

The Chloroplast *trnT-trnF* Region in the Seed Plant Lineage Gnetales

Hyosig Won,^{1*} Susanne S. Renner²

¹ Department of Biology, University of Missouri—St. Louis, 8001 Natural Bridge Road, St. Louis, MO 63121, USA

² Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166, USA

Received: 3 July 2004 / Accepted: 26 May 2005 [Reviewing Editor: Debashish Bhattacharya]

Abstract. The *trnT-trnF* region is located in the large single-copy region of the chloroplast genome. It consists of the *trnL* intron, a group I intron, and the *trnT-trnL* and *trnL-trnF* intergenic spacers. We analyzed the evolution of the region in the three genera of the gymnosperm lineage Gnetales (*Gnetum*, *Welwitschia*, and *Ephedra*), with especially dense sampling in *Gnetum* for which we sequenced 41 accessions, representing most of the 25–35 species. The *trnL* intron has a conserved secondary structure and contains elements that are homologous across land plants, while the spacers are so variable in length and composition that homology cannot be found even among the three genera. Palindromic sequences that form hairpin structures were detected in the *trnL-trnF* spacer, but neither spacer contained promoter elements for the tRNA genes. The absence of promoters, presence of hairpin structures in the *trnL-trnF* spacer, and high sequence variation in both spacers together suggest that *trnT* and *trnF* are independently transcribed. Our model for the expression and processing of the genes tRNA^{Thr}(UGU), tRNA^{Leu}(UAA), and tRNA^{Phe}(GAA) therefore attributes the seemingly neutral evolution of the two spacers to their escape from functional constraints.

Key words: Gymnosperms — Gnetales — *trnL* intron — *trnT-trnL* intergenic spacer — *trnL-trnF* intergenic spacer — Palindromic sequences

Introduction

The *trnT-trnF* region is located in the large single-copy region of the chloroplast (cp) genome. Together with its flanking genes (Fig. 1A), it forms a cistron that is conserved from mosses to seed plants (Taberlet et al. 1991). Even where chloroplast genomes have undergone extreme structural changes, such as in *Pinus thunbergii* (Wakasugi et al. 1994) (Fig. 1B), which has pseudogenized *ndhF* genes located immediately downstream from the *trnT-trnF* region, the region itself is conserved. Recent large-scale analyses have shown that the cistron comprised of the three tRNA genes is a uniquely shared characteristic of land plants (Quandt et al., 2004); in green algae, the *trnF* gene is located farther away from the other two genes. The *trnT* and *trnL* genes are lost, and the *trnL-trnF* region is disrupted, in a few nonphotosynthetic flowering plants, such as *Epifagus virginiana* (Orobanchaceae [Wolfe et al. 1992]) (Fig. 1C).

Because of its conserved gene order, its high variability, and the adequate size of its intergenic spacer (IGS) and intron, the *trnT-trnF* region is widely sequenced for phylogenetic analyses at the species and genus level in seed plants (Taberlet et al. 1991; Bakker et al. 2000; Borsch et al. 2003; Quandt et al., 2004; Shaw et al. 2005). Throughout green plants, tRNA genes are either monomeric, with putative promoter elements located upstream of the genes, or polymeric, containing other tRNAs, rRNAs, and mRNAs (Martin 1995). Examples of monomeric tRNAs are tRNA^{Ser} (Gruissem et al. 1986; Wu et al. 1997) and tRNA^{Met} (Gruissem and Zurawski 1985). Examples of polymeric tRNAs are tRNA^{Arg}, which is part of an operon encompassing 5'-16S-tRNA^{Ile}-tRNA^{Ala}-23S-4.5S-5S-tRNA^{Arg}-3' and is located within the inverted

*Current address: Seoul National University, College of Natural Sciences, School of Biological Sciences, Seoul, Korea 151-742
Correspondence to: Susanne S. Renner, Ludwig Maximilians Universität, Systematische Botanik, Menzinger Str. 67, D-80638 München, Germany; email: renner@lrz.uni-muenchen.de

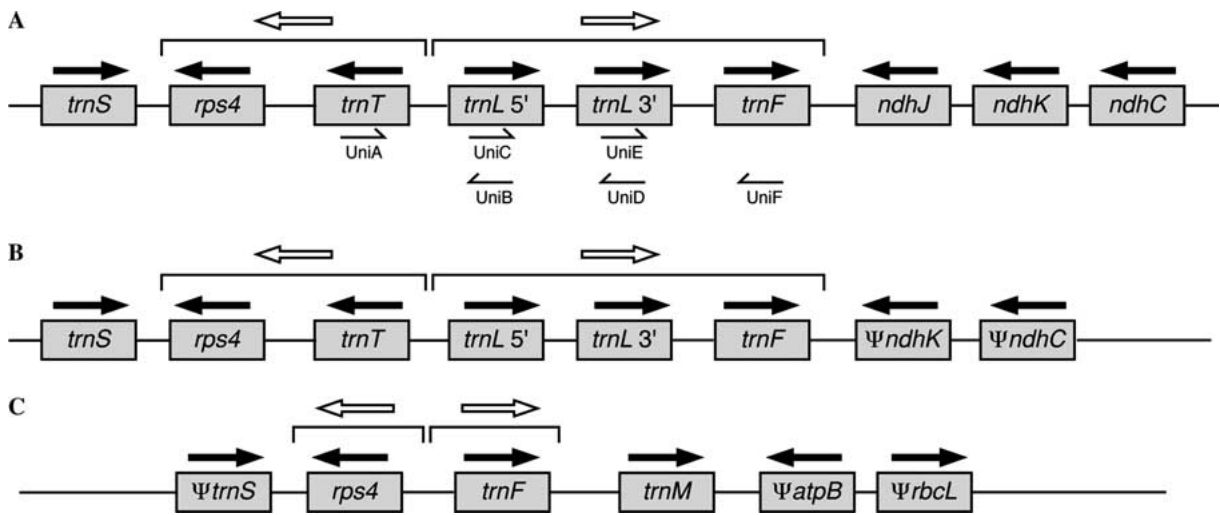


Fig. 1. Organization of the *trnT-trnL-trnF* cistron in the large single-copy region of the chloroplast genome. Open arrows indicate postulated directions of transcription; filled arrows, orientation of genes. Primers used to amplify the *trnT-trnF* region are shown as half-arrows. **A** Map representing gene order in mosses, ferns, and angiosperms. Genes flanking the *trnT-trnF* region are conserved across *Marchantia* (liverwort), *Anthoceros* (hornwort), *Psilotum* (whisk fern), and the angiosperm genera *Amborella*, *Calycanthus*,

