* intron. Dashes indicate that the node was not found, but that alternative placements of those taxa were supported by fewer than 50% of bootstrap replicates; XX indicates that the node was not found, and that alternative placements of the taxa were supported by more than 50% of bootstrap replicates. Distribution: AF = Africa, NA = North America, SA = South America.

tions combined gave a single well-resolved MP tree (CI excluding uninformative sites = 0.72) with greater than 70% bootstrap support at all nodes (Fig. 1). There was strong support for the monophyly of the three sections recognized by Landolt (1986) based on morphology and anatomy, and also by Crawford *et al.* (1997) based on allozyme analyses (Fig. 2). However, cpDNA sequence data indicated slightly different relationships within Section *Wolffiella* than those suggested by previous studies (cf. Figs 1, 2; see also Les *et al.*, 2002).

When the different cpDNA sequence partitions were

analysed independently, slightly different topologies were obtained. These differences were primarily due to a lack of resolution among particular nodes (*matK*, *trnK* and one relationship in *rpl16*). In *rbcL*, *W. caudata* was placed basal to *W. denticulata*, though with less than 50% bootstrap support. Only in *rpl16* was a topological difference supported, and this involved 60% bootstrap support for a clade containing *W. caudata* and *W. oblonga*. However, within each partition, there was no conflict between the strict consensus of all MP trees and the ML tree. Therefore, only the parsimony results are shown. Trees constructed



**Figure 2.** Previous hypotheses of relationships in *Wolffia*. (a) Redrawn from Landolt (1986) with the placement of *W. caudata* based on comments of Landolt (1992); (b) Phenogram (UPGMA) based on genetic identity at allozyme loci. Scale across bottom indicates genetic identity. Redrawn from Crawford *et al.* (1997). Designations of geographical distributions are the same as in Fig. 1.

from individual partitions had few conflicts with the tree estimated from combined data (Fig. 1). The phylogeny estimated from *rbcL* differed most from the combined data tree (Fig. 1) in that *W. caudata* rather than *W. denticulata* was basal within Section *Wolffiella*. One other difference found in analyses of *rpl16* data was that one accession of *W. oblonga* formed a clade with *W. caudata* rather than with *W. lingulata* (Fig. 1). However, neither of these conflicting nodes was supported by a bootstrap value greater than 50% (Fig. 1). Interestingly, the CIs of the non-coding regions were higher than those for either coding region (Table 3).

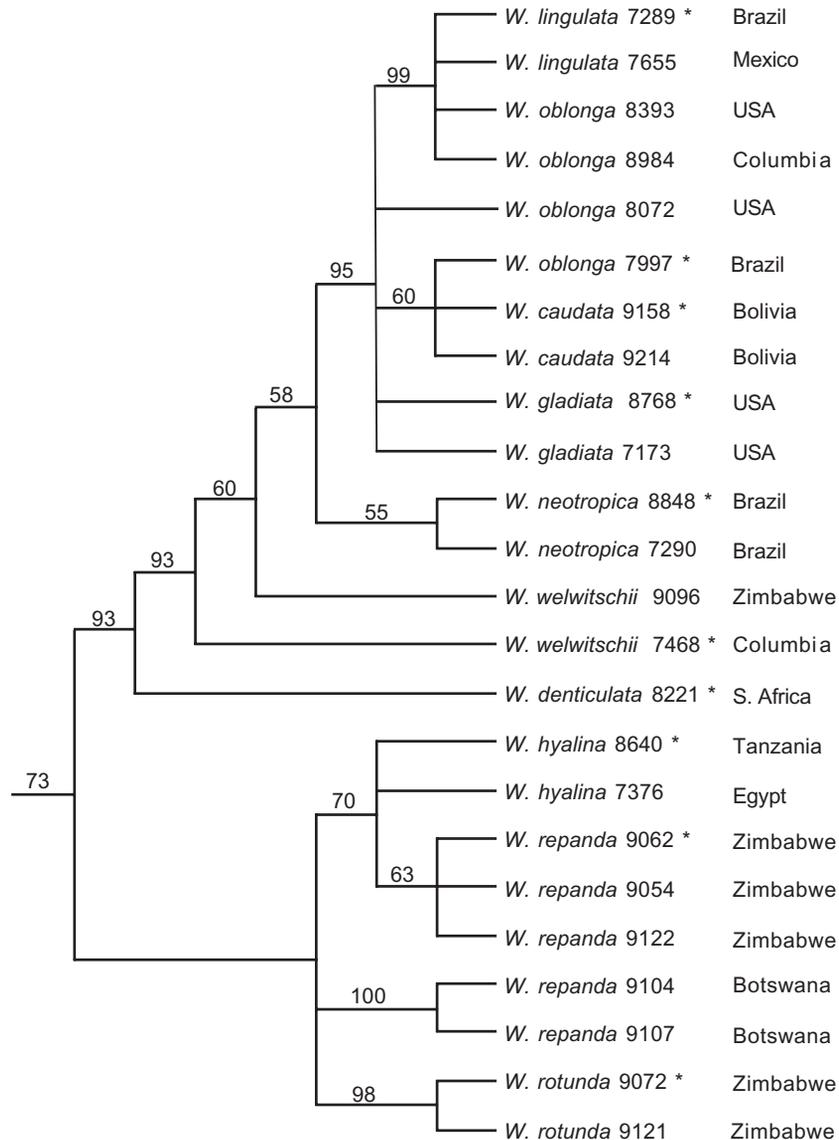
When comparing the number of parsimony steps across the best tree (obtained using all four partitions

