Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 36 (2005) 581–597
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

The internal transcribed spacer of nuclear ribosomal DNA in the gymnosperm *Gnetum*

Hyosig Won [1], Susanne S. Renner *,[2]

Department of Biology, University of Missouri-St. Louis, 8001 Natural Bridge Road, St. Louis, MO 63121, USA
Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166, USA

Received 31 July 2004; revised 8 March 2005
Available online 19 April 2005

Abstract

We analyze the structure of the internal transcribed spacers ITS1 and ITS2 of the nuclear ribosomal DNA in the gymnosperm *Gnetum*, using a phylogenetic framework derived mainly from an intron in the nuclear low-copy *LEAFY* gene. *Gnetum* comprises 25–35 species in South America, Africa, and Asia, of which we sampled 16, each with two to six clones. Criteria used to assess ITS functionality were highly divergent nucleotide substitution, GC content, secondary structure, and incongruent phylogenetic placement of presumed paralogs. The length of ITS1 ranged from 225 to 986 bp and that of ITS2 from 259 to 305 bp, the largest ranges so far reported from seed plants. *Gnetum* ITS1 contains two informative sequence motifs, but different from other gymnosperms, there are only few and short (7–13 bp) tandem repeats. *Gnetum* ITS2 contains two structural motifs, modified in different clades by shortening of stems and loops. Conspecific sequences grouped together except for two recombinant pseudogenes that had ITS1 of one clade and ITS2 of another. Most of the pseudogenic ITS copies, paralogs, and putative chimeras occurred in a clade that according to a fossil-calibrated chloroplast-DNA clock has an age of a few million years. Based on morphology and chromosome numbers, the most plausible causes of the observed high levels of ITS polymorphism are hybridization, allopolyploidy, and introgression.

Keywords: Gymnosperms; Internal transcribed spacer; *LEAFY* gene second intron; Pseudogene; Chimeric sequences; Secondary structure

1. Introduction

The internal transcribed spacer (ITS) of nuclear ribosomal DNA (rDNA) is one of the most extensively sequenced molecular markers (see Álvarez and Wendel, 2003 for recent usage statistics). The region is part of the rDNA cistron, which consists of 18S, ITS1, 5.8S, ITS2, and 26S, and is present in several hundred copies in most eukaryotes. ITS1 and ITS2 as not equivalent structures, even though they are sometimes convergent in length as well as substitutional patterns; ITS1 evolved from an intergenic spacer and ITS2 from an expansion segment in the rDNA large subunit (Hershkovitz et al., 1999). Thus, significant differences in their evolutionary trajectories can be expected. In plant genomes, thousands of copies are present and are located in one or several loci, distributed on one or several chromosomes. Nuclear rDNA copies within a genome can be highly homogeneous due to concerted evolution of intra- and interchromosomal loci. Alternatively, divergent copies can exist within a genome due to hybridization, lineage sorting, recombination among copies, and pseudogenization of cistrons (Álvarez and Wendel, 2003; Bailey et al., 2003; Baldwin et al., 1995; Buckler et al., 1997; Muir et al., 2001; Sanderson and Doyle, 1992; Suh et al., 1993; Volkov et al., 1999; Vollmer and Palumbi, 2004; Wendel et al., 1995).

* Corresponding author. Fax: +49 89 172638.
E-mail address: renner@lrz.uni-muenchen.de (S.S. Renner).
[1] Present address: School of Biological Sciences, Seoul National University, Seoul 151-747, Republic of Korea.
[2] Present address: Ludwig Maximilians Universität, Systematische Botanik, Menzinger Str. 67, D-80638 Munich, Germany.

1055-7903/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2005.03.011

The length of the ITS region of flowering seed plants is highly uniform (Baldwin et al., 1995). By contrast, that of non-flowering seed plants shows much variation. Especially the length of the ITS1 region, which ranges from 630 to 3125 bp, is strikingly more labile than it is in flowering plants (Liston et al., 1996; Maggini et al., 1998); ITS2 in gymnosperms varies only between 225 and 255 bp. The great ITS1 length variation, combined with the presence of tandem repeated sequences and paralogous copies, has limited the use of ITS for phylogeny reconstruction in non-flowering plants (but see Gernandt and Liston, 1999; Gernandt et al., 2001; Li et al., 2001; Liston et al., 1999; Soltis and Soltis, 1998).

We here investigate the evolution of ITS within the gymnosperm genus *Gnetum*. *Gnetum* consists of 25–35 species of trees and climbers that occur in tropical forests in South America, Africa, and Southeast Asia, and it is one of three genera of the ancient seed plant lineage Gnetales. Multiple nuclear, chloroplast, and mitochondrial gene sequences are available for 25 species, providing a phylogenetic framework for *Gnetum* (Won and Renner, 2003, submitted-a, submitted-b), a member of a clade that is at least 115 million years old (Rydin et al., 2003).

For comparison with the multi-copy ITS region, we chose a low-copy nuclear region, specifically the second intron of the nuclear *LEAFY/FLORICAULA* gene. In most non-flowering seed plants lineages, this gene has two copies, *PRFLL* and *NEEDLY* (Mellerowicz et al., 1998; Mouradov et al., 1998), but *Gnetum* has only the *LEAFY* homolog (Frohlich and Meyerowitz, 1997; Frohlich and Parker, 2000; Frohlich, pers. comm., 2000). *PRFLL* is a developmental control gene expressed during early stages of male cone development, and *NEEDLY* is expressed in developing female cones of *Pinus* (Mouradov et al., 1998). By comparing two biparentally inherited regions, one with thousands of copies, the other with just one copy, we aimed to examine phylogenetic utility of paralogous copies, especially given current concerns about rampant paralogy of the ITS region, at least in flowering plants (Álvarez and Wendel, 2003; Bailey et al., 2003).

2. Materials and methods

2.1. Taxon sampling

Species names and their authors, voucher information, and GenBank accession numbers for all 138 sequences generated for this study are listed in Appendices A (ITS) and B (*LEAFY*). Most samples were collected in the field by the first author, but a few came from botanical gardens or herbarium material. To represent ca. seven *Gnetum* species recognized from southern China, but whose identities need further taxonomic study (Fu et al., 1999), nine samples collected from southern China and neighboring area were included (as *G. hainanense* s. lat. and *G.* sp. Harder et al. 5621).

2.2. Sequencing

We extracted DNA from silica gel-dried leaves, using QIAGEN Plant DNeasy mini kits. Concentration and quality of extracted DNAs were checked by 1% agarose gel electrophoresis with a Lambda/*Hin*dIII/*Eco*RI size marker. We PCR-amplified the entire internal transcribed spacer (ITS) region, using primers ITSA and ITSB of Blattner (1999), and *Taq* DNA polymerase (Promega). Approximate locations of primers are shown in Fig. 1. Reaction volumes were 25 µl and contained 2.5 µl of *Taq* 10× reaction buffer, 1.5 mM $MgCl_2$, 0.4 µM primer, 0.2 mM of each dNTP, 0.625 U of *Taq* DNA polymerase, and 0.5–2 µg of template DNA. 1.5 µl dimethyl sulfoxide (DMSO) was added to some reactions (compare Section 3). PCR cycles consisted of 30 cycles of 1 min at 97 °C for template denaturation, 1 min at 53 °C for primer annealing, and 1 min 20 s at 72 °C for primer extension, followed by 7 min at 72 °C for completion of primer extension. After electrophoresis of the PCR product in a 1% agarose gel, single bands of amplified ITS were cut out and purified with QIAGEN gel extraction kits or QIAEX II gel extraction kits. Purified PCR products were cloned, using the Promega pGEM-T easy vector system according to the manufacturer's protocol. Plasmids were purified using the alkaline lysis method (Maniatis et al., 1982), after screening and proliferation of the transformed cells. To check for the

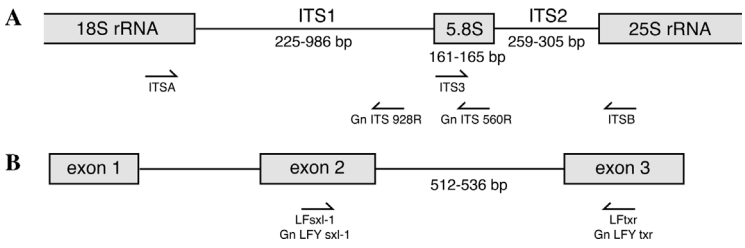

Fig. 1. Approximate locations of primers used for nuclear ribosomal ITS (A) and *LEAFY* (B) amplification. Coding regions and ribosomal RNA regions are shown as boxes, and introns and spacers as lines. Numbers below lines refer to numbers of base pairs.

presence and size of an insert, a portion of purified plasmids was digested with the restriction enzyme *Eco*RI and checked by 1.0% agarose gel electrophoresis with a Lambda/*Hin*dIII/*Eco*RI size marker. Once checked, plasmids were PCR-sequenced using the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer) with the plasmids' universal internal primers T7 and SP6, and with sequencing primers ITS3 (White et al., 1990), Gn ITS-928R (5′-CCTTGGTGC CACTCGCACCC-3′), and Gn ITS-560R (5′-TGAT TCACGGGATTCTGC-3′). The *Gnetum* (Gn) ITS- 560R primer accommodates mutations in the ITS2 primer site. At least two clones per sample were sequenced, each with both strands. Sequences were aligned and cleaned up in Seqman II (DNASTAR).

The same amplification and sequencing protocol was used for the *LEAFY* gene. We designed specific primers, *Gn LFsxl-1* (5′-GARCAYCCITTYATIGTIACIGA-3′) and *Gn LFtxr* (5′-ACRTARTGICKCATYTTIGGY TT-3′; Fig. 1), by modifying those of Frohlich and Meyerowitz (1997). PCR settings were identical to those used for nrITS, except that we used twice the amount of primers because of their degeneracy. Usually, two clones per sample were sequenced.

2.3. Sequence analysis

Gnetum ITS sequences were compared to homologous sequences from other seed plants (Hershkovitz and Lewis, 1996; Hershkovitz and Zimmer, 1996; Hershkovitz et al., 1999) and to a 5.8S rRNA sequence of *Gnetum gnemon* (Troitsky et al., 1991) to determine the boundaries of the ITS1, 5.8S, and ITS2 regions. The 3′ ends of *Gnetum* 5.8S contained indels, causing discrepancies with the previously reported 5.8S sequence, and thus we set arbitrary borders that matched the majority of available seed plant sequences. Lengths and G + C contents of the ITS regions, and pairwise Kimura-2-parameter (K2P; Kimura, 1980) divergences were calculated after excluding ambiguously aligned regions, putative paralogous, and pseudogenic sequences from the data. Plotting the G + C contents of the ITS1, 5.8S, and ITS2 regions served to assess possible escape from functional constraints and changes in substitution rates.

A data matrix of *Gnetum* ITS sequences was constructed using Se-Al software (ver. 2.0a11; http://evolve.zoo.ox.ac.uk/). Clustal X (Thompson et al., 1997) was used to align 5.8S and ITS2, with subsequent manual adjustment. A schematic view of the alignment is shown in Fig. 2. Length and G + C content of ITS1, 5.8S, and ITS2 were calculated from the aligned sequences. The Tandem Repeats Finder program (http://c3.biomath.mssm.edu/trf.html; Benson, 1999) was used to detect repeats, and the secondary structure of the ITS regions was predicted with the mfold software (ver. 3.1, http://www.bioinfo.rpi.edu/applications/mfold/; Zuker, 2003). Options of mfold were set such that each ITS regions could fold independent of adjacent rRNA because several studies have shown that the folding of 5.8S and 18S rRNA is independent from that of ITS1 and ITS2 (Gutell, 1993; Jobes and Thien, 1997; Neefs et al., 1993; Suh et al., 1992; van der Sande et al., 1992; van Nues et al., 1995). However, 3′ 5.8S and 5′ 26S rRNA must form the proximal stem (Côté and Peculis, 2001).