Oryza sativa, *Triticum aestivum*, *Zea mays*, *Lotus japonica*, *Spinacia oleracea*, *Nicotiana tabacum*, *Atropa belladonna*, and *Arabidopsis thaliana*. **B** Map of *Pinus thunbergii* (black pine). Note the pseudogenized *ndhK* and *ndhC* and the lack of *ndhJ* compared to mosses, ferns, and angiosperms. **C** Map of *Epifagus virginiana*, a nonphotosynthetic angiosperm, modified from Wolfe et al. (1992). Note the loss of *trnT*, *trnL*, and *ndh* genes and the pseudogenization of *trnS*, *atpB*, and *rbcL*.

repeat region (Leal-Klevezas et al. 2000), and the *trnE* operon of tRNA^{Glu}, tRNA^{Tyr}, and tRNA^{Asp}, which is located in the large single-copy region. The transcripts of such operons are processed to release the individual tRNA molecules (Ohme et al. 1985).

Chloroplast genes, such as the tRNA genes, usually contain upstream and downstream signals facilitating their expression. Termination of transcription is signaled by hairpin structures and T-rich regions downstream of the operon (Bogorad 1991). Steinmetz et al. (1983) compared the 5'-flanking regions of the tRNA^{Ser}, tRNA^{Phe}, tRNA^{Thr}, tRNA^{Leu}, and tRNA^{Met} genes and discovered nucleotide stretches that resembled the “-35” and “-10” promoter regions of bacterial genes, probably because of the bacterial origin of the chloroplast (see also Hanley-Bowdoin and Chua 1987; Bogorad 1991; Inokuchi and Yamao 1995). Analysis of these conserved regions together with ones seen in other plastid genes yields the consensus sequences ATTGANA at “-35” and TAAGAT at “-10.” A study of the function of these prokaryote-type promoter regions in *psbA*, *rbcL*, and *atpB* showed that they are essential for the proper expression of these genes (Gruissem and Zurawski 1985). Other tRNA genes, however, lack prokaryote-type promoter elements and instead have internal promoter elements that are highly conserved (Galli et al. 1981; Gruissem et al. 1986; Cheng et al. 1997). Putative promoter elements upstream of *trnT* as well as *trnF* (Steinmetz et al. 1983; Kanno and Hirai 1993; van Ham et al. 1994) suggest that these genes can be expressed independently, although no experimental study has been done. Independent

expression is also indicated by the opposite orientation of the tRNA^{Thr}(UGU) gene relative to the tRNA^{Leu}(UAA) and tRNA^{Phe}(GAA) genes (Figs. 1A and B). This suggests that the tRNA genes may comprise two operons, one consisting of the *trnT* and *rps4* genes and the other of *trnL* and *trnF*. Chloroplast tRNA genes or operons are transcribed with 5' leader and 3' trailer sequences, and the precursor-tRNAs are posttranscriptionally modified to form mature tRNAs. The processing events consist of 5' end maturation, formation of a 3' end, CCA addition, intron removal, and base modification (Martin 1995). The endonucleolytic cleavage of the 5' leader sequence and the 3' trailer sequence has been attributed, respectively, to endonucleases RNase P (Gegenheimer 1996; Frank and Pace 1998) and RNase Z (Frank and Pace 1998; Schiffer et al. 2001, 2002). The spacers are not needed for the processing of tRNAs, leading Bakker et al. (2000) to the conclusion that they may evolve neutrally. By contrast, the group I intron contained in the *trnL* gene maintains its secondary structure and sequence elements critical for its self-splicing from precursor-RNAs (Cech 1988; Michel and Westhof 1990; Simon et al. 2003). These sequence elements are conserved from eubacteria to chloroplasts (Kuhnel et al. 1990; Xu et al. 1990; Besendahl et al. 2000).

Here we analyze the cp *trnT-trnF* region in the three genera of Gnetales, an ancient lineage of seed plants for which well-supported species-level phylogenies are available (Ickert-Bond and Wojciechowski 2004; Huang et al. 2005; Won and Renner 2005a). The sequence characteristics discovered, together with the typical

Table 1. Length variation (bp) and G+C content (percentage) of the chloroplast *trnT-trnL* IGS, *trnL* intron, and *trnL-trnF* IGS in Gnetales

	<i>trnT-trnL</i> IGS	<i>trnL</i> intron	<i>trnL-trnF</i> IGS
<i>Gnetum</i> (<i>n</i> = 41)	292–340 (30.4–34.2)	329–384 (34.8–39.3)	132–169 (33.7–45.7)
S. America	300–304 (31.9–32.6)	329–379 (34.8–37.5)	132–169 (34.3–43.9)
Africa	292 (31.2–31.5)	384 (34.9)	165 (37.0)
Asia I	296–311 (31.1–34.2)	341–351 (36.6–38.9)	138–166 (33.7–39.4)
Asia II	296–340 (30.4–32.7)	332–347 (36.4–39.3)	133–149 (40.0–45.7)
<i>Welwitschia</i> (<i>n</i> = 1)	313 (30.4)	317 (30.4)	250 (33.2)
<i>Ephedra</i> (<i>n</i> = 3)	156 (28.8)	290 (36.6–37.2)	112* (26.7)*

*Based on one sequence provided by S. Ickert-Bond (pers. commun.).

function of enzymatic systems involved in cp tRNA processing, lead us to suggest a model for the expression and processing of the genes tRNA^{Thr}(UGU), tRNA^{Leu}(UAA), and tRNA^{Phe}(GAA).

Materials and Methods

Taxon Sampling

The samples, voucher information, and GenBank accession numbers obtained in this study are listed in Appendix 1 (Supplementary Material). Most samples were collected in the field by H.W., but a few are from botanical gardens or herbarium material.