combined), a small proportion of sites in each partition appeared to be evolving rapidly, accumulating three to four steps across the tree. Although the proportion of rapidly evolving sites was small, there was variation among partitions (1.08% *rpl16*, 1.28% *trnK*, 0.32% *matK*, 0.52% *rbcL*). These sites were not confined to third positions in the two coding partitions, as might have been expected. In particular, the seven rapidly evolving sites for *rbcL* included both a first and a second position, resulting in non-synonymous substitutions in each case. Since rapidly evolving sites may exhibit high levels of homoplasy, we reanalyzed the data after removal of these sites. With the rapidly evolving sites removed, the CI of all partitions increased (Table 3), suggesting such sites did exhibit

homoplasy. For *rbcL*, this altered the topology of the MP tree so that it was congruent with the tree obtained from the combined data set (Fig. 1). However, analysis of the more slowly evolving first, second or first and second positions of *rbcL* combined did not result in a topology congruent with Figure 1. For the other partitions, removal of rapidly evolving sites did not result in topological changes.

The larger *rpl16* alignment, which included multiple accessions for most species, was very similar in base composition and parameters to the smaller alignment (Tables 2, 3). The alignment contained 461 sites,

of which 18% were variable and 8% were parsimony informative. As with the smaller *rpl16* alignment, the best model was F81 +  $\Gamma$ . To compare this alignment with the other data partitions, we used HKY85 +  $\Gamma$ . The shape parameter was estimated to be 0.22 and the transition/transversion ratio was 0.43. Analysis of this larger data set produced three MP trees with a CI (excluding uninformative sites) of 0.78. As may occur when analyzing alignments of relatively short sequences (Fehrer, 1996), bootstrap values were low at many nodes (Fig. 3), even when sequence identity was high. For example, the two accessions of *W. neotropica*



**Figure 3.** Phylogeny estimated using the complete *rpl16* alignment with multiple accessions for most species, rooted with *Wolffia* (outgroup taxa not shown). Asterisks identify the accession included in the combined analysis above. Geographic locality of each accession is given. Values at nodes represent per cent of 1000 parsimony bootstrap replicates. Nodes with less than 50% bootstrap support were collapsed.