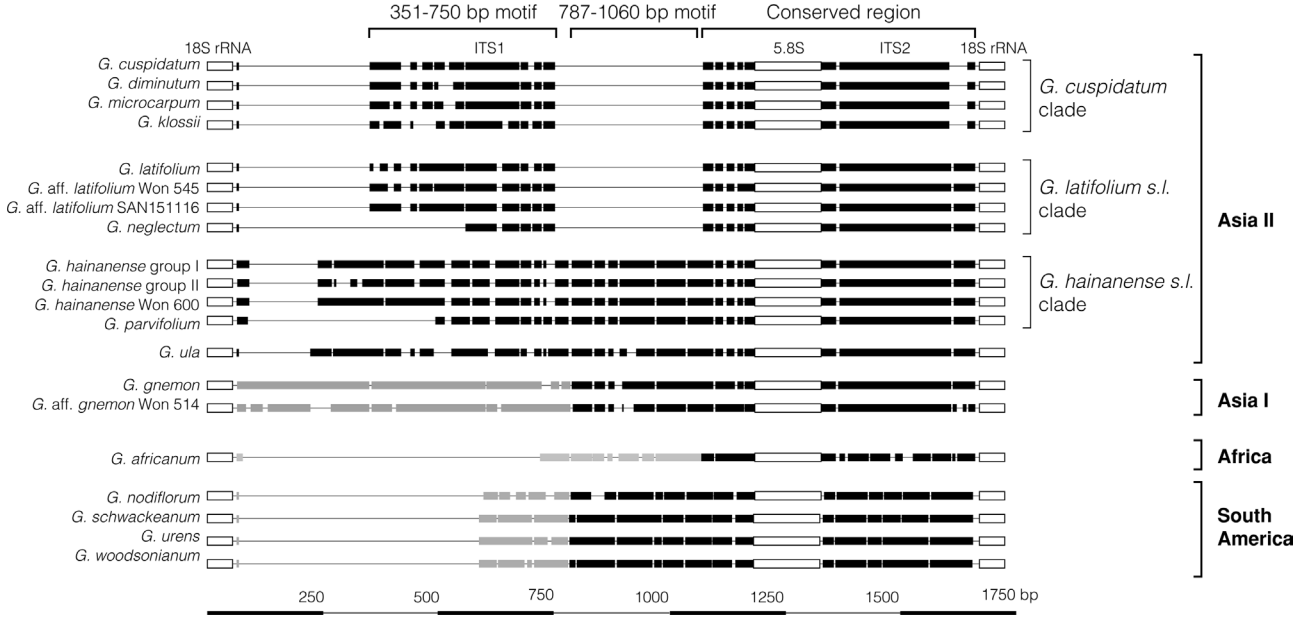

Fig. 2. Overview of the ITS alignment. Open boxes represent ribosomal RNA genes, closed boxes represent spacer sequences, and thin lines indicate gaps. Different shades of closed boxes represent sequence homologies. The bottom scale (in base pairs) indicates relative lengths.

To find *LEAFY* exon and intron boundaries, we compared our sequences to GenBank sequences of *G. gnemon*, *G. leyboldii* (unvouchered and apparently *G. woodsonianum*; H. Won, unpublished), and *Welwitschia mirabilis* (Frohlich and Meyerowitz, 1997). Sequence analysis of the *LEAFY* intron followed the same procedure as used for the ITS region.

2.4. Sequence homology assessment and phylogenetic analysis

To assess possible hybridization, lineage sorting, or cross-contamination during PCR amplification or sequencing, we analyzed multiple sequences per accession and species. Phylogenetic analyses relied on Bayesian inference, neighbor joining (NJ), parsimony, and the quartet maximum-likelihood method (QP; Strimmer and von Haeseler, 1996), using the programs and settings specified below. Trees were rooted with the South American species of *Gnetum*, based on analyses of a large combined data set (*rbcL*, *matK*, the cp *trnT-trnF* region, nuclear rITS/5.8S, the second intron of the nuclear *LEAFY* gene, and the second intron of the mitochondrial *nad1* gene) and broader outgroup sampling (Won and Renner, 2003, submitted-a, submitted-b).

The conserved part of the ITS region (consisting of the 3′ end of ITS1, 5.8S, and ITS2) and the ITS1 sequences of species from just the Asian clade of *Gnetum* were analyzed separately. We partitioned the *LEAFY* gene into intron and exon regions and estimated the shape parameter of the gamma distribution (below) separately for each. Gaps were treated as missing data.

Bayesian inference relied on MrBayes (ver. 3.0b4; Huelsenbeck and Ronquist, 2001; also Ronquist and Huelsenbeck's, 2003 on-line manual). Bayesian probabilities were obtained under the GTR + Gamma + Pinv model, with four Markov chain Monte Carlo chains run for 300,000 generations, using random trees as starting point, and sampling every 100th generation. Trees sampled before the saturation of the maximum-likelihood estimate were discarded as burn-in. Runs were repeated several times as a safeguard against spurious results. The GTR + Gamma + Pinv model was selected by Modeltest (vs. 3.6; Posada and Crandall, 1998) as the best-fit likelihood model for our data. The selected model and parameter estimates were then used for tree searches.

Parsimony, NJ, and QP relied on PAUP (vs. 4.0b10; Swofford, 2002). Heuristic searches used 100 random taxon-addition replicates, holding 100 trees at each step, with TBR branch swapping and the MulTrees, Collapse, and Steepest Descent options, and no upper limit for trees held in memory. Quartett puzzling used the model parameter values estimated in the Bayesian analyses, but for the neighbor joining analyses we relied on simple K2P distances for ready comparison with earlier studies of ITS divergences within and between taxa.

Clade support was assessed via Bayesian posterior probabilities or non-parametric bootstrapping, with 1000 replicates, using the same settings as in the respective tree searches.

Phylogenies obtained from the ITS and *LEAFY* regions were congruent, and one sequence per sample/species was therefore chosen as representative for a combined analysis. Search strategies were as above, but with separate model parameters estimated for the ITS region and the *LEAFY* exon and intron partitions.

3. Results

3.1. PCR and cloning

The ITS primer combinations described in Section 2, with few exceptions, yielded full overlap of forward and backwards strands, and we thus obtained 74 complete ITS sequences and three incomplete ones, together representing 16 species (Appendix A). PCRs with the ITSA + ITSB primers usually yielded a single band, while the combinations ITSA + Gn ITS 560R and ITS3 + ITSB sometimes yielded double or multiple bands. Sequencing verified the latter to be fungal contaminations, and they were discarded. DMSO was included in the PCR of *G. africanum*, *G. gnemon*, *G. nodiflorum*, *G. ula*, *G.* sp. Harder et al. 5621, and several *G. hainanense* s.l. samples, but was left out from other samples since it caused amplification problems. Direct sequencing of purified PCR products usually yielded weak base calls or a short readable run, and cloning of PCR products was therefore necessary (Appendix A identifies the directly sequenced vs. cloned sequences).

Restriction enzyme digestion of screened and purified plasmids by *Eco*RI revealed a single restriction site in *G. gnemon*, and sequencing verified this to lie within the ITS1 region. No other *Eco*RI restriction sites were detected in the ITS regions of *Gnetum*.

PCRs with the *LEAFY* primers yielded faint and thin single bands, and direct sequencing of the purified PCR products generated very weak base calls. We thus cloned all *LEAFY* products. Overall, we obtained 61 *LEAFY* second intron sequences, representing *W. mirabilis* and 21 species of *Gnetum* (Appendix B). We were unable to amplify *LEAFY* from *Ephedra*, but others have found two copies (*LEAFY* and *NEEDLY*; Ickert-Bond and Wojciechowski, 2002; Ickert-Bond, pers. comm.).

3.2. Length and G + C content

The length of *Gnetum* ITS1 ranged from 225 to 986 bp and that of ITS2 from 259 to 305 bp (Table 1, Fig. 2). Species in the Asia I clade (the major clades of *Gnetum* are labeled in Fig. 3) had the longest ITS1 (919–986 bp) and those in the Asia II clade the most variable (225–

Table 1
Summary of length variation (bp) and G + C content (%) of the nuclear ribosomal ITS in Gnetales after excluding putative paralogous pseudogenes (cf. Appendix A)

Group	ITS1		ITS2		5.8S	
	Length	(G + C %)	Length	(G + C %)	Length	(G + C %)
Gnetum ($n = 59$)	225–986	(59.5–68.5)	259–305	(61.9–70.6)	161–165	(50.0–53.3)
South America ($n = 10$)	447–521	(59.5–62.0)	292–299	(64.9–67.1)	164–165	(50.3–52.4)
Africa ($n = 3$)	410	(67.8–68.5)	259–261	(67.2–68.2)	164	(50.6–51.8)
Asia I clade ($n = 5$)	919–986	(59.9–61.0)	291–302	(61.9–62.8)	162–164	(50.0–51.9)
Asia II clade ($n = 41$)	225–802	(60.5–68.2)	268–305	(65.0–70.6)	161–165	(48.8–53.3)
Welwitschia[a]	~309	(64.1)	~285	(57.9)	~160	(45.6)
Ephedra	1108–1139[b]	(60.1)[c]	238[c]	(63.9)[c]	161[c]	(54.7)[c]

[a] U50740 (Liston et al., 1996) contains many ambiguous bases.
[b] Huang (2000) and Huang et al. (2005) only reported the length variation of ITS1 in *Ephedra*.
[c] Maggini et al. (1998) reported the ITS sequence of *E. fragilis*.

802 bp). South American species and the single African species had intermediate ITS1 sizes (447–521 and 410 bp). The length of the 5.8S region ranged from 161 to 165 bp, with all length variation caused by indels at the 3′ ends.

G + C contents were 59.9–68.5% for ITS1, 48.8–53.3% for 5.8S, and 61.9–70.6% for ITS2, excluding putative paralogous and chimeric sequences (Table 1, Fig. 4A).

The length of the *LEAFY* second intron was 512–536 bp for *Gnetum* and 491 bp for the *LEAFY* homolog of *Welwitschia* (Appendix B). Except for *G. ula*, which had a 14-bp deletion and thus an intron of only 512–513 bp, the introns of the remaining species were 523–536 bp long, with small size variations caused by 1- to 6-bp long gaps. The amplified exon was 230 bp long, except in three cases where it was one base pair shorter, probably due to *Taq* enzyme error. *Gnetum LEAFY* introns had higher G + C contents than found in *Welwitschia* (43.3–45.3% vs. 38.7%); exon G + C contents between genera were comparable (45.2–49.3% vs. 46.5%).

3.3. Sequence alignment, motifs, and divergence

Manual alignment, using the 'block color' option of the Se-Al program, was straightforward for 5.8S rDNA and ITS2, but highly ambiguous for ITS1 due to the fourfold length variation (225–986 bp, above). Only 69–84 bp at the 3′ end of ITS1 were relatively conserved. We detected two conserved motifs in ITS 1, one between base positions 351 and 750 present in all members of the Asia II clade (node B in Fig. 3A), and a second at positions 787 and 1060 present in all South American and some Asian species (both motifs are also marked in Fig. 2). The African species, *G. africanum*, did not contain either motif. In addition, the seed-plant-specific conserved 14-bp motif 5′-GAATTGCAGAATCC-3′ was detected in the 5.8S of all *Gnetum* sequences (Hershkovitz and Lewis, 1996; Jobes and Thien, 1997).