Gene Sequencing

We extracted DNA from silica gel-dried leaves, using QIAGEN Plant DNeasy minikits. Concentration and quality of extracted DNAs were checked by 1% agarose gel electrophoresis with a (*λ*HindIII/*Eco*RI size marker. The chloroplast tRNA^{Leu}(UAA) intron and adjacent IGSs were amplified using primers designed by Taberlet et al. (1991) (Fig. 1 shows their placement). The PCR reaction volume was 25 μ l and contained 2.5 μ l of *Taq* 10 \times reaction buffer, 1.5 mM MgCl₂, 0.4 mM primer, a 0.2 mM concentration of each dNTP, 0.625 unit of *Taq* DNA polymerase (Promega), and 0.5–2 μ g of template DNA. 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 PCR products in 1% agarose gels, PCR products were purified with QIAquick PCR product extraction kits (QIAGEN) and directly sequenced using the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer). Sequencing reactions were electrophoresed in an ABI 377 sequencer (Perkin Elmer). We were able to sequence/read both strands except for a stretch in the middle of the *trnL-F* spacer, where signals after the repeat of 'AT' started to overlap and fade out. This region was thus mainly read in only one direction. The same problem occurred after the 'A' and 'G' repeats in the middle of the *Gnetum trnL* intron. Sequences were aligned and cleaned up using Seqman II (DNASTAR).

Sequence Analysis

We obtained 41 complete sequences from about 25 species of *Gnetum*, 1 sequence from *Welwitschia*, and 3 from *Ephedra* (Appendix 1, Table 1). For *Ephedra*, we only sequenced the *trnT-trnL* IGS and the *trnL* intron. A data matrix for sequences of each chloroplast region was constructed using Se-AI software (version 2.0a11; <http://evolve.zoo.ox.ac.uk/>). Length and G+C content were calculated from the aligned sequences. Sequence divergences

were calculated using Kimura's (1980) two-parameter model. Chloroplast tRNA^{Leu}(UAA) intron and adjacent IGS sequences were compared to the sequences of tRNA^{Leu}(UGU), tRNA^{Thr}(UAA), and tRNA^{Phe}(GAA) of *Pinus thunbergii*, the chloroplast of which has been sequenced (Wakasugi et al. 1994), to determine the boundaries of the genes, intron, and spacer. To predict the secondary structure of the region, especially of palindromic (inverted repeated) sequences found in the *trnL-trnF* IGS, we used MFOLD (version 3.1; <http://www.bioinfo.rpi.edu/applications/mfold/> [Zucker 2003]), which estimates the secondary structure of DNA by thermodynamics. To detect promoter elements of the tRNA genes, we examined the IGS sequences for "-35" and "-10" promoter element homologies. The data matrices of Borsch et al. (2003) and Quandt et al. (2004), which comprise *trnL* intron and spacer sequences of numerous land plants, were also compared to that of the Gnetales sequenced here.

Phylogenetic Analysis

The flanking spacers of the tRNA^{Leu} gene showed sufficient variation in length among the three genera to make alignment difficult (see Results). We therefore designated the South American *Gnetum* group as a functional outgroup for the intra-*Gnetum* analysis, based on results of a phylogenetic analysis of Gnetales *rbcL* and *matK* sequences (Won and Renner 2003). We used PAUP* 4.0b10 (Swofford 2002) and Mr. Bayes (version 3.0b4 [Huelsenbeck and Ronquist 2001]) for phylogenetic analyses. The data matrix of the *trnL* intron and *trnT-trnL* and *trnL-trnF* spacers was analyzed by parsimony, neighbor-joining, and Bayesian inference. Parsimony analyses were run using heuristic searching, with 100 random taxon-addition replicates, holding 100 trees at each step, tree-bisection-reconnection (TBR) branch swapping, the MulTrees, Collapse, and Steepest Descent options, and no upper limit for trees held in memory. Neighbor joining used K-2-P distances and ignored gaps. Bayesian probabilities were obtained under the GTR + γ + Pinv model, with four Markov chain Monte Carlo chains run for 3 million generations, using random trees as starting point, and sampling every 10,000th generation. The trees sampled before the saturation of maximum likelihood estimates were discarded as burn-in. Nonparametric bootstrap support for parsimony and neighbor joining was obtained by resampling the data 1000 times with the same search options and model.

Results

Length and G+C Contents

The length of the *trnL* intron was 319–384 bp; that of the *trnT-trnL* spacer, 292–340 bp; and that of the *trnL-trnF* spacer, 132–169 bp (Table 1). These sizes are similar to *Welwitschia* (317, 313, and 250 bp) but

A

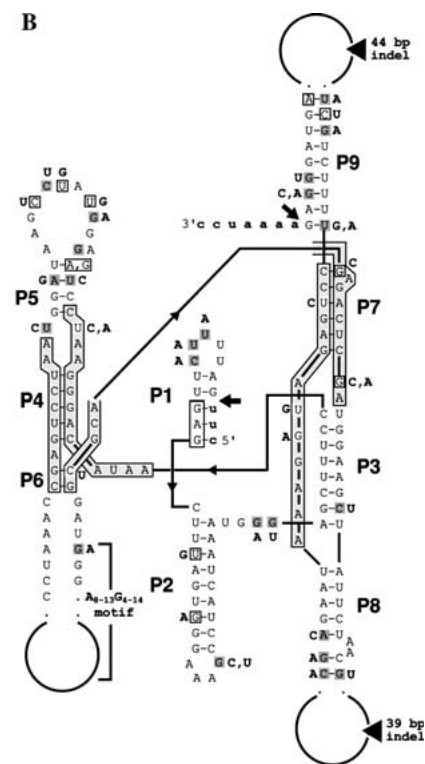
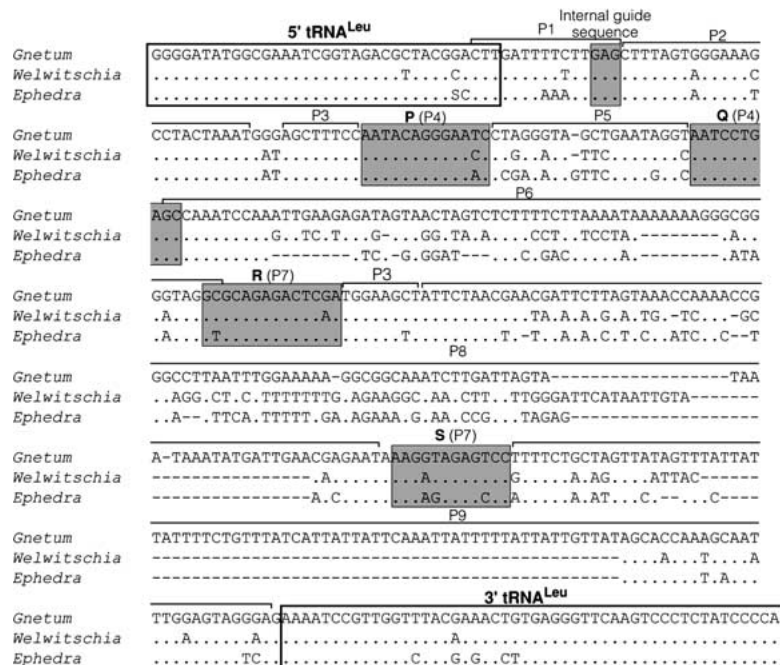


