A phylogeny of Anisophylleaceae based on six nuclear and plastid loci: Ancient disjunctions and recent dispersal between South America, Africa, and Asia

Li-Bing Zhang \textsuperscript{a,b,*}, Mark P. Simmons \textsuperscript{c}, Susanne S. Renner \textsuperscript{d}

\textsuperscript{a} The Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166-0299, USA
\textsuperscript{b} Chengdu Institute of Biology, Chinese Academy of Sciences, P.O. Box 416, Chengdu, Sichuan 610041, PR China
\textsuperscript{c} Department of Biology, Colorado State University, Fort Collins, CO 80523, USA
\textsuperscript{d} Department of Biology, Ludwig Maximilians University, Menzinger-Str. 67, D-80638 Munich, Germany

Received 27 September 2006; revised 12 February 2007; accepted 5 March 2007
Available online 13 March 2007

Abstract

The Anisophylleaceae comprise 29–34 species of shrubs and trees occurring in lowland forests and swamps in tropical Africa, Asia, and South America. These species are placed in four genera with disjunct geographic distributions; \textit{Anisophyllea} has 25–30 species in South America, Africa, and Malesia; \textit{Combretocarpus} has one species in Sumatra and Borneo; \textit{Poga} one species in equatorial Africa; and \textit{Polygonanthus} two in the Amazon Basin. Here we use a phylogeny based on six nuclear and plastid loci sequenced for 15 species representing the four genera to infer their relationships and the relative and absolute ages of the range disjunctions. \textit{Combretocarpus} is sister to the other three genera, and \textit{Polygonanthus} then sister to \textit{Poga} and \textit{Anisophyllea}. \textit{Anisophyllea}, represented by 12 species from all three continents, is monophyletic. A relaxed Bayesian clock calibrated with the oldest fossils from a relevant outgroup, Tetramelaceae, suggests that the disjunctions between \textit{Combretocarpus}, \textit{Poga}, and \textit{Polygonanthus} date back to the Cretaceous, Mid-, and Upper Eocene, whereas the intercontinental disjunctions within \textit{Anisophyllea} appear to date back only some 22–23 million years and thus probably result from long-distance dispersal.

© 2007 Elsevier Inc. All rights reserved.

Keywords: \textit{Anisophyllea}; Anisophylleaceae; Combretocarpus; Poga; Polygonanthus; Intercontinental disjunction; Long-distance dispersal

1. Introduction

Intercontinental disjunctions are among the most fascinating distribution patterns, and addressing the origin of such patterns is an important focus of historical biogeography (for plants, e.g., Engler, 1882; Good, 1947; Croizat, 1952; Carlquist, 1966; Thorne, 1972; Raven and Axelrod, 1974; Renner, 2004). Among the more challenging problems in this regard are disjunctions in the distributions of tropical tree groups, which often lack a fossil record and may have poorly known ranges even for the living species. A prime example of such a taxon is the small tree family Anisophylleaceae.

Anisophylleaceae comprise 29–34 species in four genera: \textit{Anisophyllea} with two species in South America, five to nine in mainland Africa, one in Madagascar, and 15–19 in Malesia, \textit{Combretocarpus} with one species in Sumatra and Borneo, \textit{Poga} with one species in equatorial Africa, and \textit{Polygonanthus} with two in the Brazilian Amazon Basin (Juncosa and Tomlinson, 1988a,b; Schwarzbach, in press; our Fig. 1). Except for a few economically important species, Anisophylleaceae are rarely collected and their ecology, pollination, and seed dispersal biology have not been studied in detail.

* Corresponding author. Address: The Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166-0299, USA. Fax: +1 314 577 9596.
E-mail address: Libing.Zhang@mobot.org (L.-B. Zhang).

1055-7903/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2007.03.002
been studied. *Combretocarpus rotundatus* and *Poga oleosa* are important timber trees, and the seeds of *Poga* also yield cooking oil (Hutchinson and Dalziel, 1954; Sunderland and Obama, 1999). The continental disjunctions among and within the genera (Fig. 1) led Raven and Axelrod (1974) to hypothesize a West Gondwanan history for the group. This hypothesis was also influenced by their acceptance of the then prevailing view that Anisophylleaceae belong to Rhizophoraceae, which have an abundant fossil pollen record going back to the Eocene. Anisophylleaceae and Rhizophoraceae were united mainly because they share incised petals (Tobe and Raven, 1988a,b), entire leaf margin and irregular areoles (Keating and Randrianasolo, 1988), and similar wood anatomy (Hutchinson and Dalziel, 1954; Melchior, 1964; Prance et al., 1975; Takhtajan, 1980).