The Tandem Repeats Finder program detected a few putative repeats (Table 2), but only with the most relaxed search options (alignment parameters 2, 3, 5, and minimum alignment score 30). The detected repeat sequences were 7, 8, 10, 12, 13 bp in length, had 2.2–4.3 copies, and their match points were between 0.61 and 0.85. Except for a 10-bp repeat found in the ITS2 region of *G. schwackeanum*, *G. urens*, and *G. woodsonianum* (all from South America), all repeats were located in the ITS1 region. A marker found in the ITS2 region was a 39-bp deletion synapomorphic for the *G. cuspidatum* subclade of the Asia II clade (node C in Fig. 3A).

In the conserved ITS1/5.8S/ITS2 region, K2P divergences between clades were 0.126 to 0.446, while divergences within clades were maximally 0.123 (Table 3). Divergences within species ranged from 0 to 0.080 (for >2 clones; data not shown). Sequence divergence in the ITS1 motifs within clades was considerably smaller than between clades (Table 4).

3.4. Pseudogenes and chimeras

Separate phylogenetic analyses of the variable ITS1 region and the conserved 5.8S/ITS2 and 3′ITS1 yielded largely congruent topological relationships (Figs. 3A and B). Most conspecific sequences grouped together with moderate to high statistical support, irrespective of method of phylogenetic inference. However, the following sequences (highlighted in Figs. 3A and B) jumped positions: SAN 151122-C1 (*G. diminutum*) and Won 533-C3 and Won 533-C1 (both from *G.* aff. *microcarpum*) placed differently in the tree obtained from the conserved regions (3A) than in the tree obtained from the variable region of ITS1 (3B). SAN 151122-C1 has an ITS1 region that matches sequences of *G. diminutum* and *G.* aff. *microcarpum* (368 bp) and an ITS2 region that matches *G. klossii* SAN151124-C1 (259 bp). Similarly, Won 533-C3 has an ITS1 region that matches *G. diminutum* and *G.* aff. *microcarpum* and an ITS2 region that matches most *G. klossii* ITS2 regions (268 bp). The third unusual sequence, Won 533-C1 (*G.* aff. *microcarpum*), has the longest ITS1 region in its clade (405 bp vs. 340–395 bp) and retains a 39-bp deletion in ITS2 otherwise lost in its clade (node C in Fig. 3 and Section 3.5). The conserved ITS region placed Won 533-C1 as sister to SAN151117-C1 (Fig. 3A, solid line) or

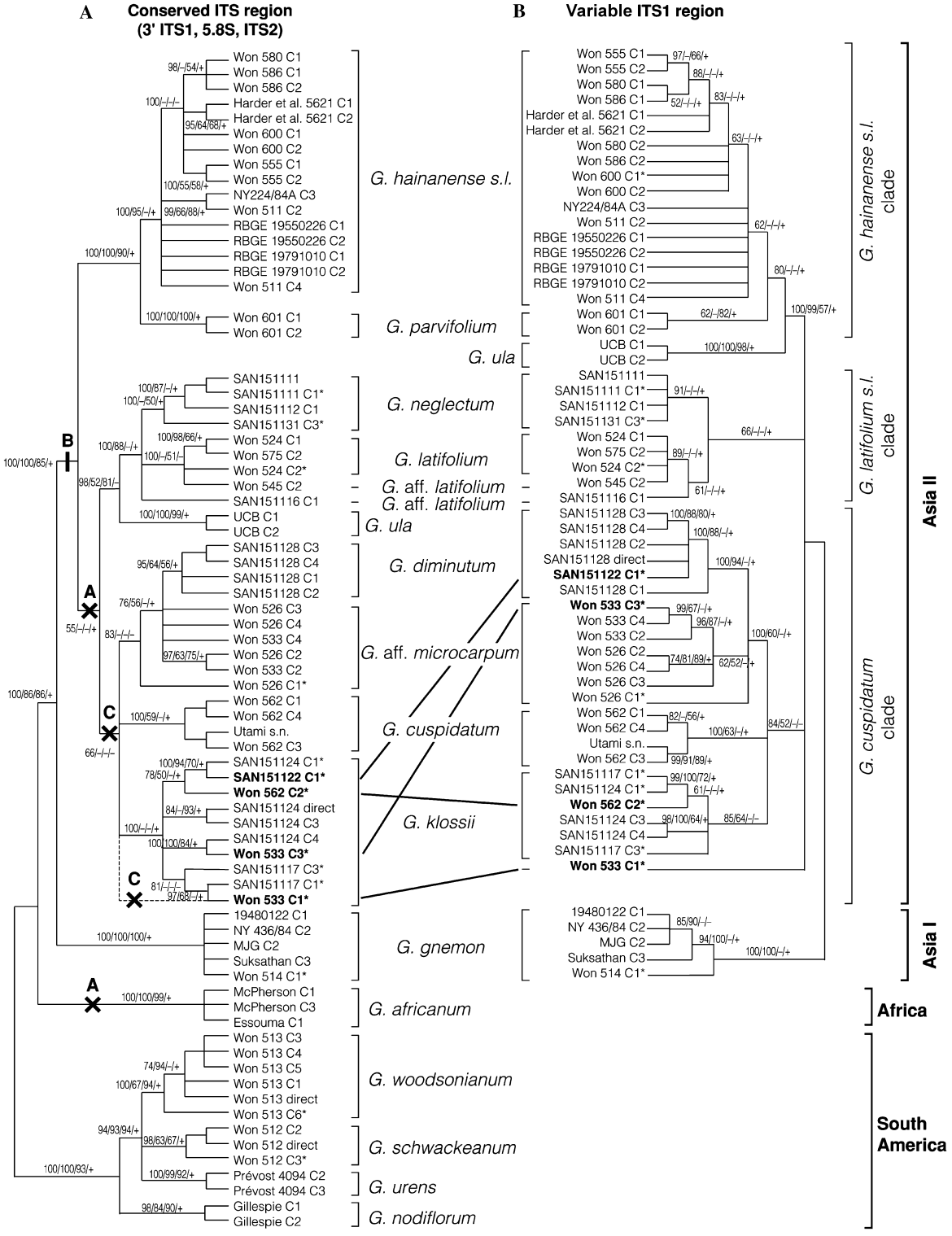

Fig. 3. Phylogeny of *Gnetum* obtained from Bayesian analysis of (A) the conserved sections of the ITS region for all taxa (3′ region of ITS1, 5.8S, and ITS2) and (B) just the variable ITS1 region for the Asian species. Sequences labeled with an asterisk are putative pseudogenes. Values next to nodes are Bayesian posterior probabilities, followed by bootstrap values under neighbor joining, quartet-puzzling support, and presence/absence of the respective node in a strict consensus tree of (A) 118,200 equally parsimonious trees (CI = 0.61, RI = 0.88, excluding uninformative characters) and (B) 23,192 equally parsimonious trees (in 98 islands, CI = 0.73, RI = 0.92, excluding uninformative characters). In tree A, an alternative position found for *G.* aff. *microcarpum* Won 533-C1 is shown by the dashed line. A marks the loss of the 787- to 1060-bp motif; B the gain of the 351- to 750-bp motif; and C the loss of the 1603- to 1641-bp (39 bp) motif, explained by a reduction in a stem-and-loop (Figs. 5C and D).

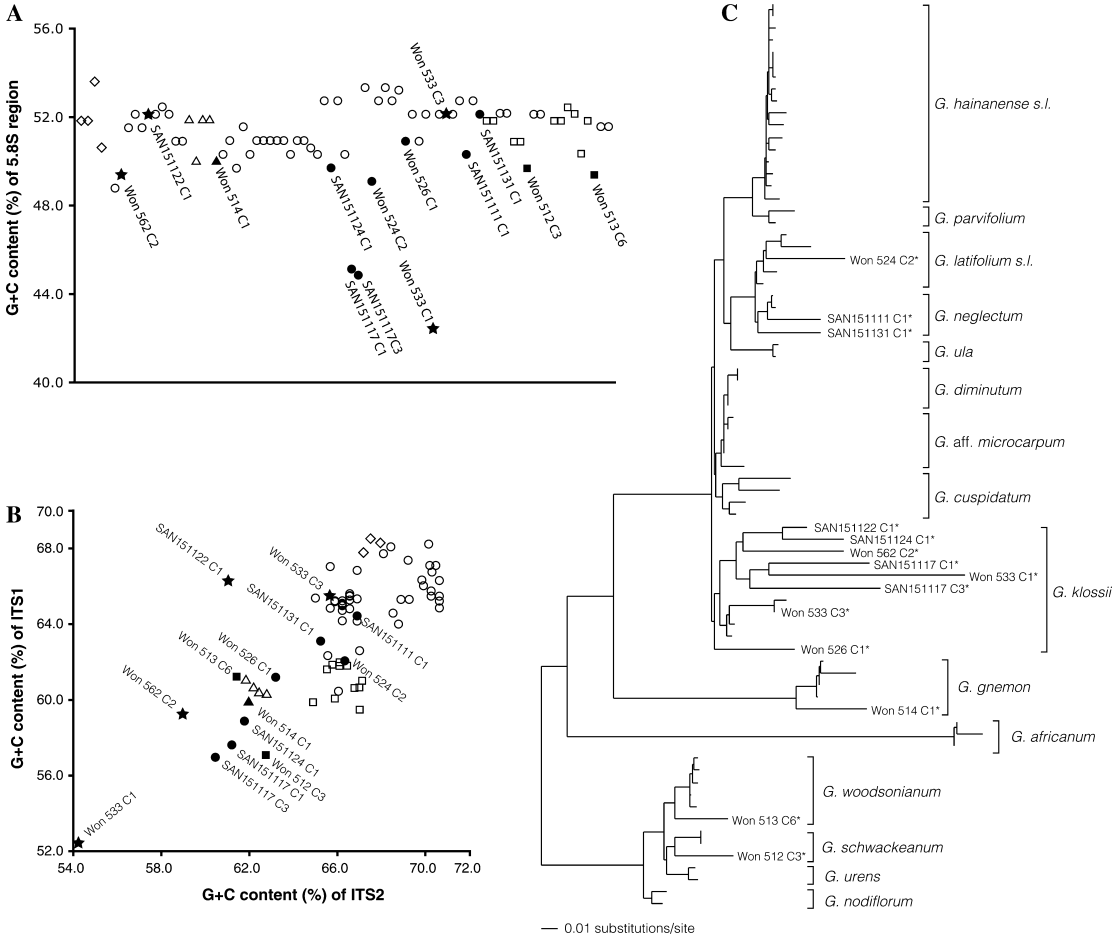

Fig. 4. (A) Plot of the G + C content of the 5.8S region in *Gnetum*. Filled circles and rectangles indicate putative pseudogene sequences; stars indicate sequences that jumped positions between the ITS1 and ITS2 phylogeny; diamonds represent *G. africanum*; triangles *G. gnemon*; open circles, sequences from the Asian clade II; and open rectangles, sequences from the South American clade. (B) Plot of the relative G + C content of ITS1 vs. ITS2. Sequences are labeled as in A. (C) Kimura-2-parameter distance tree for *Gnetum* based on conserved ITS regions. Accession numbers are shown for pseudogenic and chimeric sequences.