Fig. 2. A Sequence alignment of the chloroplast tRNA^{Leu}(UAA) gene and intron (a group I intron). Open boxes indicate the tRNA^{Leu} 5' and 3' coding regions and shaded boxes indicate conserved sequence elements (P, Q, R, S) in the intron. Nine stem-loop structures (P1–P9) are marked by brackets. See Cech (1988) for the secondary structure nomenclature of group I introns. Sequences shown are those of *Gnetum africanum*, *Welwitschia mirabilis*, and *Ephedra trifurca*. B Secondary structure model of the

chloroplast *trnL* intron in Gnetales. Intron splicing sites are marked with two arrows. Locations of the A₈₋₁₃G₄₋₁₄ motif in the P6 element, the 39-bp indel in the P8 element, and the 44-bp indel in the P9 element of *Gnetum* are indicated. Boxed bases indicate variation within *Gnetum*; shaded bases, variation among the three genera of Gnetales. Variable bases are in boldface. The secondary structure model is modified from Cech et al. (1994) and Quandt et al. (2004)

significantly longer than *Ephedra* (290, 156, and 112 bp), especially in the *trnT-trnL* spacer. Of the multiple accessions of *Gnetum*, only the four of *G. gneumon* showed distinct variation in length (one was excluded from the data matrix). Two of them, Suksathan s.n. and RBGE19480112, had shorter spacer sequences due to autapomorphic gaps at base positions 355–365 of the *trnT-trnL* spacer and positions 1104–1113 of the *trnL-trnF* spacer compared to the other two accessions, Won 514 and Won 568.

G+C contents of the *trnL* intron and spacers showed opposite trends among the three genera (Table 1). In the *trnL* intron, *Gnetum* had a higher G+C content than *Welwitschia* (34.8–39.3% vs. 30.4%) but was similar to *Ephedra* (36.6–37.2%). In the *trnT-trnL* and *trnL-trnF* spacers, *Gnetum* was comparable to *Welwitschia* (30.4–34.2% vs. 30.4% and 33.7–45.7% vs. 33.2%, respectively) but had a higher G+C content than *Ephedra* (30.4–34.2% vs. 28.8% and 33.7–45.7% vs. 26.7%, respectively).

Sequence Comparison, Alignment, and Divergence

Due to the large length and sequence variation, we could not align the spacer sequences of three genera,

nor could we find homologous motifs in their spacer regions. Like Borsch et al. (2003) and Quandt et al. (2004), we failed to detect specific homologies of Gnetales with other seed plant sequences. The *trnL* intron sequences, by contrast, contained several highly conserved motifs (Fig. 2), especially the P, Q, R, and S sequence elements and the P1, P2, and P3 elements (Figs. 2 and 3A). The P5, P6, P8, and P9 structural elements were variable. Comparison of gnetalean sequences with those of other land plants also revealed the expected conserved tRNA^{Leu} group I intron structure. By comparing our data matrix with the basal angiosperm matrix of Borsch et al. (2003), we detected a highly conserved “–35”-like element (TTGACA) and “–10”-like element (TAGGAT) between base positions –45 and –16 in the upstream of the tRNA^{Phe} gene. However, we could not detect any promoter elements in the *trnT-trnL* or *trnL-trnF* spacer regions of Gnetales.

Alignment of the *trnL* intron and IGSs of *Gnetum* required 1-, 4-, 5-, 6-, 8-, 9-, 10-, 15-, 39-, 40-, and 44-bp-long noninformative indels and 11 informative indels (Appendix 2). The P8 and P9 structural elements of *G. paniculatum* and *G. africanum* contained 39- and 44-bp-long indels (Fig. 2B), and the *trnT*–