Molecular data have shown that Anisophylleaceae are closest to the Cucurbitales (Setoguchi et al., 1999; Schwarz-bach and Ricklefs, 2000; Zhang et al., 2006), and modern systems are classifying the family as a member of that order (APG, 1998, 2003). However, only two of the four genera, represented by one or two species, were included in these molecular studies, leaving open the question of monophyly both for the family and the species-rich genus *Anisophylea*. Croizat (1943) and Hutchinson (1973) treated *Polygonanthus* as a separate family, Polygonanthaceae, based on a study by Baehni and Dansereau (1939) that stressed the distinctness of this genus. Though others did not accept this ranking, the affinities of the two species of *Polygonanthus* have remained unclear.

Here we present a phylogeny for the Anisophylleaceae that includes 15 species representing the four genera and that is based on six nuclear and plastid loci. We use these data to infer (1) whether Anisophylleaceae and *Anisophylea* are monophyletic; (2) what the generic relationships are; and (3) how old the range disjunctions may be.

### 2. Materials and methods

#### 2.1. Sampling design

We sampled 15 species from the four genera, namely 12 species of *Anisophylea* (four from Asia, one from South America and, seven from Africa), both species of the monotypic genera *Combretocarpus* and *Poga*, and one of the two species of *Polygonanthus*. Voucher information for new sequences and GenBank accession numbers are listed in Table 1. Earlier analyses have included three species of Anisophylleaceae and representatives of seven of the eight orders of the Eurosid I clade (Zhang et al., 2006; also Zhang and Simmons, 2006) and have shown that Anisophylleaceae are sister to a clade comprising the families of Cucurbitales. We therefore aligned our Anisophylleaceae sequences with the Cucurbitales data set of Zhang et al. (2006; see Section ‘Alignment’).

#### 2.2. DNA isolation, amplification, and sequencing

Total genomic DNA was isolated from silica-dried leaves or herbarium material using DNeasy Plant Mini kits (Qiagen, Valencia, CA), NucleoSpin-Plant kits (Macherey-Nagel, Düren, Germany), or the CTAB method (Doyle and Doyle, 1987), with 4% CTAB used instead of 2% CTAB. When using the kits, the time of incubation at 65 °C was extended to 20 min and 500 µl extraction buffer/100 mg material was used. EB buffer instead of AE buffer was used for elution with the DNeasy kits to avoid possible inhibition of DNA amplifications caused by EDTA. For herbarium material or material from which the DNA yield was low, several mini extractions were conducted and then combined before DNA precipitation. DNA amplification was performed following the protocol described in Zhang and Renner (2003). To increase the
Table 1
Sources of plant material and GenBank Accession numbers for ingroup taxa (Anisophylleaceae) sampled

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Plastid DNA</th>
<th>Nuclear DNA</th>
<th>Provenance/voucher</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>atpB gene</td>
<td>matK gene</td>
<td>ndhF gene</td>
</tr>
<tr>
<td>Anisophyllea R.Br. ex Sabine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. cinnamonoides (Gardn. &amp; Champ.) Alston</td>
<td>AY973431</td>
<td>AY973457</td>
<td>AY973467</td>
</tr>
<tr>
<td>A. corneri Ding Hou</td>
<td>AY9684241</td>
<td>AY9684441</td>
<td>AY9684872</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. disticha (Jack) Baillon</td>
<td>AY973432</td>
<td>AY973458</td>
<td>AY973468</td>
</tr>
<tr>
<td>A. fallax S. Elliot</td>
<td>AY9358491</td>
<td>AY9359231</td>
<td>AY9684883</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. manausensis Pires &amp; W.A. Rodrigues</td>
<td>AY973433</td>
<td>AY973459</td>
<td>AY973469</td>
</tr>
<tr>
<td>A. meniaudi Aubrev. &amp; Pellegr.</td>
<td>AY973434</td>
<td>AY973460</td>
<td>AY973470</td>
</tr>
<tr>
<td>A. myriosticta J.-J. Floret</td>
<td>AY973435</td>
<td>AY973461</td>
<td>AY973471</td>
</tr>
<tr>
<td>A. obtusifolia Engl. &amp; Brehmer</td>
<td>AY973436</td>
<td>AY973472</td>
<td>AY973483</td>
</tr>
<tr>
<td>A. polyneura Floret</td>
<td>AY973437</td>
<td>AY973462</td>
<td>AY973473</td>
</tr>
<tr>
<td>A. pomifera Engl. &amp; Brehmer</td>
<td>AY973438</td>
<td>AY973463</td>
<td>AY973474</td>
</tr>
<tr>
<td>A. purpurascens Hutchinson &amp; Dalziel</td>
<td>AY973439</td>
<td>AY973464</td>
<td>AY973475</td>
</tr>
<tr>
<td>A. sutoria Pierre</td>
<td>AY973440</td>
<td>AY973465</td>
<td>AY973476</td>
</tr>
<tr>
<td>Combretocarpus Hook. f.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combretocarpus rotundatus (Miq.) Danser</td>
<td>AY9684281</td>
<td>AY9684471</td>
<td>AY9684921</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poga Pierre</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. oleosa Pierre</td>
<td>AY973441</td>
<td>AY973466</td>
<td>AY973477</td>
</tr>
<tr>
<td>Polygonanthus Ducke</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. amazonicus Ducke</td>
<td>AY973442</td>
<td>AY973489</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
efficiency and yield, short instead of long PCRs were often performed, especially for poor-quality DNA isolations.