Table 2
Tandem repeats found in the *Gnetum* ITS sequences

Taxon	Sequences	Length	Copy numbers	Match point	Location
G. africanum	GACGGGAG	8	3.0	0.70	ITS1
G. cuspidatum	ATACGAGGGCGAC	13	3.0	0.61	ITS1
G. diminutum	ATACGAGGGCGAC	13	2.8	0.61	ITS1
G. aff. *microcarpum*	ATACGAGGGCGAC	13	2.8	0.61	ITS1
G. klossii	CGCCCGT	7	4.3	0.62	ITS1
G. parvifolium	TCCCGTTGTCGT	12	2.7	0.80	ITS1
G. schwackeanum	CGGTGCGATG	10	2.2	0.85	ITS2
G. urens	CGGTGCGATG	10	2.2	0.85	ITS2
G. woodsonianum	CGGTGCGATG	10	2.2	0.85	ITS2
G. woodsonianum	GGCAGTCCGG	10	2.4	0.79	ITS1

See Ben son (1999; http://c3.biomath.mssm.edu/trf.html) for the explanation of copy numbers and match points.

in a polytomy (Fig. 3A, dashed line), while its variable ITS1 region placed it in a large polytomy (Fig. 3B). The G + C content of Won 533-C1 is the lowest obtained from *Gnetum*, even in its 5.8S region, indicating that it is a pseudogene. By contrast, the G + C content of Won 533-C3 (which also jumped position) lies in the range of functional copies (Figs. 4A and B). For SAN151122-C1, the G + C content of ITS2 lies in the pseudogenic range, while that of ITS1 is in the normal range. Another likely pseudogene, Won 562-C2 of *G. cuspidatum*, had ITS1 and ITS2 characteristic of *G. klossii* with which it grouped (Figs. 3A and B), but a low G + C content, especially in its ITS2 region (59.0% vs. 65.6–69.3% in other ITS2 from *G. cuspidatum*; Fig. 4B). Other sequences with low G + C con-

Table 3
Sequence divergences (K2P) for the conserved region of ITS1/5.8S/ITS2 (590 aligned characters) and the ITS2 region (333 aligned characters)

	South America	Africa	SE Asia I clade	SE Asia II clade
Conserved region of ITS1/5.8S/ITS2 ($n = 62$)				
South America ($n = 13$)	0–0.086			
Africa ($n = 3$)	0.333–0.382	0.004–0.016		
SE Asia I clade ($n = 5$)	0.276–0.334	0.388–0.446	0.004–0.079	
SE Asia II clade ($n = 41$)	0.166–0.320	0.255–0.415	0.126–0.326	0–0.123
ITS2 ($n = 72$)				
South America ($n = 13$)	0–0.120			
Africa ($n = 3$)	0.642–0.747	0–0.012		
SE Asia I clade ($n = 5$)	0.456–0.565	0.802–0.893	0.003–0.095	
SE Asia II clade ($n = 51$)	0.273–0.465	0.618–0.842	0.329–0.550	0–0.215

Table 4
Sequence divergences (K2P) for the conserved sequence motifs in the ITS1 region

	South America	Asia I clade	Asia II clade	
Motif between bp 787 and 1060 ($n = 33$)				
South America ($n = 12$)	0–0.119			
Asia I clade ($n = 5$)	0.388–0.504	0–0.076		
Asia II clade ($n = 16$)	0.384–0.465	0.321–0.463	0.004–0.194	
	G. cuspidatum	*G. latifolium* s.l.	*G. hainanense* s.l.	*G. ula*
Motif between bp 351 and 750 ($n = 45$)				
G. cuspidatum subclade ($n = 20$)	0–0.223			
G. latifolium s.l. subclade ($n = 9$)	0.082–0.268	0–0.115		
G. hainanense s.l. subclade ($n = 14$)	0.351–0.704	0.368–0.570	0.007–0.254	
G. ula ($n = 2$)	0.134–0.290	0.142–0.272	0.292–0.457	0.007

tents and divergent ITS1 lengths were SAN151124-C1, SAN151117-C1, and SAN151117-C3, all of *G. klossii*. Their G + C contents in ITS2 were 60.4–61.8% compared to 65.7–68.5% in presumably functional *klossii* ITS2 (Fig. 4B). Even the 5.8S of two of these sequences, SAN151117-C1 and SAN151117-C3, had a significantly low G + C content (Fig. 4A). Thus, they appear to be pseudogenes.

Further sequences with long branches in the NJ tree (Fig. 4C) were *G.* aff. *microcarpum* Won 526-C1, *G. neglectum* SAN151111-C1 and SAN151131-C3, *G. latifolium* Won 524-C2, *G. woodsonianum* Won 513-C3, and *G. schwackeanum* Won 512-C3 (see also Appendix A). Although these sequences represented long branches and their G + C contents were in the pseudogenic range (Figs. 4A and B), they still clustered with conspecific sequences. We detected no pseudogenic copies of the *LEAFY* second intron.

3.5. Secondary structure of Gnetum ITS

As expected from the proximal stem's role in the processing of pre-rRNA (at least in yeast: Côté and Peculis, 2001), we found most of it strongly conserved, especially the distal and medial segments. However, the indels near the 3′ end of *Gnetum* 5.8S and the length variation in ITS2 resulted in inconsistent folding; only the sequences of *G. africanum* and *G. gnemon* formed clear proximal stem segments (Figs. 5A and B). ITS2 contained well-conserved structural motifs consisting of two stem-and-loop structures (Figs. 5C–F). The 39-bp deletion in the *G. cuspidatum* subclade (Fig. 3A, node C) represents a drastic shortening of one of these structures (Figs. 5C and D), and the 9-bp deletion in the pseudogenic sequences of *G. diminutum* and *G. klossii* represents another stem–loop shortening (Figs. 5E and F).

Estimated secondary structures of the ITS1 RNA transcripts were too variable to account for the 351–750 and 787–1060 bp motifs in the ITS1 region (see Section 3.3). Given that software designed to fold RNA molecules into minimum-free-energy configurations can generate quite different structural predictions for the same sequence, we used comparative sequence analysis and tried to find evidence for compensatory base changes (CBC) within our alignments (Campbell et al., 2005; Coleman et al., 1998; Gottschling et al., 2001). However, we were unable to distinguish compensatory mutations from background noise, a statistical problem also encountered by Goertzen et al. (2003). Notably, Campbell et al. (2005) found only a single putative CBC among 24 cloned ITS1 sequences of *Picea*.

3.6. Phylogenetic analyses

All data partitions supported the monophyly of three geographical species groups in *Gnetum* (South America, Africa, and Asia; Figs. 3 and 6). Statistical support for several of the nodes within the Asian clade increased after exclusion of the putative recombinant and paralogous pseudogenes in the *G. klossii* clade (data not shown). Manual alignment of the *LEAFY* sequences was

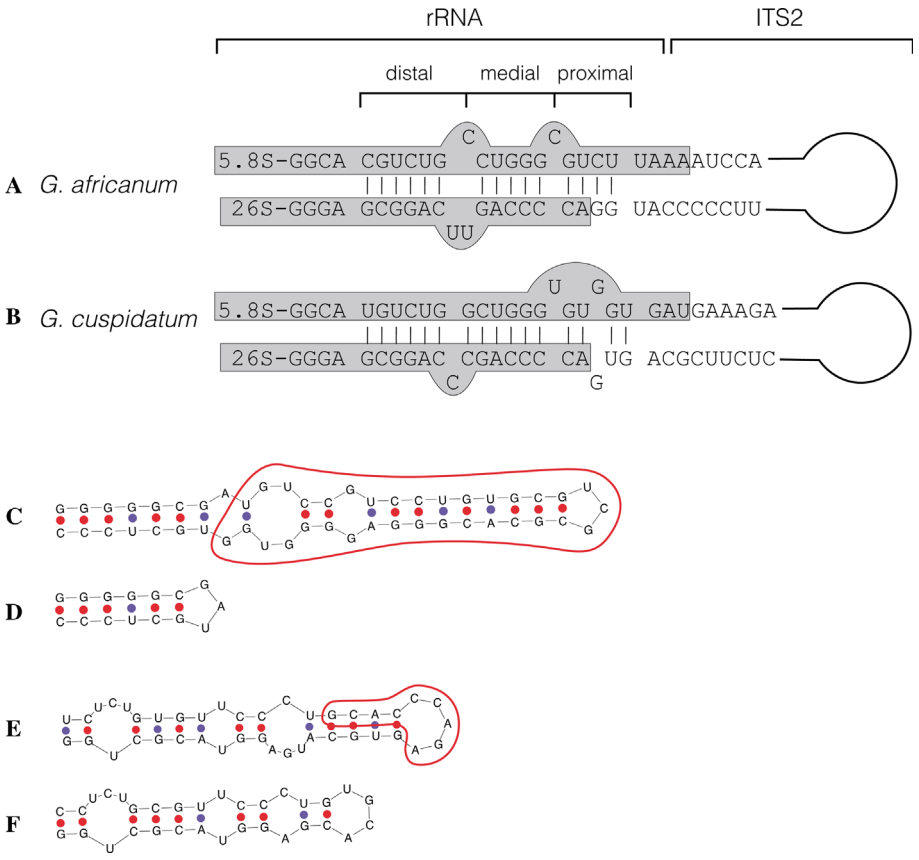

Fig. 5. Structural motifs and changes in their structure predicted by mfold. (A and B) The arbitrary boundaries of 5.8S and 26S rRNA are indicated by shaded boxes. (A and B) The conserved proximal stem segments of *G. africanum* and *G. gnemon*. (C and D) The 39-bp deletion in species of the *G. cuspidatum* subclade (Fig. 3A) likely presents the reduction of a hairpin structure, compared to the predicted secondary structure in other *Gnetum* sequences. (E and F) A 9-bp deletion in *G. diminutum* SAN151122-C1 and *G. klossii* SAN151124-C1 likely represents another hairpin shortening.

straightforward within *Gnetum*, but not between *Gnetum* and *Welwitschia*, resulting in the exclusion of the *Welwitschia LEAFY* intron from phylogenetic analyses. The tree resulting from *LEAFY* was highly congruent with the ITS trees (Figs. 3 and 6, the latter also shows all indels informative within *Gnetum*). Exceptions were a sequence from *G. gnemon* Won 514-C2, which placed with *G. costatum* rather than the second accession of *G. gnemon*, and two sequences of *G. parvifolium*, which placed in a polytomy with *G. hainanense* s.l. Other differences between the *LEAFY* and ITS topologies concerned the placement of *G. klossii* and *G. ula* (Figs. 3 vs. 6). Analysis of combined ITS and *LEAFY* data for 38 representative accessions/species resulted in a generally well-supported topology (Fig. 7) that was congruent with topologies obtained from ITS or *LEAFY*.

4. Discussion

4.1. Gnetum ITS compared to other seed plants

The length variation in ITS1 reported here (225–986 bp) is among the largest observed in seed plants. Only two Pinaceae have similarly large intrageneric length heterogeneity: *Abies* ITS1 varies from 1300 to 2100 bp (Xiang et al., 2000) and *Picea* from 2747 to 3271 bp (Campbell et al., 2005; Maggini et al., 2000). By contrast, *Larix* (1365–1375 bp), *Pseudotsuga* (1170–1177 bp; Gernandt and Liston, 1999), *Pinus* subgenus *Strobus* (2532–2751 bp; Gernandt et al., 2001), and one of *Gnetum*'s closest relative *Ephedra* (1108–1139 bp; Huang et al., 2005) all have relatively conserved ITS1 lengths. Interestingly, the conserved ITS1 motif GGCRY-(4 to 7N)-GYGYCAAGGAA, formerly thought to be angiosperm-specific but now known from *Welwitschia*, *Ephedra fragilis*, *Cycas revoluta*, and *Taxus baccata* (Liston et al., 1996; Liu and Schardl, 1994; Maggini et al., 1998), is absent from *Gnetum*.