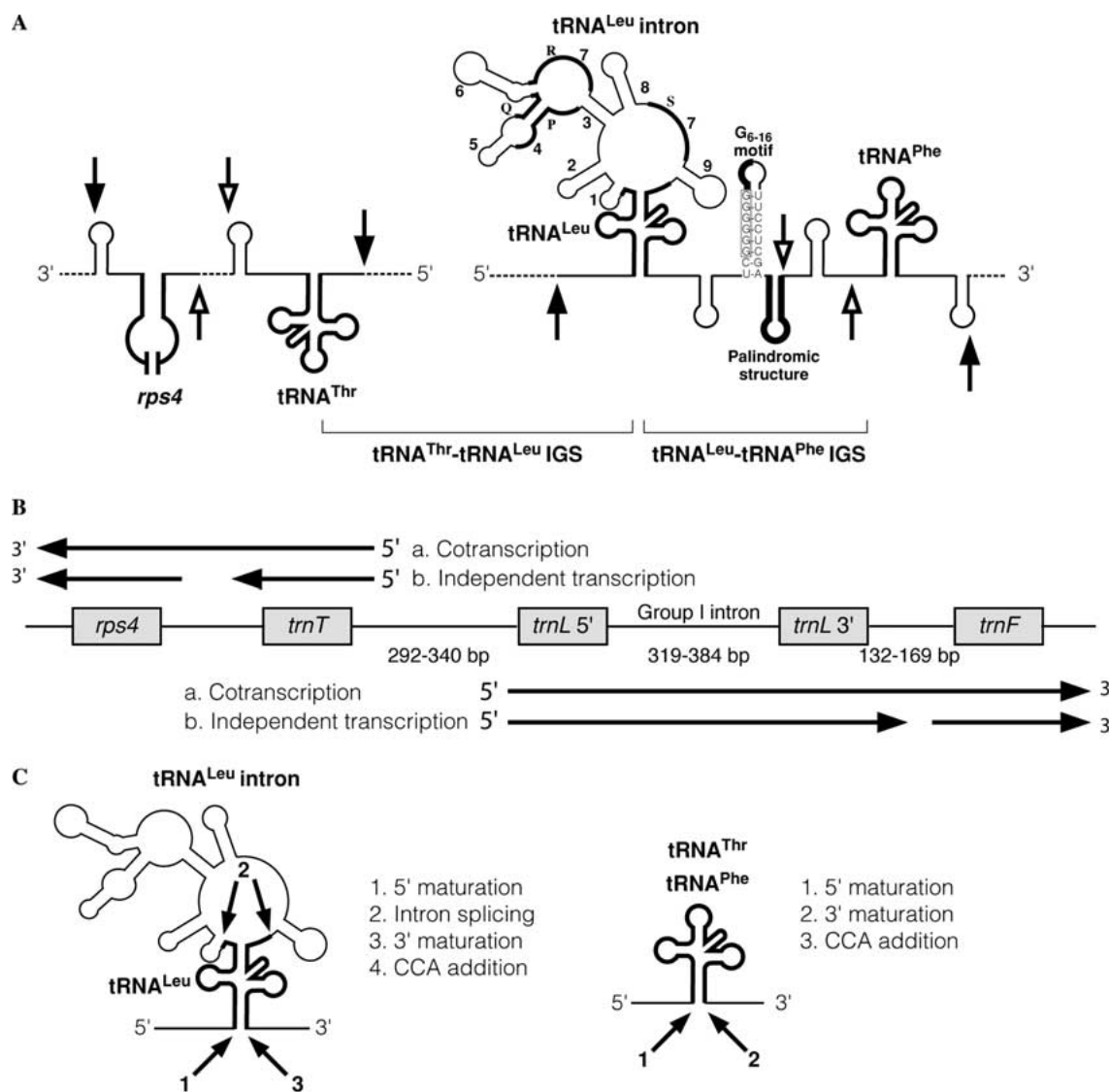


Fig. 3. Secondary structure model for *Gnetum* chloroplast tRNAs and inferred mode of tRNA transcription and processing. **A** Proposed secondary structure for *Gnetum* chloroplast tRNA^{Thr}(UGU), tRNA^{Leu}(UAA), and tRNA^{Phe}(GAA). Solid lines indicate transcribed regions; dashed lines indicate nontranscribed regions upstream or downstream of the tRNA transcription start/termination points. Filled arrows indicate the transcription start/termination point; open arrows indicate additional putative transcription

termination/start points for the independent transcription model. The secondary structure model for *trnL* intron is modified from that of Kuhse et al. (1990). The structure of the *rps4* gene is arbitrary. **B** Proposed transcription of *Gnetum* *trnT*, *trnL*, and *trnF* genes. *TrnT* and *rps4* are transcribed in the opposite direction from *trnL* and *trnF*. Both *trnT* and *rps4*, and *trnL* and *trnF*, could be cotranscribed or independently transcribed. **C** Proposed order of tRNA processing.

trnL IGS region of *G. aff. latifolium* Beaman 9408 and *G. klossi* contained 40- and 15-bp-long indels. The P6 structural element of the intron also contained an A₈₋₁₃G₄₋₁₄ motif with varying repeat numbers of 'A' and 'G' nucleotides (Fig. 2B), and the *trnL*–*trnF* spacer a G₆₋₁₆ motif at base positions 977–994 and palindromic sequences with AT repeats at base positions 1008–1061 (Fig. 3A; see Palindromic Structures in the *trnL*–*trnF* IGS Region, below). There was intraspecific variation in these mono- and dinucleotide repeats.

Sequence divergences among the *gnetalean* genera and among *Gnetum* species are specified in Appendix 3.

Palindromic Structures in the *trnL*–*trnF* IGS Region

The palindromic sequences in the *trnL*–*trnF* spacer were first identified by sequence comparison and then verified by looking at the secondary structure predicted by MFOLD. They are listed in Table 2. Palindromes were found in all *Gnetum* sequences. They formed hairpin structures, with tandem repeats of 'AT' making up the stem of the hairpin (Fig. 3A). The secondary structure of the *trnL*–*trnF* spacer suggests that the G₆₋₁₆ motif is also involved in a conserved hairpin immediately upstream of the palindromic sequences, with

Table 2. Palindromic sequences in the *Gnetum* chloroplast tRNA^{Leu}-tRNA^{Phe} IGS region

A. AT repeat types

5'-AAGTT-(AT)_n-AACTT-3'

n = 4; *G. parvifolium*, *G. sp.* Harder 5621

n = 5; *G. costatum*, *G. cuspidatum* Won 562, *G. diminutum*, *G. hainanense s.l.*, *G. klossii*, *G. latifolium* Won 524, Won 575, *G. macrostachyum*, *G. tenuifolium*, *Gnetum* sp. nov. (*Vinkiella* nom. nud.)

n = 6; *G. acutum*, *G. aff. latifolium* Won 545, *G. aff. latifolium* SAN151116, *G. neglectum*, *G. microcarpum*

n = 10; *G. gnemon* RBGE19480122, Suksathan *s.n.*

5'-AAGTT-(AT)₂ATAA(AT)₂-AACTT-3'

G. gnemon Won 514, *G. ula*

5'-AAGTT-(AT)₂AT(AT)₂AC(AT)₂-AACTT-3'

G. aff. latifolium Beaman 9408

5'-AAGTT-AGTTG-(AT)₂AT(AT)₂AG(AT)₂GT(AT)₂-CAACT-AACTT-3'

G. raya

B. AT and AAAT repeat type

5'-AAGTT-(AAAT)₃-AT-(AT)₄-TT-(ATTT)₃-AACTT-3'

G. gnemonoides SAN151121, SAN151130

5'-AAGTT-(AAAT)₃-AA-(AT)₃-AT-(ATTT)₃-AACTT-3'