The entire nuclear ITS-1, 5.8S, and ITS-2 region was amplified using external primers ITS-A and ITS-B and internal primers ITS-C and ITS-D (Battner, 1999). The boundaries of ITS1, 5.8S and ITS2 were determined by comparison with those of Soldanella (Zhang et al., 2001). The plastid ATP synthase β subunit (atpB) gene was amplified with forward primers S385R, S766R, S1186R and S1494R and reverse primers rbcl-1, S2, S20, S335, S611 and S1022 (Hoot et al., 1995). The plastid matK (matK) gene was amplified with primers trnK710 (Johnson and Solitis, 1995), trnK-2R (Steele and Vilgalys, 1994), matK-AF and matK-8R (Ooi et al., 1995; matK-8R is complement to matK8 of Steele and Vilgalys, 1994), and primers matK-299F, matK-880F, matK-441R, matK-699R and matK-1018R (Zhang et al., 2006). One portion of approximately 630 basepairs of the plastid NADH dehydrogenase subunit F (ndhF) gene was amplified with primers 972F, 1318F, 1318R (complement to 1318F) and 1603R, 1955R (Olmstead and Sweere, 1994). When these primers did not work, primers 924F and 951F as forward primers, 1318F, 1318R (complement to 1318F) and 1603R, 1955R were determined for this study and have been deposited in GenBank as Accession Nos. AY973422 to AY973489 (Table 1).

Three additional loci (matK, trnK, and rbcL) were included in the analysis, following Davis et al. (1998). A total of 28 ambiguously-aligned positions were excluded from analyses, two in the trnL-F spacer and 26 in the ITS sequences. Ambiguously-aligned nucleotides of individual sequences in regions that could be unambiguously aligned for the remaining sequences were scored as uncertain (“?”). Reading frames of the four protein-coding genes were determined by using MacClade 4.03 (Maddison and Maddison, 2001), with reference to annotated sequences in GenBank.

2.4. Phylogenetic analysis

Gap characters, the inclusion of which often affects the inferred tree topology and increases branch-support values (Simmons et al., 2001), were scored using modified complex indel coding (Simmons and Ochoterena, 2000; Müller, 2006). This modification only applies when asymmetrical step matrices would have been coded by the original complex indel coding. None of the step matrices violated the triangle inequality. Only parsimony–informative gap characters were scored from unambiguously aligned regions. Individual sequences for which the gap was considered ambiguously aligned relative to gaps in other sequences were scored as uncertain for the gap character in question. However, ambiguously-aligned gaps that moved together in each of the alternative equally optimal alignments were included in the analysis, following Davis et al. (1998). A total of 48 gap characters were included (atpB: 0; matK: 5; ndhF: 6; rbcl: 1; trnL-F spacer: 22; ITS: 14).

As a means of data exploration, several alternative potential process partitions (Bull et al., 1993) of the characters were analyzed. Each of the six loci was first analyzed (using nucleotide and gap characters) independently, and the four protein-coding loci (atpB, matK, ndhF, and rbcl) were analyzed using nucleotide as well as amino acid characters (together with their gap characters, when applicable). Linkage units assumed to trace a single evolutionary history, namely the plastid genome and the nrDNA, were analyzed to check for well-supported, contradictory signal. To compare the phylogenetic signal in nucleotide, amino acid, and gap characters, separate analyses were conducted on the nucleotide characters (both from all loci and only the protein-coding genes), the amino acid characters (from all protein-coding genes), and the gap characters (from all loci). Polygonanthus was excluded from the gap-characters-
only analysis because the two loci for which it was sequenced (\textit{atpB} and \textit{rbcL}) contained no gaps. Analyses were also conducted on all nucleotide and gap characters from the protein-coding genes separately from the non-coding regions to compare their relative phylogenetic signal. A simultaneous analysis (Kluge, 1989; Nixon and Carpenter, 1996) of nucleotide and gap characters from all loci was conducted as the primary basis for phylogenetic inference of the Anisophylleaceae. Data matrices have been deposited in TreeBASE (http://www.treebase.org/treebase/) as study Accession No. = S1690 and matrix Accession No. = M3052–M3058. There was a total of 22% missing/inapplicable data in the simultaneous-analysis data matrix.

Equally weighted parsimony tree searches were conducted for each data matrix using 1000 tree-bisection-reconnection (TBR) searches in PAUP + 4.0b10 (Swofford, 2001) with a maximum of 1000 trees held per TBR search. Parsimony jackknife analyses (Farris et al., 1996) were conducted using PAUP* with the removal probability set to approximately e^{-1} (36.7879%), and “jac” resampling emulated. One thousand replicates were performed with ten TBR searches per replicate and a maximum of 100 trees held per TBR search.