**Table 4.** Number of nucleotide differences (excluding indels) in *rpl16* sequences between two groups of closely related species

	<i>W. caudata</i>		<i>W. gladiata</i>		<i>W. lingulata</i>		<i>W. oblonga</i>			
	9158	9214	7173	8768	7289	7655	7997	8072	8393	8984
<i>W. caudata</i> 9158	–									
<i>W. caudata</i> 9214	2	–								
<i>W. gladiata</i> 7173	2	2	–							
<i>W. gladiata</i> 8768	2	2	0	–						
<i>W. lingulata</i> 7289	6	6	4	4	–					
<i>W. lingulata</i> 7655	6	6	4	4	0	–				
<i>W. oblonga</i> 7997	2	2	2	2	6	6	–			
<i>W. oblonga</i> 8072	4	4	2	2	6	6	4	–		
<i>W. oblonga</i> 8393	6	6	4	4	0	0	6	6	–	
<i>W. oblonga</i> 8984	6	6	4	4	0	0	6	6	0	–
	<i>W. hyalina</i>		<i>W. repanda</i>		<i>W. rotunda</i>					
	7376	8640	9054	9062	9104	9107	9122	9072	9121	
<i>W. hyalina</i> 7376	–									
<i>W. hyalina</i> 8640	0	–								
<i>W. repanda</i> 9054	1	1	–							
<i>W. repanda</i> 9062	2	2	1	–						
<i>W. repanda</i> 9104	6	6	7	8	–					
<i>W. repanda</i> 9107	6	6	7	8	0	–				
<i>W. repanda</i> 9122	1	1	0	1	7	7	–			
<i>W. rotunda</i> 9072	5	5	6	7	11	11	6	–		
<i>W. rotunda</i> 9121	5	5	6	7	10	10	6	0	–	

were identical, yet *W. neotropica* was supported in only 55% of bootstrap replicates (Fig. 3) and the two identical accessions of *W. gladiata* did not even form a clade (Fig. 3).

In addition to instances of similar *rpl16* sequences that did not cluster together, there were species with some divergent accessions that showed a high degree of similarity to other species (Fig. 3; Table 4). For example, two accessions of *W. oblonga* (8393 and 8984) were identical to each other and also to both accessions of *W. lingulata*. Yet these two accessions differed from putative conspecifics (7997 and 8072) by six nucleotide substitutions. *Wolffiella oblonga* (7997) showed greater similarity to two accessions of *W. caudata* (Table 4), whereas *W. oblonga* (8072) was not closely related to any other accession.

#### BIOGEOGRAPHY

Different biogeographical hypotheses were indicated when relationships within *Wolffiella* were estimated using different types of data (morphology, allozymes,

cpDNA). Morphological data (Fig. 2a) indicated that *Wolffiella* originated either in Africa (distribution of Sections *Stipitatae* and *Rotundae*) or South America (distribution of *W. neotropica*). Whether an African or South American origin is assumed, dispersal between these two continents would have occurred at least three times if the morphological tree were correct. In contrast, both the allozyme and the cpDNA data (Figs 1, 2b) indicated an African origin, followed by two dispersal events between Africa and South America. The allozyme and cpDNA hypotheses differ in that the allozyme data suggest that *W. welwitschii* originated in South America, and then dispersed to Africa while the combined cpDNA data are equivocal on this point (Fig. 1; but see Fig. 3). Regardless of the topology, there have been at least three dispersal events between North and South America, assuming both *W. lingulata* and *W. oblonga* are each monophyletic. Results of the SOWH test indicated that the phylogeny estimated from the cpDNA data (Fig. 1), and the biogeographical hypothesis it supported, were significantly more likely than those supported by either morphological data ( $2\delta = 249.27$ ,  $2\delta$ -critical = 11.84,

$P < 0.002$ ) or allozyme data ( $2\delta = 57.817$ ,  $2\delta$ -critical = 5.21,  $P < 0.002$ ).