G. gnemonoides Utami *s.n.*

C. Doubly palindromic—*G. africanum*

5'-AAGT-T-(AT)₅-AT-ACCT-TTTTGG

GAAAGT-TATGATATCTATATATCATA-ACCTTTC-3'

D. South American type

5'-TTAT-(AT)₃-ATAA-3'

G. microstachyum

5'-TATTT-CT-(AT)₈-AG-AAATA-3'

G. nodiflorum

5'-TTATGT-(AT)₃-CT-(AT)₃-AG-(AT)₃-ACATAA-3'

G. paniculatum

5'-TTAT-(AT)₄-TATAT-(AT)₄-ATAA-3'

G. urens

5'-(AT)₄-TTCT-(AT)₁₀-AGAA-(AT)₄-3'

G. schwackeanum

5'-TATAT-(AT)₃-AGAAAT-(AT)₃-AAATA-3'

G. woodsonianum

UCGGGGGG//AGCUCCUU forming a stem (Fig. 3A). Prediction of the secondary structure of the *Welwitschia trnL-trnF* spacer detected several hairpins, although the sequence itself does not have palindromic sequences. Palindromic structures found in *Gnetum* coincided with geographical clades; Asian and African palindromic sequences started with AAGTT and ended in AACTT, which form a stem by complementary base-pairing. The AAGTT motif was conserved throughout *Gnetum*, while the AACTT motif was absent from the South American species (Table 2; Fig. 4, gap I). However, although the AAGTT motif was present in South America, it was not involved in hairpin formation there. Asian species had AT repeats except for *G. gnemonoides*, which had AT plus AAAT repeats, perhaps as a result of mutation from ATAT, with compensatory downstream changes of ATAT to ATTT. *Gnetum africanum* had two consecutive palindromic structures, referred to as “doubly palindromic” in Table 2, separated by five bases (TTTTG). The length of the palindromes 5'-AAGTT-(AT)_n-AACTT-3' ranged from 18 bp (*n* = 4) to 30 bp (*n* = 10), and there were several cases with substitutions in the AT repeats (Table 2). The C's and G's in the middle of AT repeats of the South American and African sequences

probably result from compensatory base changes in the stem region of hairpin structures.

Phylogenetic Analyses

Alignment of chloroplast *trnL* intron and adjacent spacer sequences required 1 to 44 bp indels as described above; number of aligned characters, parsimony-informative sites, and gaps are shown in Table 3, and the data file has been submitted to TreeBase. Sequences of *Gnetum* could not be aligned with those of *Welwitschia* and *Ephedra* except in the *trnL* intron region (Fig. 2). One of the tree topologies estimated by Bayesian analysis (in four repeat runs) of the complete intron and spacer sequences is shown in Fig. 4. Parsimony analysis yielded 71 trees on three islands. Accessions sorted geographically into a South American clade, an African clade, and two Asian clades (labeled Asia I and Asia II in Fig. 4) that were well supported except for the monophyly of the Asian group. Within the Asia II clade, two subclades were resolved, a *G. cuspidatum* clade and a *G. hainanense s.l.* clade (Fig. 4). Except for two homoplasious gaps marked A and G in Fig. 4, gaps (B–K) were informative. A neighbor-joining tree from

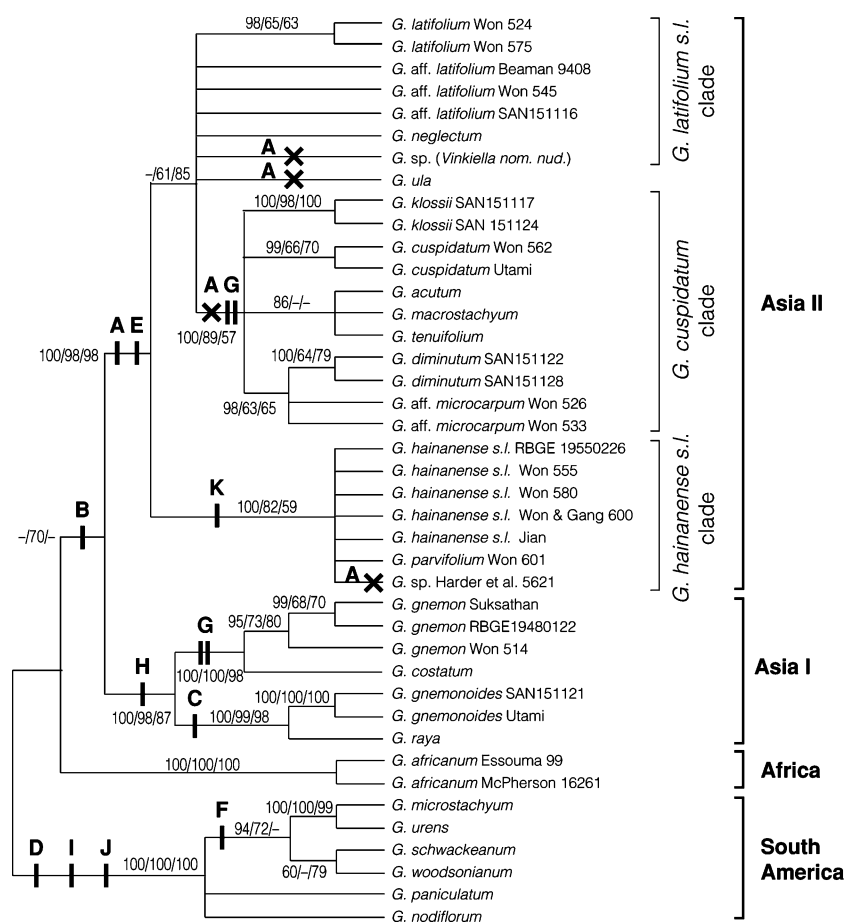


Fig. 4. Phylogeny of *Gnetum* obtained from Bayesian analysis of the chloroplast *trnL* intron and spacer sequences. Values next to nodes are Bayesian posterior probabilities, followed by bootstrap values under parsimony (CI = 0.55, RI = 0.45, excluding uninformative characters) and neighbor joining. Base positions 142–181, 246–260, 297–301, 576–606, 706–745, 810–853, 976–993, and 1012–1070 were excluded from the analyses and 11 informative gaps (A–K) were binary-coded (see Appendix 2).