Maximum likelihood (Felsenstein, 1973) analyses of nucleotide characters from each of the nine loci were performed as (not infallible; Gaut and Lewis, 1995; Siddall, 1998; Sanderson and Kim, 2000) tests for long-branch attraction (Felsenstein, 1978). One hundred jackknife replicates were performed with one TBR search per replicate and a maximum of 100 trees held per TBR search. Modeltest 3.6 (Posada and Crandall, 1998) was used to select the best-fit likelihood model (of the 56 examined) for maximum likelihood analyses. The Akaike Information Criterion (Akaike, 1974) was used to select among models. The models selected were GTR + I + \Gamma (\textit{atpB}, \textit{ndhF}, \textit{rbcL}, and the analysis of all nucleotide characters), GTR + \Gamma (\textit{ITS}), TVM + \Gamma (\textit{matK}), or TVM + I (\textit{trnL-F}) spacer. The selected model and parameter estimates were then used for tree searches.

2.5. Divergence time estimation

The fossil record for Anisophylleaceae consists of Middle and Late Miocene pollen from Borneo assigned to \textit{Combretocarpus} (Anderson and Muller, 1975; Morley, 1977; R. Morley, personal communication to SR, August 2005). This is insufficient to reliably constrain a molecular clock, and we therefore decided to include the oldest fossils representing the sister clade of Anisophylleaceae, that is, the remaining six families of Cucurbitales (Zhang et al., 2006). Adding these genetically divergent groups, however, worsened the problem of rate heterogeneity, which we attempt to circumvent by using a “relaxed clock” approach (Thorne et al., 1998).

A data set containing only the four protein-coding plastid loci (\textit{atpB}, \textit{matK}, \textit{ndhF}, and \textit{rbcL}) and excluding all gapped sites was used for time estimation; it included 4639 characters. The topology used for dating purposes was the tree found with the full combined data analyzed under parsimony, and the software employed was the Bayesian relaxed clock program of Thorne and Kishino (2002); available at http://statgen.ncsu.edu/thorne/). Parameter values under the F84 + \Gamma model with five rate categories were estimated with PAML’s baseml (version 3.14, available at http://abacus.gene.ucl.ac.uk/software/paml.html; Yang, 1997), this being the most parameter-rich model implemented in Thorne’s program. The latter’s \textit{estbranches} component was used to calculate branch lengths and their variance, given the sequence data, the model parameter output from PAML, and the specified rooted topology. Branch lengths from \textit{estbranches} became the priors for memc searches in \textit{multidivtime} that sought to find the most likely model of rate change (with rate change assumed to be log-normally distributed), given the tree topology, time constraints on nodes, and a Brownian motion parameter (\upsilon) that controls the magnitude of autocorrelation per million years (my) along the descending branches of the tree. Prior gamma distributions on parameters of the relaxed clock model were as follows: the mean and SD of the prior distribution for the root age were set to 100 my, based on fossils (below). The mean and SD of the prior distribution for the ingroup root rate were set to 0.0005 substitutions/site/my by dividing the median of the distances between the ingroup root and the tips by 100 my. The prior and SD for \upsilon were set to 0.02. Markov chains in \textit{multidivtime} were run for 1 million generations, sampling every 100th generation for a total of 10,000 trees, with a burn-in of 10,000 generations before the first sampling of the Markov chain. To check for convergence, we ran several analyses.

To study the effect of the missing data in \textit{Polygonanthus} (3462 nucleotides), we reran the clock analysis with and without this taxon.

We obtained absolute time estimates by constraining the root node of Cucurbitales to maximally 135 my old, based on the onset of angiosperm radiation (Hughes, 1994; Brenner, 1996), and the age of the split between \textit{Datisca} and \textit{Octomeles}/\textit{Tetrameles} to minimally 68 my old, based on woods described as \textit{Tetramelesxylon} \textit{prenudiflora} from the Deccan Intertrappean beds at Moghaonkalan in India (Lakhanpal and Verma, 1965; Lakhanpal, 1970). These woods have key characteristics of \textit{Tetrameles} (E. Wheeler, Department of Wood and Paper Science at North Carolina State University, personal communication to SR, August 2005) and are the oldest fossils of any member of Cucurbitales. The Intertrappean beds at Moghaonkalan are now accepted as Maastrichtian (70.6–65.5 my; Khajuria et al., 1994; S. Chitaley, Cleveland Museum of Natural History, pers. comm. to SR, August 2006). All other fossils of Cucurbitales are considerably younger (see Zhang et al. (2006) for a summary).