## DISCUSSION

### MOLECULAR EVOLUTION

This study is similar to other published results (e.g. Steele & Vilgalys, 1994; Johnson & Soltis, 1995; Manos & Steele, 1997; Xiang, Soltis & Soltis, 1998) in showing that *matK* is more variable and diverges more rapidly than *rbcL* for both synonymous and non-synonymous substitutions. Although *rbcL* appeared to be under greater constraint at non-synonymous sites, it was surprising to find a rapidly evolving first and second position that in both cases resulted in non-synonymous substitutions in the *rbcL* data set. The different amino acids encoded at these sites are those typical of *rbcL* sequences (e.g. Kellogg & Juliano, 1997), and it may be that there is selection for one of several amino acids at these sites, but little selection for a specific one of those amino acids among the taxa examined here. Analysis of the *rbcL* data set that included rapidly evolving first and second position sites resulted in a topology that was inconsistent with all other analyses, suggesting that even non-synonymous sites in *rbcL* may exhibit problematic levels of homoplasy (see also Manos & Steele, 1997).

Although the degree of divergence among non-coding chloroplast regions varies greatly (e.g. Small *et al.*, 1998), the two regions examined here both appeared to be diverging at relatively high rates, and this, combined with a lesser degree of constraint, might be expected to lead to greater homoplasy in these regions. However, the CIs of both non-coding regions were higher than for the coding regions, and removal of rapidly evolving (and potentially more homoplasious) sites had little effect on the resulting topology, indicating that the non-coding regions were not excessively homoplasious at the taxonomic level examined in this study.

### PHYLOGENETIC RELATIONSHIPS IN WOLFFIELIA

Phylogenetic relationships in *Wolffiella* inferred from the present study may be compared to those hypothesized previously for the genus. In an attempt to resolve relationships within *Wolffiella*, Landolt (1986) used primitive and derived states of 26 anatomorphological characters and ranked species phylogenetically according to an 'index of primitivity' for those characters (Fig. 2a). We placed the subsequently named *W. caudata* as the sister to *W. lingulata* in Fig. 2a because Landolt (1992) viewed it as most closely related to that species. Landolt (1986) recognized a basal split within the genus, with one lineage consisting of Sections *Rotundae* and *Stipitatae* and the other

comprising Section *Wolffiella*. Les *et al.* (1997a,b) used 41 anatomorphological characters (including many of the same ones employed by Landolt, 1986) to conduct a phylogenetic analysis using maximum parsimony. That analysis showed the same basal split as Landolt (1986) hypothesized with monophyly of the three sections. However, the genus was not shown to be monophyletic. The uniformity of the anatomorphological traits used by Les *et al.* (1997a,b) precluded resolution of relationships among species. Crawford *et al.* (1997) inferred relationships within *Wolffiella* using allozyme data (Fig. 2b) and recovered the same basal split in the genus as well as clustering of the sections (Fig. 2a,b). Our combined cpDNA sequence phylogeny also supports the basal split as well as the sectional relationships depicted by Landolt (1986) and supported by later studies (Crawford *et al.*, 1997; Les *et al.*, 1997a,b). Of interest may be the tree based on the larger *rpl16* alignment, in which Section *Stipitatae* is not monophyletic (Fig. 3; see also Table 4). Whether this result is real or due to an artefact remains to be determined.

Species in the two lineages resulting from the basal split in *Wolffiella* differ in several ecological features (Landolt, 1986). Members of Sections *Rotundae* and *Stipitatae* live on the surface of seasonal waters and survive dry periods by producing seeds. The seeds germinate quickly, and rapid vegetative reproduction produces large populations that cover the surface of the water with the onset of the wet period (Landolt, 1994). In contrast, members of Section *Wolffiella*, both in the Americas and in Africa, live submersed in permanent waters. When overgrown by other plants, they survive by using organic substances from the water. They are also capable of sinking to the bottom of the water and using nutrients released from the soil. With the exception of *W. welwitschii*, members of Section *Wolffiella* live in permanent waters and do not rely on seeds for reproduction (Landolt, 1986). Thus, the initial phylogenetic split in *Wolffiella* reflects two lineages differing by several fundamental life history and ecological attributes.