It has been suggested that tandem repeats, causing non-homologous recombination and/or unequal crossing-over, are responsible for the great ITS length variation in gymnosperms. Tandem repeats of 68–237 bp occur in several Pinaceae (Campbell et al., 2005; Gernandt et al., 2001), and *Picea abies* has a 480-bp long repeat (Maggini et al., 2000). At least one subsection of *Pinus* also has a 183- to 189-bp motif with a highly conserved 77-bp subrepeat (Gernandt et al., 2001). The ITS1 of *E. fragilis* con-

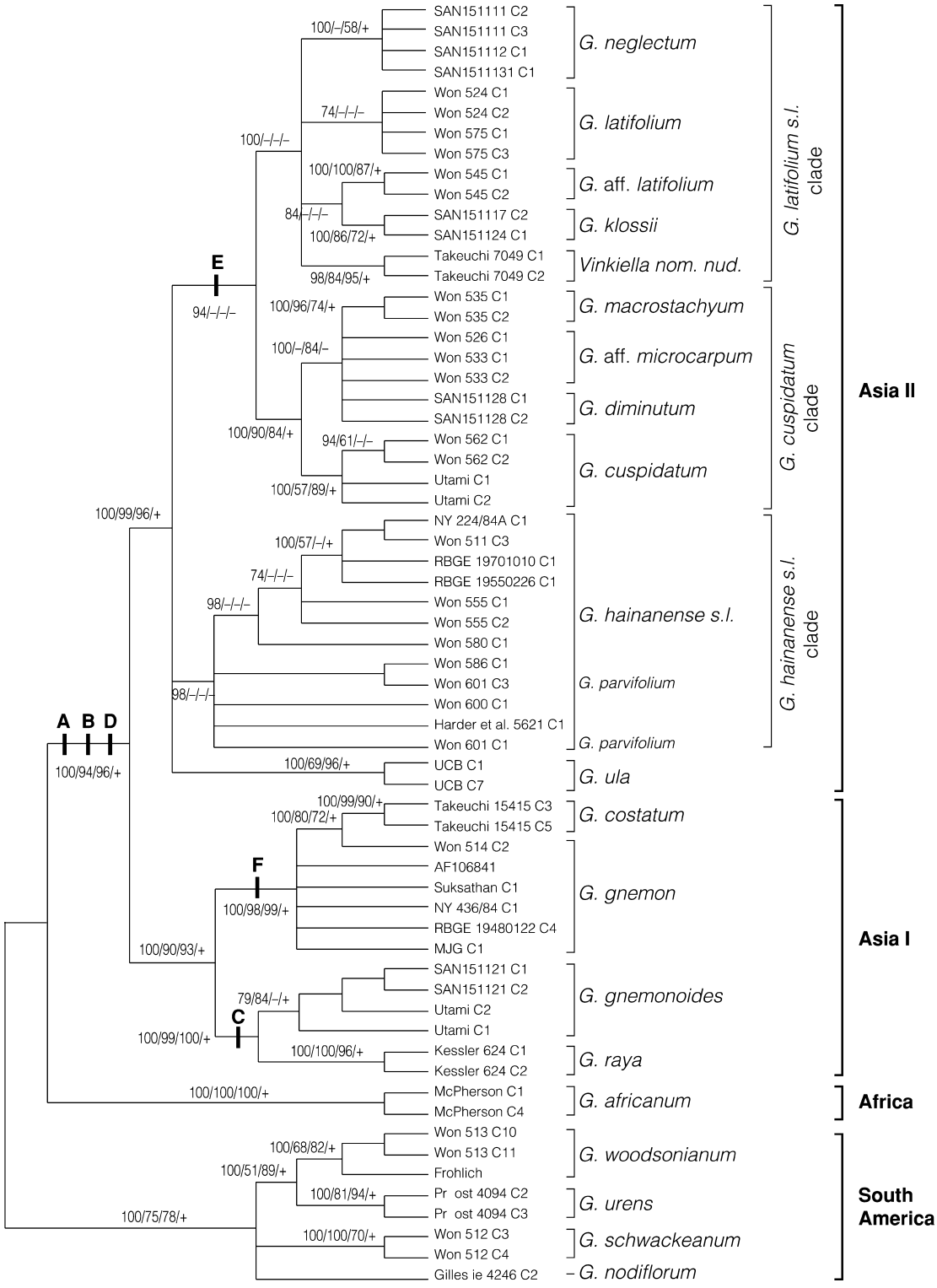

Fig. 6. Phylogeny of *Gnetum* obtained from Bayesian analysis of the second intron and adjacent exons of the nuclear *LEAFY* gene. Values next to nodes are Bayesian posterior probabilities, followed by bootstrap values under neighbor joining, quartet-puzzling support, and presence/absence of the respective node in a strict consensus tree of 108,600 equally parsimonious trees (CI = 0.55, RI = 0.45, excluding uninformative characters). A indicates a gap at bp 440–441; B and C, insertions of "G" and "T" at bp 445, respectively; D a gap at bp 477–482; E a gap at bp 526; and F a gap at bp 686–689.

tains a tandem repeat of 68–71 bp (Maggini et al., 1998), while *Welwitschia* contains none (Liston et al., 1996). In *Gnetum*, we found only short repeats of 7–13 bp (using the most relaxed search options of the Tandem Repeat Finder program) and only in about half of the species (Table 2). With one exception, repeats were limited to single species.

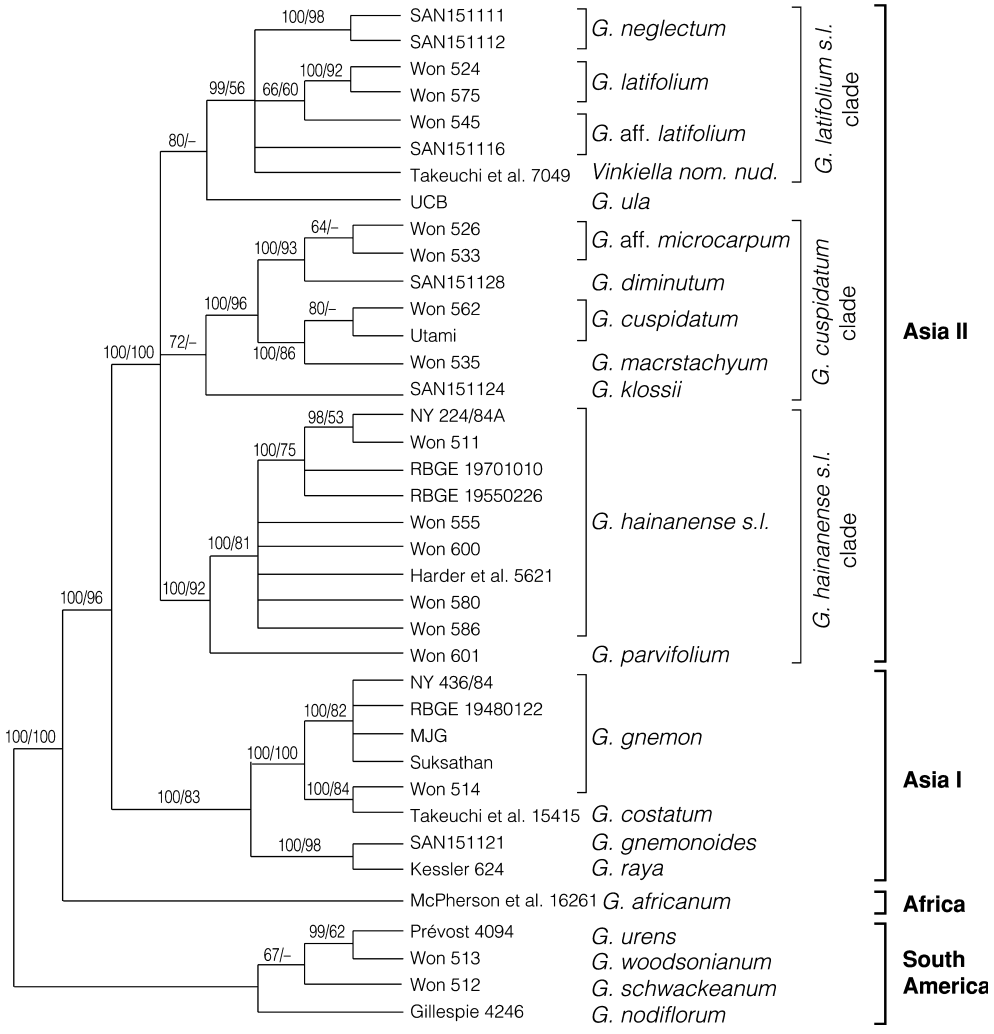

Fig. 7. Phylogeny of *Gnetum* obtained from Bayesian analysis of combined ITS and *LEAFY* sequences. Values next to nodes are Bayesian posterior probabilities, followed by bootstrap values under parsimony, which yielded 37,742 trees (CI = 0.67 and RI = 0.86, excluding uninformative characters).

Tandem repeats therefore are not the cause of the observed huge ITS length variation in *Gnetum*.

Persistent ITS polymorphism can be due to several biological phenomena. Among them are allopolyploid genomes (Muir et al., 2001; Volkov et al., 1999; Wendel et al., 1995), numerous nucleolar organizer regions (Bobola et al., 1992; Karvonen and Savolainen, 1993), and highly divergent ITS pseudogenes no longer capable of undergoing normal concerted evolution (Buckler et al., 1997; Muir et al., 2001; Razafimandimbison et al., 2004). Nothing is known about the number of nucleolar organizers in *Gnetum*, but speculation about the other two factors is appropriate. The three species of *Gnetum* whose chromosomes have been counted (*G. costatum*, *G. gnemon*, and *G. montanum*) have $2n = 44$ (Fagerlind, 1941; Hizume et al., 1993). Fagerlind suggested that *G. gnemon* might be allotetrapolyploid because he observed 11 larger and 11 smaller bivalents. The closest relatives of *Gnetum* also have high chromosome numbers relative to other gymnosperms, namely $2n = 28$ in *Ephedra* (Choudry, 1984; Florin, 1932), a genus in which as many as 40% of the species appear to be polyploid (Delevoryas, 1980); *W. mirabilis* has $2n = 42$ (Fernandes, 1936; Florin, 1932). Whether these numbers reflect ancient or recent polyploidy, perhaps in connection with hybridization, is unknown in *Gnetum*, but in *Ephedra* interspecific hybridization is well documented (Ickert-Bond and Wojciechowski, 2004; Wendt, 1993). Allopolyploidy may thus partly explain the observed high levels of ITS polymorphism in *Gnetum*. This study has also documented the existence of ITS pseudogenes in some species, and perhaps they are divergent enough to escape normal concerted evolution.

4.2. Ribosomal pseudogenes in Gnetum

Recombinant ITS copies, such as discovered here, may originate from simple crossing-over, more complex gene conversion followed by multiple crossing-over, or by PCR mediation (Buckler et al., 1997; Cronn et al., 2002). Of the recombinant sequences in *Gnetum*, one (SAN151122-C1)

has a normal G + C content in ITS1 and 5.8S, but a lower than normal G + C content in ITS2, while others had normal G + C contents in both regions. This suggests that at least SAN151122-C1 was formed by recombination of a functional ITS1 with a paralogous pseudogenic ITS2. Other recombinant ITS sequences appear to unite functional copies of different origin. There were no consistent length differences between pseudogenes and presumably functional sequences (as judged by G + C content).