For absolute ages we relied on the geologic time scale of Gradstein et al. (2004).
3. Results and discussion

3.1. Comparison of single and combined data sets

The parsimony jackknife trees with parsimony jackknife values above each branch, and maximum likelihood jackknife values below each branch for each of the six loci are presented in Figs. S1–S6 as supplemental data at: http://www.biology.colostate.edu/Research/. One of the two most parsimonious trees of the simultaneous analysis is presented in Fig. 2. The most parsimonious, parsimony jackknife, and likelihood jackknife trees for all analyses are available as supplemental data. Data-matrix and tree statistics for all analyses are presented in Table 2. The four genera of Anisophylleaceae form a strongly supported monophyletic group, sister to all other Cucurbitales (tree not shown, but consistent with that found in Zhang et al. (2006) from five of the six loci used here; the exception being ITS). We assessed congruence between trees obtained from individual data matrices and the simultaneous data by mapping each matrix’s parsimony–informative characters onto the topology found in the simultaneous analysis (Table 2). The nucleotide-based \textit{trnL-F} spacer matrix and the amino acid-based \textit{atpB} and \textit{rbcL} matrices were outliers in requiring 14, 21, and 10% additional steps, respectively, on the ingroup simultaneous-analysis tree. Partitioned analyses yielded a high degree of taxonomic congruence with the simultaneous-analysis tree, and no conflicting clade received more than 65% jackknife support in any partitioned analysis, whether under parsimony or likelihood. The ITS data and the plastid genes each required a single additional step when mapped onto the simultaneous-analysis tree for Anisophylleaceae, but there was no conflict between the parsimony-based ITS and plastid-gene jackknife trees, and only a single clade with 54% jackknife support contradicted in the likelihood-based ITS gene tree. There is thus no reason to suspect different histories between the nrDNA and plastid genes (through lineage sorting, introgression, or unrecognized paralogy without concerted evolution; Doyle, 1992; Wendel et al., 1995; Álvarez and Wendel, 2003).

3.2. Nucleotide-frequency heterogeneity

None of the plastid loci exhibited significant nucleotide heterogeneity among different terminals, as measured by the chi-square test implemented in PAUP* (which ignores phylogenetic correlations). This held both within Anisophylleaceae, and among all ingroup and outgroup terminals in the data set of Zhang et al. (2006) with which we aligned our Anisophylleaceae sequences. In contrast, ITS, which was only sampled for Anisophylleaceae, exhibited highly significant heterogeneity ($x^2 = 153.8$; 140 parsimony–informative nucleotide characters). \textit{Combretocarpus rotundatus}, \textit{Anisophyllea corneri}, and \textit{A. disticha}, and were obvious outliers with an average of 68% GC content in contrast to the 39% average GC content amongst the remaining terminals. These two species of \textit{Anisophyllea} were sister to each other in the ITS tree (Fig. S6), the plastid gene tree (available as supplemental data), and the simultaneous analysis (with 81–100% jackknife support). Since the plastid genes independently supported the relationship, we do not consider their sister status an artifact but rather a case where a shift in nucleotide composition reinforces the phylogenetic signal (Lockhart et al., 1992).

![Fig. 2](#) One of the two most parsimonious trees of the simultaneous analysis as phylogram based on fast optimization. Simultaneous-analysis jackknife tree with parsimony jackknife values are shown above branches, and maximum likelihood jackknife values below branches. Clades in the parsimony jackknife tree that were contradicted by clades in the likelihood jackknife tree are indicated by “•XX•” with the jackknife support for the highest contradictory likelihood clade listed. Dotted lines indicate the unreal branch lengths.
Table 2
Data-matrix and tree statistics for each of the analyses