Relationships among species in Section *Wolffiella* have been difficult to infer. Les *et al.* (1997a,b) were unable to clarify these relationships using anatomorphological data. Greater resolution was achieved in our combined cpDNA phylogeny and in the allozyme dendrogram which agreed topologically in most respects, differing only in the placement of *W. welwitschii* (cf. Figs 1, 2b). Landolt (1986) portrayed close phylogenetic relationships among *W. gladiata*, *W. lingulata* and *W. oblonga*. He also viewed the newly described *W. caudata* as closely related to these three species, and especially to *W. lingulata* (Landolt, 1992). A close relationship hypothesized between *W. caudata* and the other three

species is concordant with our cpDNA phylogeny. However, *W. caudata* is sister to the other three species in the cpDNA phylogeny whereas Landolt considered it to be most closely related to *W. lingulata*. The phylogeny of Landolt (1986) also differed in the placement of *W. denticulata* and *W. neotropica* (cf. Figs 1, 2a).

The *rpl16* sequences failed to group all accessions of *Wolffiella oblonga* (Fig. 3), with accessions occurring in three different regions of the tree (see also Table 4). The clustering of two accessions of *W. oblonga* (8393 and 8394) with *W. lingulata* is concordant with allozyme data; Crawford *et al.* (1997) found very high allozyme identities between these taxa (the highest yet found between any two Lemnaceae species) with no alleles unique to either species. This result is also supported by morphology, as Landolt (1986) observed that *W. gladiata*, *W. oblonga* and *W. lingulata* were 'very difficult to recognize' and that distinctions between *W. oblonga* and *W. lingulata* were 'especially unclear' due to extensive variability in key, defining characteristics. He further stated that it was difficult to determine whether certain collections contained both species, or a single species displaying different developmental forms (Landolt, 1986). Given this, it is possible that *W. oblonga* and *W. lingulata* may not form distinct species or may have diverged very recently. Also of interest is the accession of *W. oblonga* (7997) that grouped with *W. caudata* in the *rpl16* analysis (Fig. 3, Table 4), though the reasons for this anomalous placement are unknown. Another accession of *W. oblonga* (8072) was distinct in the *rpl16* analysis (Fig. 3) and did not show a high degree of similarity to any other accession. Too few data exist to determine whether this accession represents a novel lineage or whether its position is due to incomplete coalescence. Whatever the case, it is clear that a more detailed study of *W. oblonga* and other closely related species could be fruitful for elucidating their relationships.

#### MOLECULAR BIOGEOGRAPHY OF *WOLFFIELLA*

Determining the biogeographical history of Lemnaceae is challenging because they are readily dispersed by virtue of their minute size. Furthermore, there are several documented examples of species recently introduced by humans (Landolt, 1986). These factors have led to widespread distributions of *Lemna*, *Spirodela* and *Wolffia*. In contrast, *Wolffiella* is restricted to North and South America and Africa. Landolt (1986) suggested that *Wolffiella* originated in the warmer regions of South America because the 'most primitive and probably most ancient' species (*W. neotropica*) occurred there, although his tree was equivocal with regard to the origins of *Wolffiella* (Fig. 2a). Relationships inferred from divergence at

allozyme loci and cpDNA data indicated an African origin for *Wolffiella*, with dispersal to America, followed by a later dispersal of *W. welwitschii* back into Africa (Figs 2b, 3). An African origin is more parsimonious than is a South American origin, as it requires only two dispersal events between Africa and the Americas, rather than the three required if *Wolffiella* originated in South America. We assumed a synonymous substitution rate of 0.12% per Myr for *rbcL*, which is similar to rates calculated for a variety of flowering plants (e.g. Xiang *et al.*, 2000) to estimate the timing of the dispersal from Africa to the Americas. The divergence between *W. denticulata* and *W. welwitschii* was 2.35% at synonymous sites, which sets the estimated time for the divergence between American and African species at approximately 9.8 million years.