Pseudogenic and chimeric sequences can result from introgression between species, with one parental genome mainly transcriptionally active and that of the other parent gradually transformed into pseudogenes, or from shared ancestral variation before the divergence of species (i.e., incomplete lineage sorting). In rare cases, one or the other explanation is favored by topological differences between nuclear and chloroplast tree topologies that suggest different uniparental vs. biparental histories (but see Vollmer and Palumbi, 2004). This is not the case in *Gnetum*, however. Nuclear ITS and *LEAFY* yielded essentially the same topology as obtained from chloroplast regions (Won and Renner, submitted-b). The mode of chloroplast inheritance in *Gnetum* and *Welwitschia*, its sister genus, is unknown; *Ephedra*, which is sister to both, has maternal chloroplast inheritance (Mogensen, 1996). Assuming maternal chloroplast inheritance also in *Gnetum*, introgression from a pollen donor from the *G. latifolium* subclade into a *G. klossii* mother would explain why *G. klossii* in chloroplast trees (Won and Renner, submitted-b) places in the *G. cuspidatum* clade while in the nuclear *LEAFY* tree (Fig. 6) it places in the *G. latifolium* clade. However, in the nuclear ITS tree (Fig. 3) it also placed in the *G. cuspidatum* clade. Perhaps we detected only the paternal *LEAFY* allele, but missed the paternal ITS allele. It is equally easy to imagine, however, that *G. klossii* genome might have introgressed into the *G. cuspidatum* clade because members of this clade have partly overlapping ranges, for example in Borneo, on the Malay Peninsula, and on Sumatra. Bornean *G. diminutum* and Malayan *G.* aff. *microcarpum* are genetically close (Figs. 3 and 6) and share morphological similarities, such as lanceolate leaves, fruiting axes with short internodes, and small seeds, suggesting that they may have evolved from each other, with the introgression of *klossii* genes predating the divergence of *microcarpum.* In either case, concerted evolution in *G. cuspidatum* clade so far has not prevented continued presence of chimeric sequences.

Chimeric sequences can also result from cross-contamination of DNA templates, mediated by PCR (Buckler et al., 1997; Cronn et al., 2002); *Taq* enzyme error is estimated as 7.2×10^{-5}–2×10^{-4} (Cha and Thilly, 1995). Because we cloned most PCR products (see Section 3.1), any base wrongly incorporated by the *Taq* enzyme would have been propagated. The low sequence divergence detected among multiple accessions of the same species (maximally 0.080), however, indicates that this problem was limited, although its scope is impossible to assess in a non-coding region. A likely *Taq* error is the single-base deletion we found in the *LEAFY* exon.

Adding the denaturant DMSO to the PCRs had no consistent effect on PCR results, although all putative pseudogenes came from samples amplified without DMSO. This agrees with findings in *Gossypium* and Winteraceae, where the presence of DMSO interfered with PCR, while in *Nicotiana* and *Tripsacum* it helped amplification of ITS alleles, some with high secondary structure stability and low free energy, others with low stability and high free energy (Buckler et al., 1997).

The absence of any *Eco*RI restriction sites in the 5.8S rRNA of *Gnetum* indicates that our sequences were not the result of contamination with fungal DNA. Fungi and non-seed plants have extra *Eco*RI restriction sites in their 5.8S rRNA (Jobes and Thien, 1997).

4.3. Timing of ITS evolution within Gnetum

Gnetum shows the greatest length variation in ITS sequences so far reported from any seed plant genus. This raises the questions of how this variation formed and for how long it has persisted. As argued in the preceding section, tandem repeats do not appear to have played a great role in *Gnetum* ITS evolution. Other structural changes, by contrast, seem to have been important, as exemplified by the shortened stem-and-loop structures (Figs. 5D–F). Loss of sequence motifs and presence of indels played a role particularly in the evolution of ITS1 (summary: Fig. 2). An example is the apparently independent loss of the 787- to 1060-bp motif from the African clade and within the Asia II clade (the two nodes marked A in Fig. 3A; although one of this two nodes has little statistical support in the tree shown here, both are solidly supported by chloroplast intron and spacer sequences; Won and Renner, submitted-b).

The ancestral *Gnetum* ribosomal DNA region likely contained the 787- to 1060-bp motif in the ITS1 5′ region and the conserved 3′ ITS1, 5.8S, and ITS2 regions. It then appears to have evolved mainly via loss of sequence and structural motifs, high rates of substitution, and hybridization or introgression (compare the indirect evidence for allotetraploidy in *G. gnemon* cited above [Fagerlind, 1941]). A new ITS allele is expected to go to fixation by concerted evolution, a process that can take 1–2 million years or just a few generations, e.g., in herbaceous cotton and tobacco (Senchina et al., 2003; Skalická et al., 2003; Wendel et al., 1995). In *Gnetum*, a chloroplast gene molecular clock calibrated with gnetalean macrofossils suggests that extant *Gnetum* is young and that the Asian clade, which contains all paralogous ITS copies discovered, is only a few million years old (Won and Renner, 2003, submitted-a). Perhaps this time has been insufficient for complete concerted evolution and elimination of ancestral

polymorphisms; alternatively, hybridization may be ongoing among the Southeast Asian species.

Acknowledgments

We thank M. Frohlich for *LEAFY* sequences and primers; H. Meimberg and two anonymous reviewers for comments on the manuscript; and the American Society of Plant Taxonomy, the International Association of Plant Taxonomy, and the International Center for Tropical Ecology at University of Missouri-St. Louis (UM-SL) for financial support. This paper represents a portion of a dissertation submitted by the first author in partial fulfillment of the requirements for a Ph.D. degree from UM-SL.

Appendix A

Species sampled, vouchers, GenBank accession numbers, and summary for length (bp) and G + C (%) content for nuclear ribosomal ITS sequences

Taxon	Voucher information	Clone number	GenBank Accession No.	ITS1 length (GC)	ITS2 length (GC)	5.8S length (GC)
G. africanum Welw.	McPherson et al. 16261 (MO)	C1	AY449543	410 (68.3)	259 (68.0)	164 (51.8)
	McPherson et al. 16261 (MO)	C2	AY449545*	—	261 (68.2)	97 (53.6)*
	McPherson et al. 16261 (MO)	C3	AY449544	410 (68.5)	259 (67.5)	164 (51.8)
	Essouma 99 (MO)	C1	AY449546	410 (67.8)	259 (67.2)	164 (50.6)
G. cuspidatum Bl.	Won 562 (MO)	C1	AY449547	392 (60.5)	270 (66.0)	165 (48.8)
	Won 562 (MO)	C2	AY449548[a]	395 (59.2)	268 (59.0)	165 (49.4)
	Won 562 (MO)	C3	AY449550	392 (65.3)	270 (69.3)	165 (52.1)
	Won 562 (MO)	C4	AY449551	393 (62.3)	270 (65.6)	165 (51.5)
	Utami s.n. (BO)	Direct	AY449549	392 (65.3)	270 (68.9)	165 (51.5)
G. diminutum F. Markgraf	Postar & Won SAN151122	C1	AY449552[b]	368 (66.3)	259 (61.0)	165 (52.1)
	Postar & Won SAN151128	Direct	AY449553*	368 (66.0)	206 (69.9)*	165 (52.1)
	Postar & Won SAN151128	C1	AY449554	370 (64.9)	269 (70.6)	165 (52.5)
	Postar & Won SAN151128	C2	AY449555	368 (66.3)	269 (70.6)	165 (52.1)
	Postar & Won SAN151128	C3	AY449556	368 (65.5)	269 (70.6)	165 (50.9)
	Postar & Won SAN151128	C4	AY449557	368 (65.5)	269 (70.6)	165 (50.9)
G. gnemon L.	RBGE 19480122 (E)	C1	AY449558	986 (60.6)	299 (62.2)	162 (51.9)
	NY 436/84 (MO)	C1	AY449559	980 (59.9)	302 (61.9)	162 (50.0)
	Suksathan (QSB)	C3	AY449560	984 (60.4)	301 (62.5)	162 (51.9)
	U. Mainz Bot. Gard. (MJG)	C2	AY449561	985 (60.3)	301 (62.8)	162 (51.9)
	Won 514 (MO)	C1	AY449562[a]	919 (61.0)	291 (61.9)	164 (50.0)
G. hainanense s.l.	RBGE 19550226 (E)	C1	AY449563	790 (64.2)	299 (66.2)	161 (50.3)
	RBGE 19550226 (E)	C2	AY449564	790 (65.2)	299 (65.9)	161 (50.9)
	NY 224/84A (MO)	C3	AY449565	794 (65.2)	299 (65.9)	161 (49.7)
	RBGE 19791010 (E)	C1	AY449566	792 (64.7)	299 (66.2)	161 (51.6)
	RBGE 19791010 (E)	C2	AY449567	790 (64.8)	299 (66.6)	161 (50.3)
	Won 511 (MO)	C2	AY449568	791 (65.2)	299 (66.2)	161 (50.9)
	Won 511 (MO)	C4	AY449569	791 (65.0)	299 (66.2)	161 (50.9)
	Won 555 (MO)	C1	AY449570	746 (65.3)	299 (66.6)	161 (50.9)
	Won 555 (MO)	C2	AY449571	746 (65.0)	299 (66.2)	161 (50.9)
	Won & Maxwell 580 (MO)	C1	AY449572	745 (64.2)	299 (66.9)	161 (50.9)
	Won & Maxwell 580 (MO)	C2	AY449573	745 (65.1)	299 (66.2)	161 (50.3)
	Won & Maxwell 586 (MO)	C1	AY449574	756 (65.3)	299 (66.9)	161 (50.9)
	Won & Maxwell 586 (MO)	C2	AY449575	756 (65.5)	299 (66.6)	161 (50.9)
	Won & Gang 600 (MO)	C1	AY449576	802 (62.6)	300 (67.0)	161 (50.6)
	Won & Gang 600 (MO)	C2	AY449577	753 (65.1)	299 (66.2)	161 (50.3)
G. klossii Merr. ex F. Markgraf	Postar & Won SAN151117	C1	AY449582[a]	394 (57.6)	268 (61.2)	164 (45.1)
	Postar & Won SAN151117	C3	AY449583[a]	395 (57.0)	268 (60.4)	165 (44.8)
	Postar & Won SAN151124	Direct	AY449578*	135 (64.6)*	268 (68.5)	165 (52.7)
	Postar & Won SAN151124	C1	AY449579[a]	394 (58.9)	259 (61.8)	165 (49.7)
	Postar & Won SAN151124	C3	AY449580	340 (68.2)	268 (70.2)	165 (52.7)
	Postar & Won SAN151124	C4	AY449581	352 (67.0)	268 (65.7)	165 (50.3)
G. latifolium Bl. s.l.	Won 524 (MO)	C1	AY449584	374 (67.4)	302 (69.2)	165 (53.3)
	Won 524 (MO)	C2	AY449585[a]	377 (62.1)	303 (66.3)	165 (49.1)
	Won 575 (MO)	C2	AY449586	380 (66.8)	302 (66.9)	165 (52.7)

(*continued on next page*)

Appendix A (*continued*)