<table>
<thead>
<tr>
<th>Matrix</th>
<th>No. of chars.</th>
<th>No. of PI chars.</th>
<th>MPT length</th>
<th>No. of MPTs</th>
<th>No. of IG jackknife clades</th>
<th>Avg. IG jackknife support (%)</th>
<th>CI</th>
<th>RI</th>
<th>Increase in IG steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>atpB DNA</td>
<td>1473</td>
<td>248</td>
<td>977</td>
<td>1798</td>
<td>4/3</td>
<td>58/60</td>
<td>.48</td>
<td>.72</td>
<td>3/6%</td>
</tr>
<tr>
<td>atpB amino acid</td>
<td>490</td>
<td>49</td>
<td>208</td>
<td>884,009</td>
<td>1</td>
<td>57</td>
<td>.62</td>
<td>.77</td>
<td>5/21%</td>
</tr>
<tr>
<td>matK DNA</td>
<td>1351</td>
<td>437</td>
<td>1726</td>
<td>12</td>
<td>9/11</td>
<td>86/73</td>
<td>.51</td>
<td>.75</td>
<td>1/2%</td>
</tr>
<tr>
<td>matK amino acid</td>
<td>415</td>
<td>217</td>
<td>1017</td>
<td>40</td>
<td>9</td>
<td>81</td>
<td>.62</td>
<td>.78</td>
<td>1/4%</td>
</tr>
<tr>
<td>ndhF DNA</td>
<td>646</td>
<td>227</td>
<td>973</td>
<td>11,872</td>
<td>5/7</td>
<td>92/81</td>
<td>.51</td>
<td>.73</td>
<td>1/4%</td>
</tr>
<tr>
<td>ndhF amino acid</td>
<td>209</td>
<td>86</td>
<td>451</td>
<td>980,002</td>
<td>3</td>
<td>91</td>
<td>.65</td>
<td>.75</td>
<td>0</td>
</tr>
<tr>
<td>rbcL DNA</td>
<td>1508</td>
<td>228</td>
<td>953</td>
<td>8099</td>
<td>6/6</td>
<td>81/82</td>
<td>.42</td>
<td>.74</td>
<td>2/3%</td>
</tr>
<tr>
<td>rbcL amino acid</td>
<td>474</td>
<td>46</td>
<td>235</td>
<td>960</td>
<td>5</td>
<td>62</td>
<td>.40</td>
<td>.73</td>
<td>3/10%</td>
</tr>
<tr>
<td>rml-F spacer</td>
<td>565</td>
<td>159</td>
<td>367</td>
<td>2420</td>
<td>3/2</td>
<td>77/82</td>
<td>.74</td>
<td>.91</td>
<td>2/14%</td>
</tr>
<tr>
<td>ITS of nrDNA</td>
<td>860</td>
<td>153</td>
<td>569</td>
<td>1</td>
<td>8/9</td>
<td>84/80</td>
<td>.60</td>
<td>.57</td>
<td>2/1%</td>
</tr>
<tr>
<td>Non-coding genes</td>
<td>4775</td>
<td>1137</td>
<td>4660</td>
<td>12</td>
<td>11</td>
<td>85</td>
<td>.48</td>
<td>.73</td>
<td>1/1%</td>
</tr>
<tr>
<td>Plastid genes</td>
<td>1628</td>
<td>315</td>
<td>950</td>
<td>33</td>
<td>11</td>
<td>82</td>
<td>.66</td>
<td>.81</td>
<td>1/0.2%</td>
</tr>
<tr>
<td>Gap chars. only</td>
<td>5543</td>
<td>1299</td>
<td>5049</td>
<td>12</td>
<td>11</td>
<td>88</td>
<td>.50</td>
<td>.75</td>
<td>1/1%</td>
</tr>
<tr>
<td>DNA chars. only</td>
<td>48</td>
<td>47</td>
<td>109</td>
<td>872,000</td>
<td>3</td>
<td>67</td>
<td>.81</td>
<td>.90</td>
<td>2/5%</td>
</tr>
<tr>
<td>DNA exons only</td>
<td>6355</td>
<td>1405</td>
<td>5504</td>
<td>3</td>
<td>13/13</td>
<td>92/94</td>
<td>.50</td>
<td>.74</td>
<td>0</td>
</tr>
<tr>
<td>AA chars. only</td>
<td>4764</td>
<td>1126</td>
<td>4629</td>
<td>18</td>
<td>11</td>
<td>85</td>
<td>.48</td>
<td>.73</td>
<td>1/1%</td>
</tr>
<tr>
<td>Simultaneous</td>
<td>1386</td>
<td>395</td>
<td>1994</td>
<td>64</td>
<td>7</td>
<td>82</td>
<td>.58</td>
<td>.74</td>
<td>2/2%</td>
</tr>
<tr>
<td></td>
<td>6403</td>
<td>1452</td>
<td>5621</td>
<td>2</td>
<td>13</td>
<td>93</td>
<td>.51</td>
<td>.74</td>
<td>N/A</td>
</tr>
</tbody>
</table>

The number of characters for each data matrix is reported after exclusion of alignment–ambiguous regions, if any. PI, parsimony–informative; MPT, most parsimonious tree(s); IG, ingroup; CI, ensemble consistency index (Kluge and Farris, 1969) on the most parsimonious tree(s) for the parsimony–informative characters; RI, ensemble retention index (Farris, 1989). “Increase in IG steps” was determined by mapping the PI characters from the respective data matrix onto the most parsimonious trees from the simultaneous analysis for the ingroup taxa; the lowest increases in steps were reported.

3.3. Parsimony vs. likelihood analyses

The parsimony and likelihood jackknife trees for the six loci analyzed independently of one another contained no contradictory clades except for a conflict between the likelihood jackknife tree of all nucleotide characters and the parsimony-based simultaneous-analysis jackknife tree. In the likelihood tree, the South American Anisophyllea manausensis was resolved as the sister group of the southern Asian clade (A. cinnamomoides, A. corneri, A. disticha) with 64% jackknife support, whereas in the parsimony tree it was resolved as sister to a clade of all other species of Anisophyllea with 52% jackknife support. Because of this contradictory and weakly supported resolution, the placement of A. manausensis as an early-derived lineage within Anisophyllea remains ambiguous.