Although it appears most likely that *Wolffiella* originated in Africa, it is less clear where *W. welwitschii* (which is distributed in both South America and Africa) evolved. Allozyme data (and morphology) suggest that *W. welwitschii* originated in South America, and then dispersed to Africa, while the combined cpDNA data are equivocal (Figs 1, 2). The accession of *W. welwitschii* that we sequenced for the combined data set was from South America. If the origin of the species were in Africa with dispersal to America, then it should be at least as divergent from the American species as it is from *W. denticulata*. In contrast, if the species originated in America with dispersal back to Africa the American accession should be less divergent from exclusively American species than from the African *W. denticulata* (Fig. 1). The accession of *W. welwitschii* from South America was more divergent from *W. denticulata* than from any American species of *Wolffiella* (about twice as divergent in *matK* and *rbcL*), supporting an American origin for *W. welwitschii* with dispersal back to Africa. Additional support for an American origin occurred in the *rpl16* data, for which both an African and American accession were sequenced (Table 1; Fig. 3). The two accessions of *W. welwitschii* differed by only two substitutions (0.43%) whereas the mean divergence between *W. welwitschii* and the South American *W. neotropica* was 1.1% and divergence between *W. welwitschii* (whether from America or Africa) and its closest African species, *W. denticulata*, was over 2.5%. In *Wolffiella*, the rate of *rpl16* sequence divergence was 1.33 times faster than at *rbcL* synonymous sites (see Results), giving an estimated time of 1.35 Myr for dispersal of *W. welwitschii* back to Africa. Among the species of *Wolffiella*, *W. welwitschii* is best adapted for dispersal because it flowers and sets seed rather frequently, and Lemnaceae seeds are able to survive out of water much longer than can their fronds (Landolt, 1997).

The distribution of *Wolffiella* differs markedly from the distributions of other Lemnaceae genera because it can be explained by only two major dispersal events. In contrast, attempts to reconstruct distributions of other genera using the methods employed in the present study have given ambiguous results (D.H. Les, unpubl. data) due to widespread distributions of several species in other genera. There are several historical and ecological factors that could account for the fewer dispersal events in *Wolffiella* than in the other genera. Molecular data (Crawford & Landolt, 1995; Crawford *et al.*, 1997; Les *et al.*, 2002) showed that species of *Wolffiella* are less divergent on average than are species of other genera of Lemnaceae, suggesting that species of *Wolffiella* are 'younger' than those of other genera. Thus, the lack of dispersal could be a reflection of time since origin. As indicated earlier, *Wolffiella* is restricted to tropical and subtropical regions where distances between continents are greater than in the boreal and nemoral regions of the northern hemisphere, and low dispersal may reflect the greater distances between continents. The fronds of *Wolffiella* are very thin, and cuticles function less efficiently against dehydration than in other Lemnaceae, making it difficult for the fronds to be transported long distances even when covered by feathers (Landolt, 1986). Seeds, which allow for more efficient transport than fronds, are rare in all species except members of Sections *Rotundae* and *Stipitatae* and in *W. welwitschii* (Landolt, 1986). The specialized adaptations of Sections *Rotundae* and *Stipitatae* to seasonally dry, local pools in Africa make them poor colonizers, so dispersal events would likely not result in effective colonization (Landolt, 1994).

#### ACKNOWLEDGEMENTS

We would like to thank E. L. Braun for assisting us with the SOWH test and allowing us to use his computer facilities. We thank two anonymous reviewers for critically reading an earlier version of the manuscript. Research was funded in part by NSF grant DEB-9806537 to D.H.L. and D.J.C.