Taxon	Voucher information	Clone number	GenBank Accession No.	ITS1 length (GC)	ITS2 length (GC)	5.8S length (GC)
G. aff. *latifolium* Bl.	Won 545 (MO)	C2	AY449587	397 (66.8)	303 (70.3)	165 (53.3)
G. aff. *latifolium* Bl.	Postar & Won SAN151116	C1	AY449588	404 (66.3)	305 (69.8)	165 (52.7)
G. aff. *microcarpum* Bl.	Won 526 (MO)	C1	AY449589[b]	384 (61.2)	269 (63.2)	165 (50.9)
	Won 526 (MO)	C2	AY449590	368 (65.5)	269 (70.6)	165 (52.1)
	Won 526 (MO)	C3	AY449591	375 (64.0)	269 (68.8)	165 (50.9)
	Won 526 (MO)	C4	AY449592	368 (65.5)	269 (70.3)	165 (52.1)
	Won 533 (MO)	C1	AY449593[a]	405 (52.3)	301 (54.2)	165 (42.4)
	Won 533 (MO)	C2	AY449594	368 (65.2)	269 (70.6)	165 (52.1)
	Won 533 (MO)	C3	AY449595[b]	368 (65.5)	268 (65.7)	165 (52.1)
	Won 533 (MO)	C4	AY449596	368(65.8)	269 (70.3)	165 (52.1)
G. neglectum Bl.	Postar & Won SAN151111 (MO)	Direct	AY449597	225 (67.1)	302 (70.2)	165 (52.7)
	Postar & Won SAN151111 (MO)	C1	AY449598[a]	225 (64.4)	302 (66.9)	165 (50.3)
	Postar & Won SAN151112 (MO)	C1	AY449599	225 (67.1)	302 (70.5)	165 (52.7)
	Postar & Won SAN151131 (MO)	C1	AY449600[a]	225 (63.1)	302 (65.2)	165 (52.1)
G. nodiflorum Brongn.	Gillespie 4246 (US)	C1	AY449601	447 (60.6)	292(66.8)	164 (51.8)
	Gillespie 4246 (US)	C2	AY449602	449 (61.0)	295 (67.1)	164 (51.8)
G. parvifolium (Warb.) Cheng ex Chun	Won & Gang 601 (MO)	C1	AY449603	569 (65.4)	300 (65.0)	161 (52.2)
	Won & Gang 601 (MO)	C2	AY449604	569 (64.9)	300 (65.7)	161 (52.2)
G. aff. *schwackeanum* Taub. ex F. Markgraf	Won 512 (MO)	Direct	AY449605	521 (60.7)	297 (67.0)	165 (50.9)
	Won 512 (MO)	C2	AY449606	521 (59.5)	297 (67.0)	165 (50.9)
	Won 512 (MO)	C3	AY449607[a]	517 (57.1)	298 (62.8)	165 (49.7)
G. ula Brongn.	Maheshwari s.n. (UCB)	C1	AY449608	727 (68.1)	304 (68.4)	165 (52.1)
	Maheshwari s.n. (UCB)	C2	AY449609	719 (67.7)	304 (68.1)	165 (52.1)
G. urens (Aubl.) Bl.	Prévost 4094 (MO)	C2	AY449610	516 (60.1)	299 (65.5)	164 (51.8)
	Prévost 4094 (MO)	C3	AY449611	516 (59.9)	299 (64.9)	164 (51.8)
G. woodsonianum (F. Markgraf) H. Won *stat. nov.*	Won 513 (MO)	Direct	AY449612	513 (61.8)	295 (66.4)	164 (51.8)
	Won 513 (MO)	C1	AY449613	513 (61.8)	295 (66.1)	164 (52.4)
	Won 513 (MO)	C3	AY449614	514 (61.9)	295 (65.8)	164 (52.1)
	Won 513 (MO)	C4	AY449615	514 (62.0)	295 (66.1)	164 (50.3)
	Won 513 (MO)	C5	AY449616	513 (61.6)	296 (65.5)	164 (51.8)
	Won 513 (MO)	C6	AY449617[a]	513 (61.2)	293 (61.4)	164 (49.4)
G. sp.	Harder et al. 5621 (MO)	C1	AY449618	756 (65.6)	299 (66.6)	161 (51.6)
	Harder et al. 5621 (MO)	C2	AY449619	754 (65.5)	299 (66.6)	161 (51.6)

[a] A putative pseudogene.
[b] A recombinant pseudogene.
[*] An incomplete sequence.

Appendix B

Species sampled, vouchers, GenBank accession numbers, lengths, and G + C contents (in parentheses) of nuclear *LEAFY* second intron and adjacent partial exons

Taxon	Voucher/source	Clone	GenBank Accession No.	Intron length (G + C)	Exon length (G + C)	Notes
Gnetaceae						
Gnetum africanum Welw.	McPherson et al. 16261 (MO)	C1	AY296575	535 (43.6)	230 (48.7)	
		C4	AY296576	534 (43.6)	230 (48.7)	
G. costatum K. Schum.	Takeuchi et al. 15415 (MO)	C3	AY296577	525 (44.4)	230 (47.0)	
		C5	AY296578	525 (44.2)	230 (47.8)	
G. cuspidatum Bl.	Utami s.n. (BO)	C1	AY296579	528 (43.8)	230 (48.3)	
		C2	AY296580	528 (43.6)	230 (47.8)	
	Won 562 (MO)	C1	AY296581	527 (43.8)	230 (46.5)	
		C2	AY296582	528 (43.8)	230 (47.4)	
G. diminutum F. Markgraf	Postar & Won SAN 151128 (MO)	C1	AY296583	528 (43.4)	230 (48.3)	
		C2	AY296584	526 (43.5)	230 (47.8)	

Appendix B (*continued*)

Taxon	Voucher/source	Clone	GenBank Accession No.	Intron length (G + C)	Exon length (G + C)	Notes
G. gnemon L.	Frohlich & Meyerowitz (1997)	—	AF106841	525 (44.0)	230 (45.7)	
	RBGE 19480122 (E)	C4	AY296585	525 (44.0)	230 (48.3)	
	NY 436/84 (MO)	C1	AY296586	525 (44.2)	230 (47.4)	
	Mainz Botanical Garden (MJG)	C1	AY296587	525 (44.0)	230 (47.8)	
	Suksathan s.n. (QSB)	C1	AY296588	525 (44.2)	230 (48.7)	
	Won 514 (MO)	C1	AY296589	525 (44.5)	230 (48.7)	
G. gnemonoides Brongn.	Utami s.n. (BO)	C1	AY296590	531 (44.2)	230 (48.7)	
		C2	AY296591	531 (43.7)	230 (48.3)	
	Postar & Won SAN151121 (MO)	C1	AY296592	531 (43.7)	230 (46.5)	
		C2	AY296593	531 (43.7)	230 (47.1)	
G. hainanense s.l.	RBGE 19550226 (E)	C1	AY296594	529 (44.8)	229 (49.3)	
	NY 224/84A (MO)	C1	AY296595	529 (44.6)	230 (48.3)	
	RBGE 19791010 (E)	C1	AY296596	529 (44.6)	230 (47.8)	"*G. montanum*"
	Won 511 (MO)	C3	AY296597	529 (44.4)	230 (47.8)	
	Won 555 (MO)	C1	AY296600	529 (44.8)	230 (47.4)	
		C2	AY296601	529 (45.0)	230 (47.4)	
	Won & Maxwell 580 (MO)	C1	AY296598	529 (44.4)	230 (48.3)	
	Won & Maxwell 586 (MO)	C1	AY296599	529 (44.6)	230 (48.7)	
	Won & Gang 600 (MO)	C1	AY296602	529 (44.2)	230 (46.5)	
G. klossii Merrill ex F. Markgraf	Postar & Won SAN 151117 (MO)	C2	AY296603	528 (44.3)	230 (47.8)	
	Postar & Won SAN 151124 (MO)	C1	AY296604	528 (44.5)	229 (48.0)	
G. latifolium Bl. s.l.	Won 524 (MO)	C1	AY296605	528 (44.5)	229 (47.2)	
		C2	AY296606	528 (44.7)	230 (47.8)	
	Won et al. 575 (MO)	C1	AY296607	529 (44.2)	230 (47.8)	
		C3	AY296608	528 (44.7)	230 (49.1)	
G. aff. *latifolium* Bl.	Won 545 (MO)	C1	AY296609	532 (44.0)	230 (48.3)	
		C2	AY296610	532 (44.0)	230 (48.3)	
G. macrostachyum Hook. f.	Won 535 (MO)	C1	AY296612	530 (43.4)	230 (47.0)	
		C2	AY296613	530 (43.6)	230 (46.5)	
G. aff. *microcarpum* Bl.	Won 526 (MO)	C1	AY296614	528 (43.8)	230 (49.1)	
	Won 533 (MO)	C1	AY296615	528 (43.9)	230 (47.8)	
		C2	AY296616	527 (43.3)	230 (47.4)	
G. neglectum Bl.	Postar & Won SAN151111 (MO)	C2	AY296617	528 (45.3)	230 (48.3)	
		C3	AY296618	528 (44.7)	230 (47.8)	
	Postar & Won SAN151112 (MO)	C1	AY296619	528 (44.7)	230 (48.3)	
	Postar & Won SAN 151131 (MO)	C1	AY296620	523 (45.1)	230 (48.3)	
G. nodiflorum Brongn.	Gillespie 4246 (US)	C2	AY296621	536 (44.4)	230 (47.4)	
G. parvifolium (Warb.) Cheng ex Chun	Won & Gang 601 (MO)	C1	AY296622	529 (45.0)	230 (47.4)	
		C3	AY296623	529 (44.8)	230 (47.8)	
G. raya F. Markgraf	Kessler et al. 624 (MO)	C1	AY296624	531 (43.7)	230 (47.8)	
		C4	AY296625	531 (43.9)	230 (47.8)	
G. aff. *schwackeanum* Taub. ex F. Markgraf	Won 512 (MO)	C3	AY296626	536 (43.7)	230 (47.4)	
		C4	AY296627	536 (43.7)	230 (47.4)	
G. ula Brongn.	Maheshwari s.n. (UCB)	C1	AY296628	512 (44.5)	230 (47.8)	
		C7	AY296629	513 (44.8)	230 (47.8)	
G. urens (Aubl.) Bl.	Prévost 4094 (MO)	C2	AY296630	536 (44.2)	230 (47.0)	
		C3	AY296631	536 (44.4)	230 (47.4)	
G. woodsonianum (F. Markgraf) H. Won *stat. nov.*	Frohlich & Meyerowitz (1997)	—	—	534 (43.8)	230 (45.2)	"*G. leyboldii*"
	Won 513 (MO)	C10	AY296632	534 (43.3)	230 (47.4)	
		C11	AY296633	534 (43.6)	230 (47.8)	
G. sp.	Harder et al. 5621 (MO)	C3	AY296634	531 (44.6)	230 (47.4)	
Vinkiella R. Johns *nom. nud.*	Takeuchi et al. 7049 (MO)	C1	AY296635	528 (44.9)	230 (47.4)	
		C2	AY296636	528 (44.5)	230 (48.7)	
Welwitschiaceae						
Welwitschia mirabilis Hook. f.	Frohlich & Meyerowitz (1997)	—	AF109130	491 (38.7)	230 (44.8)	
	Olson s.n. (MO)	C1	AY296637	491 (38.7)	230 (46.5)	

References

Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29, 417–434.

Bailey, C.D., Carr, T.G., Harris, S.A., Hughes, C.E., 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Mol. Phylogenet. Evol. 29, 435–455.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., Donoghue, M.J., 1995. The ITS regions of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Mo. Bot. Gard. 82, 247–277.

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580.

Blattner, F.R., 1999. Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR. Biotechnology 27, 1180–1185.

Bobola, M.S., Smith, D.E., Klein, A.S., 1992. Five major nuclear ribosomal DNA repeats represent a large and variable fraction of the genomic DNA of *Picea rubens* and *P. mariana*. Mol. Biol. Evol. 13, 612–622.

Buckler IV, E.S., Ippolito, A., Holtsford, T.P., 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. Genetics 145, 821–832.

Cha, R.S., Thilly, W.G., 1995. Specificity, efficiency, and fidelity of PCR. In: Dieffenbach, C.W., Dreksler, G.S. (Eds.), PCR Primer: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 37–51.

Campbell, C.S., Wright, W.A., Cox, M., Vining, T.F., Smoot Major, C., Arsenault, M.P., 2005. Nuclear ribosomal DNA internal transcribed spacer 1 (ITS1) in *Picea* (Pinaceae): sequence divergence and structure. Mol. Phylogenet. Evol. 35, 165–185.

Choudry, A.S., 1984. Karyomorphological and cytogenetical studies in *Ephedra*. J. Sci. Hiroshima Univ. Ser. B. Div. 2 (19), 57–109.

Coleman, A.W., Preparata, R.M., Mehrotra, B., Mai, J.C., 1998. Derivation of the secondary structure of the ITS-1 transcript in Volvocales and its taxonomic correlations. Protist 149, 135–146.