Table 3. Number of characters, parsimony-informative sites, and phylogenetically informative indels in the chloroplast *tRNA^{Leu}* intron and intergenic spacers in *Gnetum*

	<i>trnT</i> – <i>trnL</i> IGS	<i>trnL</i> intron	<i>trnL</i> – <i>trnF</i> IGS
Aligned characters	371	436	189
Parsimony-informative sites	46 (12.4%)	33 (7.6%)	46 (24.3%)
Informative sites after excluding gaps	45 (12.1%)	26 (6.0%)	18 (9.5%)
Phylogenetically informative indels	5	3	3

just the *trnL* intron sequences also identified the South American species as the basal branch and sister to the Asian and African clades (data not shown).

Discussion

Behavior of the trnL Intron and Adjacent Spacers in Gnetum

Topologies obtained from the *trnL* sequences were congruent with phylogenies from separate and combined analyses of nuclear sequences (rRNA Internal Transcribed Spacer [ITS] and the second intron of *LEAFY*), chloroplast *matK* and *rbcL* genes, and the conserved region of the mitochondrial *nadI* second intron and partial exon sequences (Won and Renner 2005a, b). This is in agreement with other studies that

show that fast evolving regions can sometimes provide cost-efficient phylogeny reconstruction even of deep divergences (Hershkowitz and Lewis 1996; Källersjö et al. 1998; Borsch et al. 2003; Hilu et al. 2003; Quandt et al. 2004). Based on fossil-calibrated molecular clocks, the split between *Gnetum* and *Welwitschia* had already occurred by 115 million years ago (Won and Renner 2003). This great age partly explains the high sequence variation in the two spacers; in addition, the absence of conserved promoter elements for tRNA genes from both spacers suggests that neither is under functional constraints as suggested earlier for the *trnL*–*trnF* spacer but not the *trnT*–*trnL* spacer (Bakker et al. 2000; Quandt et al. 2004).

We found several palindromes in the *Gnetum trnL*–*trnF* IGS region (Table 2). These sequences were involved in hairpin formation by intrastrand

base-pairing. In addition, two more hairpin structures were commonly formed between the 3' end of the tRNA^{Leu} gene and the palindromic structure (Fig. 3A). Such stem-and-loop structures can be substrates for structure-specific nucleases and/or mismatch repair enzymes (Leach 1994) and can cause double-strand breakage during DNA replication (Nasar et al. 2000). Slippage during replication is also the likely cause of the many 'A,' 'G,' and 'AT' repeats in the intron and spacer regions (Levinson and Gutman 1987; Kelchner 2000; Graham et al. 2000). So far, no enzymes that would recognize the hairpin structures in the tRNA spacers have been reported. Instead, the enzymes that process precursor tRNAs, such as RNase P, RNase Z, and ATP (CTP):tRNA nucleotidyl transferase, recognize only the tRNA's tertiary structure (Martin 1995; Frank and Pace 1998; Kunzmann et al. 1998; Schiffer et al. 2002), suggesting that the nucleotide sequence between the tRNA genes may be almost irrelevant to tRNA enzymatic maturation. Circumstantial evidence, however, indicates that the palindromes in the *trnL-trnF* spacer serve two functions: By forming thermodynamically stable hairpins, they may reduce the chance that stem-loops interfere with tertiary structure formation of adjacent tRNAs. By forming hairpin structures, they may also function as transcription termination signals as has been found in operons originating from bacteria, including chloroplast operons (Bogorad 1991). Hairpin structures/inverted repeated sequences have been reported from the downstream of maize tRNA^{His}(GUG) (Schwarz et al. 1981) and tRNA^{Phe}(UGU) (Steinmetz et al. 1983), spinach tRNA^{Ser}(UGA) (Holschuh et al. 1984), tobacco tRNA^{Glu}(UUC)-tRNA^{Tyr}(GUA)-tRNA^{Asp}(GUC) (Ohme et al. 1985), and the *Brassica napus* rRNAs-tRNA^{Arg}(ACG) operon (Leal-Klevezas et al. 2000).

In many seed plants, the *trnL-trnF* spacer appears to harbor a bacterial-type promoter for the tRNA^{Phe} gene (Steinmetz et al. 1983). Indeed, we detected the "-35" and "-10" promoters in several conifers and basal angiosperms, and Quandt et al. (2004) found them in most land plant lineages except lycophytes and leptosporangiate ferns. In Gnetales, however, we detected no promoters in either of the two spacers. Studies on the function of tRNA upstream sequences have shown that either they contain prokaryote-type promoters (Gruissem et al. 1983; Steinmetz et al. 1983; Wu et al. 1997) or there are eukaryote-type internal promoters directly in the tRNA (Galli et al. 1981; Gruissem et al. 1986; Cheng et al. 1997). The loss of promoter elements from the nonphotosynthetic *Epifagus* tRNA^{Glu}(UUC) (Morden et al. 1991) suggests, however, that prokaryote-type promoters are not always required for transcription of tRNA genes, and a few studies of tRNA transcription in lineages with functioning chloroplasts also suggest

this (Gruissem et al. 1986; Jahn 1992; Wu et al. 1997). We therefore interpret the lack of promoter elements in the upstream regions of Gnetales tRNA^{Thr}, tRNA^{Leu}, and tRNA^{Phe} genes as indicating that these genes have internal promoters, relieving their spacer regions from functional constraints that would come from harboring functional promoters. Release from involvement in the tRNA promoting/processing mechanism may explain why the *trnT-trnL* and *trnL-trnF* intergenic spacers contain so many point mutations and indels.

In contrast to the two spacers, the *trnL* intron maintains highly conserved structural motifs that allow it to form the characteristic secondary structure of group I introns of land plants (Kuhse et al. 1990; Xu et al. 1990; Lambowitz et al. 1999; Belfort et al. 2002; Quandt et al., 2004) (Fig. 2). Unlike the self-splicing introns of *Anabaena* and *Tetrahymena* (Zaug et al. 1993), chloroplast *trnL* introns of land plants require helper proteins for splicing (Xu et al. 1990; Lambowitz et al. 1999; Simon et al. 2003), in this regard resembling group II introns in tRNA genes (Vogel et al. 1999). The three gaps that we detected in the *trnL* intron all belonged to the P8 stem-and-loop, one of the autapomorphic indels also belonged to the P8 element, and another occurred in the P9 element. Borsch et al. (2003) similarly found that 68% of the characters in their basal angiosperm *trnL* intron matrix were contributed by the P6 and P8 stem-loops, while the P, Q, R, and S structural elements were highly conserved.