3.4. Nucleotide vs. gap and amino acid characters

Three clades, with an average of 67% jackknife support, were resolved by the gap-characters-only analysis (Table 2; tree available as supplemental data). All three of these clades are also present in the nucleotide-characters-only tree, which is identical in topology to the simultaneous-analysis tree. We therefore consider the nucleotide and gap characters to have congruent phylogenetic signal.

Of the 18 clades found in the four trees constructed using amino acid characters (available as supplemental data), only one contradicted a clade found in the corresponding tree from nucleotide characters. The contradictory clade received lower jackknife support (51%) than did the 17 matching or congruent clades (on average 78%). The amino acid-based rbcL gene tree resolved an unusual clade, with Combretocarpus and Poga being sister groups with 51% jackknife support, whereas the two genera were separated by two clades with 93% and 99% jackknife support in the nucleotide-based tree (Fig. S4). This sister group relationship was not seen in any of the other parsimony or likelihood jackknife trees (and was generally contradicted). No convergent amino acids (Simmons, 2000) or artifacts caused by composite characters (Simmons and Freudenstein, 2002) were identified upon inspection of the amino acid and nucleotide data matrices, suggesting that in this case silent substitutions were more phylogenetically informative than replacement substitutions (Simmons et al., 2002). One composite character (amino acid character 142) was identified and investigated. This composite character resulted in a different character state (valine) in Fagus (Fagales) than in Combretocarpus and Poga (isoleucine). The jackknife analysis was re-run with the valine in Fagus re-scored as isoleucine, but Combretocarpus and Poga were still resolved as a clade with 52% jackknife support (tree available as supplemental data).

3.5. Monophyly of Anisophylleaceae and position within Cucurbitales

Before the advent of molecular data, Anisophylleaceae were generally included in the Rhizophoraceae because of their shared incised petals (Tobe and Raven, 1988a,b), entire leaf margin and irregular areoles (Keating and Randriansolo, 1988), and similar wood anatomy (Hutchinson
and Dalziel, 1954; Melchior, 1964; Prance et al., 1975; Takhtajan, 1980). However, these families differ in many other characters (Juncosa and Tomlinson, 1988a,b; Behnke, 1988; Tober and Raven, 1988a,b; Vezey et al., 1988), and molecular data now place Rhizophoraceae in the Malpighiales (Schwarzbach and Ricklefs, 2000; APG, 2003). Takhtajan (1969) and Hutchinson (1973) were among the few botanists who early on ranked Anisophylleaceae as a family (as first done by Ridley (1922)), and both placed Anisophylleaceae and Rhizophoraceae in Myrtales, in agreement with embryological characters (summarized in Tobe and Raven, 1987). Their Anisophylleaceae did not include Polygonanthus because of a study by Baehni and Dansereau (1939) that stressed the differences between Polygonanthus and the other Anisophylleaceae. (Baehni and Dansereau’s study, which suggested that Polygonanthus was a Saxifragaceae, also was the reason why Croizat (1943) and Hutchinson (1973) placed Polygonanthus in a separate family, Polygonanthaceae.) Our data strongly support the monophyly of Anisophylleaceae (Fig. 2). Indeed, all loci that could be aligned across ingroup and outgroup by themselves supported the family with 100% jackknife support (Fig. 1S–6S). Most morphological, anatomical, palynological, and embryological studies also have concluded that the four genera of Anisophylleaceae are closely related (Tober and Raven, 1987; Dahlgren, 1988; Juncosa and Tomlinson, 1988a,b; Keating and Randrianasolo, 1988; Vezey et al., 1988). The genetic distance between Anisophylleaceae and the remaining Cucurbitales, however, is very great, and we know of no morphological synapomorphies for the order (morphological characters are discussed in Zhang et al., 2006). Classification systems with narrowly circumscribed orders, such as that of Takhtajan (1997), still accord Anisophylleaceae ordinal rank.

3.6. Infra-familial relationships of Anisophylleaceae

Little has been published on the generic relationships within Anisophylleaceae. Keys usually first separate Combretocarpus from the other genera based on its winged fruits (Hutchinson, 1973; Schwarzbach, in press), and this basal split is confirmed by our data (Fig. 2; all loci either placed Poga as sister to Anisophyllea or were consistent with this placement: Figs. 1S–6S). Combretocarpus has monomorphic leaves with pinnate venation, while the other genera have mostly dimorphic leaves with divergent venation. Poga, in turn, differs from Anisophyllea and Polygonanthus by having 3–4 seeds instead of one, and the last two genera are distinguished from each other by lobed vs. unlobed petals (Dahlgren, 1988; Schwarzbach, in press).