#### REFERENCES

- Crawford DJ, Landolt E. 1995. Allozyme divergence among species of *Wolffia* (Lemnaceae). *Plant Systematics and Evolution* **197**: 59–70.
- Crawford DJ, Landolt E, Les DH, Tepe E. 1997. Allozyme variation and the taxonomy of *Wolffiella* (Lemnaceae). *Aquatic Botany* **58**: 43–54.
- Farris JS, Källersjö M, Kluge AG, Bult C. 1995. Testing significance of incongruence. *Cladistics* **10**: 315–319.
- Fehrer J. 1996. Conflicting character distribution within different data sets of cardueline finches: Artifact or history? *Molecular Biology and Evolution* **13**: 7–20.
- Goldman N, Anderson JP, Rodrigo AG. 2000. Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* **49**: 652–670.
- Johnson LA, Soltis DE. 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *matK* sequences. *Annals of the Missouri Botanical Garden* **82**: 149–175.
- Jordan WC, Courtney MW, Neigel JE. 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). *American Journal of Botany* **83**: 430–439.
- Kellogg EA, Juliano ND. 1997. The structure and function of rubisco and their implications for systematic studies. *American Journal of Botany* **84**: 413–428.
- Kumar S, Tamura K, Nei M. 1993. *MEGA: molecular evolution genetic analysis*, Vers. 1.0. Pennsylvania: Pennsylvania State University.
- Landolt E. 1986. Biosystematic investigations in the family of duckweeds (Lemnaceae) volume 2. The family of Lemnaceae – a monographic study, volume 1. *Veröffentlichungen des Geobotanischen Institutes der ETH, Stiftung Rubel Zurich* **71**.
- Landolt E. 1992. *Wolffiella caudata*, a new Lemnaceae species from the Bolivian Amazon region. *Berichte des Geobotanischen Institutes der ETH, Stiftung Rubel, in Zurich* **58**: 121–123.
- Landolt E. 1994. The Lemnaceae of Zimbabwe and Botswana. *Berichte des Geobotanischen Institutes der ETH, Stiftung Rubel, in Zurich* **60**: 110–136.
- Landolt E. 1997. How do Lemnaceae (duckweeds) survive dry conditions? *Bulletin of the Geobotanical Institute ETH*. **63**: 25–31.
- Landolt E. 2000. Contribution on the Lemnaceae of Ecuador. *Fragmenta Floristica et Geobotanica* **45**: 221–237.
- Les DH, Crawford DJ. 1999. *Landoltia* (Lemnaceae), a new genus of duckweeds. *Novon* **9**: 530–533.
- Les DH, Crawford DJ, Landolt E, Aakjar R, Tepe E. 1997a. Systematics of Lemnaceae revisited. *American Journal of Botany* **84**: 211.
- Les DH, Crawford DJ, Landolt E, Gabel JD, Kimball RT. 2002. Phylogeny and systematics of Lemnaceae, the duckweed family. *Systematic Botany* **27**: 221–240.
- Les DH, Landolt E, Crawford DJ. 1997b. Systematics of Lemnaceae: inferences from micromolecular and morphological data. *Plant Systematics and Evolution* **204**: 161–177.
- Manos PS, Steele KP. 1997. Phylogenetic analyses of 'higher' Hamamelididae based on plastid sequence data. *American Journal of Botany* **84**: 1407–1419.
- Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**: 418–426.
- Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.
- Rambaut A, Grassly NC. 1997. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution

- along phylogenetic trees. *Computer Applications in Biosciences* **13**: 235–238.
- Simmons MP, Ochoterena H. 2000.** Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* **49**: 369–381.
- Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF. 1998.** The tortoise and the hare: choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany* **85**: 1301–1315.
- Steele KP, Vilgalys R. 1994.** Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. *Systematic Botany* **19**: 126–142.
- Swofford DL. 1999.** *PAUP\*: Phylogenetic analysis using parsimony (\*and other methods)*, Version 4.0. Sunderland, MA: Sinauer Associates.
- Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996.** Phylogenetic inference. In: Hillis DM, Moritz C, Marble BK, eds. *Molecular systematics*, 2nd edn. Sunderland, MA: Sinauer Associates, 404–514.
- Thompson JD, Higgins DG, Gibson TJ. 1994.** Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673–4680.
- Xiang Q-Y, Soltis DE, Soltis PS. 1998.** Phylogenetic relationships of Cornaceae and close relatives inferred from *matK* and *rbcL* sequences. *American Journal of Botany* **85**: 285–297.
- Xiang Q-Y, Soltis DE, Soltis PS, Manchester SR, Crawford DJ. 2000.** Timing the eastern Asian–eastern North American floristic disjunction: molecular clock corroborates paleontological estimates. *Molecular Phylogenetics and Evolution* **15**: 462–472.
- Yang Z. 2002.** *Phylogenetic analysis by maximum likelihood (PAML)*. Vers. 3.12. London: University College.
- Yang Z, Nielsen R. 2000.** Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* **17**: 32–43.