Côté, C.A., Peculis, B.A., 2001. Role of the ITS2-proximal stem and evidence for indirect recognition of processing sites in pre-rRNA processing in yeast. Nucleic Acids Res. 29, 2106–2116.

Cronn, R.C., Cedroni, M., Haselkorn, T., Grover, C., Wendel, J.F., 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. Theor. Appl. Genet. 104, 482–489.

Delevoryas, T., 1980. Polyploidy in gymnosperms. In: Lewis, W.H. (Ed.), Polyploidy: Biological Relevance. Plenum Press, New York, pp. 215–218.

Fagerlind, F., 1941. Bau und Entwicklung der *Gnetum*-Gametophyten. Kongl. Svenska Vetensk. Akad. Handl. 19, 1–55.

Fernandes, A., 1936. Sur la caryologie de *Welwitschia mirabilis* Hook. f. Bol. Soc. Brot. 11, 267–280.

Florin, R., 1932. Die Chromosomenzahlen bei *Welwitschia* und einigen *Ephedra* Arten. Svensk Bot. Tidskr. 26, 205–214.

Frohlich, M.W., Meyerowitz, E.M., 1997. The search for homeotic gene homologs in basal angiosperms and Gnetales: a potential new source of data on the evolutionary origin of flowers. Int. J. Plant Sci. 158, S131–S142.

Frohlich, M.W., Parker, D.S., 2000. The mostly male theory of flower evolutionary origins: from genes to fossils. Syst. Bot. 25, 155–170.

Fu, L.-K., Yu, Y., Gilbert, M.G., 1999. Gnetaceae. In: Wu, Z.-Y., Raven, P.H. (Eds.), Flora of China. Vol. 4. Cycadaceae through Fagaceae. Missouri Botanical Garden Press, St. Louis, pp. 102–105.

Gernandt, D.S., Liston, A., 1999. Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga*. Am. J. Bot. 86, 711–723.

Gernandt, D.S., Liston, A., Piñero, D., 2001. Variation in the nrDNA ITS of *Pinus* subsection *Cembroides*: implication for molecular systematic studies of pine species complexes. Mol. Phylogenet. Evol. 21, 449–467.

Goertzen, L.R., Cannone, J.J., Gutell, R.R., Jansen, R.K., 2003. ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the Asteraceae. Mol. Phylogenet. Evol. 29, 216–234.

Gottschling, M., Hilger, H.H., Wolf, M., Diane, N., 2001. Secondary structure of the ITS1 transcript and its application in a reconstruction of the phylogeny of Boraginales. Plant Biol. 3, 629–636.

Gutell, R.R., 1993. Collections of small subunit (16S- and 16S-like) ribosomal RNA structures. Nucleic Acids Res. 21, 3051–3054.

Hershkovitz, M.A., Lewis, L.A., 1996. Deep-level diagnostic value of rDNA-ITS region. Mol. Biol. Evol. 13, 1276–1295.

Hershkovitz, M.A., Zimmer, E.A., 1996. Conservation patterns in angiosperm rDNA ITS2 sequences. Nucleic Acids Res. 24, 2857–2867.

Hershkovitz, M.A., Zimmer, E.A., Hans, W.J., 1999. Ribosomal DNA sequences and angiosperm systematics. In: Hollingsworth, P.M., Bateman, R.M., Gornall, R.J. (Eds.), Molecular Systematics and Plant Evolution. Taylor and Francis, London, pp. 268–326.

Hizume, M., Gu, Z., Yue, Z., Kondo, K., 1993. Karyomorphological study of *Gnetum montanum* collected in Yunnan, China. CIS Chromosome Inform. Serv. 54, 23–25.

Huang, J., 2000. Molecular systematics and evolution of the genus *Ephedra*. Ph.D. Dissertation. University of Georgia.

Huang, J., Giannasi, D.E., Price, R.A., 2005. Phylogenetic relationships in *Ephedra* (Ephedraceae) inferred from chloroplast and nuclear DNA sequences. Mol. Phylogenet. Evol. 35, 48–59.

Huelsenbeck, J.P., Ronquist, F.R., 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Ickert-Bond, S.M., Wojciechowski, M.F., 2002. Assessing the utility of the *LEAFY* gene for phylogeny reconstruction in *Ephedra* (Ephedraceae, Gnetales). J. Arizona-Nevada Acad. Sci. (Proc. Suppl.) 37, 5.

Ickert-Bond, S.M., Wojciechowski, M.F., 2004. Phylogenetic relationships in *Ephedra* (Gnetales): evidence from nuclear and chloroplast DNA sequence data. Syst. Bot. 29, 834–849.

Jobes, D.V., Thien, L.B., 1997. A conserved motif in the 5.8S ribosomal RNA (rRNA) gene is a useful diagnostic marker for plant internal transcribed spacer (ITS) sequences. Plant Mol. Biol. Rep. 15, 326–334.

Karvonen, P., Savolainen, V.O., 1993. Variation and inheritance of ribosomal DNA in *Pinus sylvestris* L.: chromosomal organization and structure. Heredity 71, 614–622.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

Liston, A., Robinson, W.A., Oliphant, J.M., Alvarez-Buylla, E.R., 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed plants. Syst. Bot. 21, 109–120.

Liston, A., Robinson, W.A., Piñero, D., Alvarez-Buylla, E.R., 1999. Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. Mol. Phylogenet. Evol. 11, 95–109.

Li, J., Davis, C.C., Donoghue, M.J., Kelley, S., Tredici, P.D., 2001. Phylogenetic relationships of *Torreya* (Taxaceae) inferred from sequences of nuclear ribosomal DNA ITS region. Harvard Pap. Bot. 6, 275–281.

Liu, J.-S., Schardl, C.L., 1994. A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. Plant Mol. Biol. 26, 775–778.

Maggini, F., Frediani, M., Gelati, M.T., 2000. Nucleotide sequence of the internal transcribed spacers of ribosomal DNA in *Picea abies* Karst. DNA Seq. 11, 87–89.

Maggini, F., Marrocco, R., Gelati, M.T., De Dominicis, R.I., 1998. Lengths and nucleotide sequences of the internal spacers of nuclear ribosomal DNA in gymnosperms and pteridophytes. Plant Syst. Evol. 213, 199–205.

Maniatis, T., Fritsch, E.F., Sambrook, J., 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Publications, Cold Spring Harbor, NY.

Mellerowicz, E.J., Horgan, K., Walden, A., Coker, A., Walter, C., 1998. *PRFLL*—a *Pinus radiata* homologue of *FLORICAULA* and *LEAFY* is expressed in buds containing vegetative shoot and undifferentiated male cone primordia. Planta 206, 619–629.

Mogensen, H.L., 1996. The hows and whys of cytoplasmic inheritance in seed plants. Am. J. Bot. 83, 383–404.

Mouradov, A., Glassick, T., Hamdorf, B., Murphy, L., Fowler, B., Marla, S., Teasdale, R.D., 1998. *NEEDLY*, a *Pinus radiata* ortholog of *FLORICAULA*/*LEAFY* genes, expressed in both reproductive and vegetative meristems. Proc. Natl. Acad. Sci. USA 95, 6537–6542.

Muir, G., Fleming, C.C., Schlötterer, C., 2001. Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. Mol. Biol. Evol. 18, 112–119.

Neefs, J.-M., van de Peer, Y., de Rijk, P., Chapelle, S., de Wachter, R., 1993. Compilation of small subunit RNA structures. Nucleic Acids Res. 21, 3025–3049.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Razafimandimbison, S.G., Kellogg, E.A., Bremer, B., 2004. Recent origin and phylogenetic utility of divergent ITS putative pseudogenes: a case study from Naucleeae (Rubiaceae). Syst. Biol. 53, 177–192.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rydin, C., Mohr, B.A., Friis, E.M., 2003. *Cratonia cotyledon* gen. et sp. nov.: a unique Cretaceous seedling related to *Welwitschia*. Proc. R. Soc. Lond. Ser. B 270 (Suppl.), 29–32.

Sanderson, M.J., Doyle, J.J., 1992. Reconstruction of organismal phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. Syst. Biol. 41, 4–17.

Senchina, D.S., Álvarez, I., Cronn, R.C., Liu, B., Rong, J., Noyes, R.D., Paterson, A.H., Wing, R.A., Wilkins, T.A., Wendel, J.F., 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol. Biol. Evol. 20, 633–643.

Skalická, K., Lim, K.Y., Matyásek, R., Koukalová, B., Leitch, A.R., Kovarík, A., 2003. Rapid evolution of parental rDNA in a synthetic tobacco allotetraploid line. Am. J. Bot. 90, 988–996.

Soltis, D.E., Soltis, P.S., 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. In: Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), Molecular Systematics of Plants II: DNA Sequencing. Kluwer Academic Publishers, Boston, pp. 21–24.

Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Suh, Y., Thien, L.B., Zimmer, E.A., 1992. Nucleotide sequences of the internal transcribed spacers and 5.8S rRNA gene in *Canella winterana* (Magnoliales; Canellaceae). Nucleic Acids Res. 20, 6101–6102.

Suh, Y., Thien, L.B., Reeve, H., Zimmer, E.A., 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. Am. J. Bot. 80, 1042–1055.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Troitsky, A.V., Melekhovets, Y.F., Rakhimove, G.M., Bobrova, V.K., Valiejo-Roman, K.M., Antonov, A.S., 1991. Angiosperm origin and early stages of seed plant evolution deduced from rRNA sequence comparisons. J. Mol. Evol. 32, 253–261.

van der Sande, C.A.F.M., Kwa, M., van Nues, R.W., van Heerikhuizen, H., Raué, H.A., Planta, R.J., 1992. Functional analysis of internal transcribed spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA. J. Mol. Biol. 223, 899–910.

van Nues, R.W., Rientjes, J.M.J., Morré, S.A., Mollee, E., Planta, R.J., Venema, J., Raué, H.A., 1995. Evolutionarily conserved structural elements are critical for processing of internal transcribed spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA. J. Mol. Biol. 250, 24–36.

Volkov, R.A., Borisjuk, N.V., Panchuk, I.I., Schweizer, D., Hemleben, V., 1999. Elimination and rearrangement of parental rDNA in the allotetraploid *Nicotiana tabacum*. Mol. Biol. Evol. 16, 311–320.

Vollmer, S.V., Palumbi, S.R., 2004. Testing the utility of internally transcribed spacer sequences in coral phylogenetics. Mol. Ecol. 14, 2763–2772.

Wendel, J.F., Schnabel, A., Seelanan, T., 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc. Natl. Acad. Sci. USA 92, 280–284.

Wendt, T., 1993. A new variety of *Ephedra torreyana* (Ephedraceae) from West Texas and Chihuahua, with notes on hybridization in the *E. torreyana* complex. Phytologia 74, 141–150.

White, T.J., Birns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols: A Guide to Methods and Applications. Academic Press, New York, pp. 315–322.

Won, H., Renner, S.S., 2003. Horizontal gene transfer from flowering plants to *Gnetum*. Proc. Natl. Acad. Sci. USA 100, 10824–10829.

Won, H., Renner, S.S., submitted-a. Evolution and biogeography of *Gnetum*. Am. J. Bot.

Won, H., Renner, S.S., submitted-b. Molecular evolution of the chloroplast *trnT-trnF* region in the seed plant lineage Gnetales. J. Mol. Evol.

Xiang, Q.-P., Xiang, Q.-Y., Liston, A., Fu, L.-K., Fu, D.-Z., 2000. Length variation of the nuclear ribosomal DNA internal transcribed spacer in the genus *Abies*, with references to its systematic utility in Pinaceae. Acta Bot. Sin. 42, 946–951.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.