Polygonanthus has two species, both endemic to the Amazon basin of Brazil (Prance et al., 1975), and given that the genus has been ranked as a separate family (above), one would have expected it to place as sister to the other three genera or at least as forming a long branch. This is not the case (Figs. 2, S5). However, since Polygonanthus lacked several sequences (Table 1), the short branch leading to this taxon in Fig. 1 must underestimate the true number of substitutions.

Anisophyllea, the most diverse genus in the family, contains 25–30 species (Juncosa and Tomlinson, 1988a,b; Schwarzbach, in press) of which we sampled several representatives each from Africa and Asia and one of its two South American species for a total of 12. With the exception of the slowly evolving atpB gene, all datasets and analyses strongly support Anisophyllea monophyly (Figs. 2 and S1–S6). Within Anisophyllea, the South American A. manausensis was ambiguously resolved as sister to all other species (see Section ‘Parsimony vs. Likelihood Analyses’), and the Asian and African clades were resolved as sister to each other. Within the African clade, the Madagascar A. fallax and the West African A. myristicta together were sister to the rest of the African species (mostly endemic to West Africa). However, further species sampling and detailed morphological studies are needed to reliably infer relationships within Anisophyllea.

3.7. Biogeography of Anisophylleaceae

Divergence time estimates with 95% confidence intervals for nodes of interest are listed in Table 3, and Fig. 3 shows a chronogram summarizing the temporal unfolding of Anisophylleaceae. As discussed, the only known fossils of Anisophylleaceae are 13–5 my old Mid- and Upper Mio- cene pollen grains from Borneo (Anderson and Muller, 1975; Morley, 1977). While documenting the Miocene presence of Combretocarpus on Borneo, they are clearly much younger than the split between Combretocarpus and the remainder of the family, which is estimated from molecular distances as having occurred in the Cretaceous, some 85 my ago (Table 3). A Cretaceous age for the earliest divergence within Anisophylleaceae is consistent with the oldest fossils of their sister clade, the Cucurbitales, which are Maastrichtian (70.6–65.5 my; cf. Materials and methods). These Maastrichtian fossils represent Tetramelaceae, a member of the core Cucurbitales (Zhang et al., 2006), and so likely underestimate the true age of the order. The sister group of the Cucurbitales appears to be the Fagales (Chase et al., 1993; Zhang et al., 2006), which have a fossil record dating back 98 my (Crepet et al., 2004). Based on this overall fossil context and the molecular clock estimates obtained from the plastid genes analyzed here (Table 3), it appears likely that the deepest divergence in Anisophylleaceae occurred as Gondwana was breaking up, while Polygonanthus and Poga both may date back some 40 my. Our finding (Table 3) that relaxed clock estimates are relatively robust against the inclusion or exclusion of single taxa with large amounts of missing data agrees with other studies (e.g., Zhou et al., 2006). In contrast to the Cretaceous and Eocene ages of Combretocarpus, and Polygonanthus, and Poga, the splits between the South American, African, and Asian species of Anisophyllea are only some 22–23 my old (Table 3), implying long distance dispersal across water. Little is known about fruit dispersal in Anisophyllea, but several
Table 3
Divergence time estimates in million years (with 95% confidence intervals in parentheses) for nodes of interest obtained from a Bayesian relaxed clock constrained with the oldest fossils of the sister clade of Anisophylleaceae

<table>
<thead>
<tr>
<th>Dated nodes</th>
<th>Time (my) including Polygonanthus</th>
<th>Time (my) excluding Polygonanthus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combretocarpus vs. remaining Anisophylleaceae</td>
<td>85 (67–107)</td>
<td>87 (68–108)</td>
</tr>
<tr>
<td>Polygonanthus vs. remaining Anisophylleaceae</td>
<td>44 (30–63)</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Poga vs. remaining Anisophylleaceae</td>
<td>37 (26–53)</td>
<td>40 (28–54)</td>
</tr>
<tr>
<td>A. manausensis vs. remaining Anisophylleaceae</td>
<td>23 (15–34)</td>
<td>23 (15–33)</td>
</tr>
<tr>
<td>African vs. Asian clades of Anisophyllea</td>
<td>22 (14–32)</td>
<td>22 (14–31)</td>
</tr>
</tbody>
</table>

Nodes are labeled as in Fig. 3. Polygonanthus lacks 3462 of the 4639 nucleotides used for time estimation (Table 1), and the analyses were therefore repeated with and without this taxon.

Fig. 3. Chronogram obtained from a relaxed clock approach applied to the four plastid-coding genes and constrained with the oldest fossils of Tetramelaceae (associated with node X).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2007.03.002.

References


Acknowledgments

We thank N. Fujii, L.G. Saw, S. Lee, D. Kenfack and S. Moses, H. Setoguchi, H.T.W. Tan, and curators of MO and NY for plant material, and two anonymous reviewers for helpful suggestions that improved the manuscript. NSF grant DEB-0344837 to L.A. Johnson provided support to LBZ during much of the writing.