Expression and Processing of the cp tRNA^{Thr}, tRNA^{Leu}, and tRNA^{Phe} Genes: A Model

Based on our results, the secondary structure of the cp *trnT-trnF* region is as shown in Fig. 3A. As discussed above and shown in Figs. 1A and B and Fig. 3, the cp tRNA^{Thr} gene and the tRNA^{Leu}-tRNA^{Phe} genes are transcribed in opposite directions. Two earlier studies have presented indirect evidence for cotranscription of the tRNA^{Leu} and tRNA^{Phe} genes: Bonnard et al. (1984) saw the absence of a termination signal in the *Vicia faba trnL-trnF* spacer as such evidence, and Koch et al. (2005) the presence of duplicated tRNA^{Phe} genes between the two bacterial promoter elements present in the Brassicaceae they studied. However, the structural aspects of the *trnL-trnF* spacer in *Gnetum*, especially the presence of palindromic sequences and hairpin structures, favor an interpretation of independent transcription (previous section). Regardless of whether tRNA cistrons or operons including tRNA(s) are cotranscribed or independently transcribed, the processing machinery of the chloroplast has to successfully mature each tRNA (Steinmetz et al. 1983; Ohme et al. 1985; Marion-Poll et al. 1988; Christopher and Hallick 1990; Tonkyn and Gruissem 1993; Leal-Klevezas et al. 2000). The tRNA process-

ing pathway includes (1) 5' end maturation by excision of 5' leader sequence by RNase P, (2) splicing of the *trnL* intron (group I intron), (3) excision of the 3' trailer sequence by RNase Z, (4) terminal 'CAA' addition by ATP (CTP):tRNA nucleotidyl transferase, and (5) RNA editing. The order of steps 1, 2, and 3 in the processing of the tRNA^{Leu} gene has not been studied, but the splicing of group II introns in the tRNA^{Ile}(GAU), tRNA^{Val}(UAC), tRNA^{Ala}(UGC), and tRNA^{Gly}(UCC) genes indicates that 5' endonucleolytic cleavage precedes intron splicing and 3' end maturation (Vogel and Hesse 2001; but see Delp et al. 1991). Because group I and II intron splicing mechanisms in plant chloroplasts have common features, such as mediation by intron RNA and a requirement for helper proteins (Belfort et al. 2002) or chloroplast ribonucleoproteins (cpRNPs [Nakamura et al. 1999]), it is likely that processing of the tRNA^{Leu} gene also follows the above pathway. In tRNAs without introns, 5' maturation also usually precedes 3' maturation (Wang et al. 1988; Martin 1995; Gegenheimer 1996). The processing and splicing pathway proposed here for the cp *trnL* intron now needs to be tested via experiments and comparison across a much broader taxonomic range.

Acknowledgments. This paper represents a portion of a dissertation submitted by the first author in partial fulfillment of the requirements for a Ph.D. from University of Missouri—St. Louis. Financial support from the American Society of Plant Taxonomy, the International Association of Plant Taxonomy, and the International Center for Tropical Ecology at the University of Missouri—St. Louis is gratefully acknowledged.

References

- Bakker FT, Culham A, Gomez-Martinez R, Carvalho J, Compton J, Dawtrey R, Gibby M (2000) Patterns of nucleotide substitution in angiosperm cpDNA *trnL* (UAA)-*trnF* (GAA) regions. *Mol Biol Evol* 17:1146–1155
- Belfort M, Derbyshire V, Parker MM, Cousineau B, Lambowitz AM (2002) Mobile introns: pathways and proteins. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) *Mobile DNA II*. ASM Press, Washington, DC, pp 761–783
- Besendahl A, Qiu Y-L, Lee J, Palmer JD, Bhattacharya D (2000) The cyanobacterial origin and vertical transmission of the plastid tRNA^{Leu} group-I intron. *Curr Genet* 37:12–23
- Bogorad L (1991) Replication and transcription of plastid DNA. In: Bogorad L, Vasil IK (eds) *The molecular biology of plastids: Cell culture and somatic cell genetics of plants*, Vol. 7A. Academic Press, San Diego, pp 93–124
- Bonnard G, Michel F, Weil JH, Steinmetz AA (1984) Nucleotide sequence of the split tRNA^{Leu} gene from *Vicia faba* chloroplasts: evidence for structural homologies of the chloroplast tRNA^{Leu} intron with the intron from the autosplicable *Tetrahymena* ribosomal RNA precursor. *Mol Gen Genet* 194:330–336
- Borsch T, Hilu KW, Quandt D, Wilde V, Neihuis C, Barthlott W (2003) Noncoding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms. *J Evol Biol* 16:558–576
- Cech TR (1988) Conserved sequences and structures of group I introns: building an active site for RNA catalysis—a review. *Gene* 73:259–271
- Cech TR, Damberger SH, Gutell RR (1994) Representation of the secondary and tertiary structure of group I introns. *Nature Struct Biol* 1:273–280
- Cheng YS, Lin C-H, Chen L-J (1997) Transcription and processing of the gene for spinach chloroplast threonine tRNA in a homologous *in vitro* system. *Biochem Biophys Res Comm* 233:380–385
- Christopher DA, Hallick RB (1990) Complex RNA maturation pathway for a chloroplast ribosomal protein operon with an internal tRNA cistron. *Plant Cell* 2:659–671
- Delp G, Igloi GL, Kössel H (1991) Identification of *in vivo* processing intermediates and of splice junctions of tRNAs from maize chloroplasts by amplification with the polymerase chain reaction. *Nucleic Acids Res* 19:713–716
- Frank DN, Pace NR (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu Rev Biochem* 67:153–180
- Galli G, Hofstetter H, Birnstiel ML (1981) Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature* 294:626–631
- Gegenheimer P (1996) Structure, mechanism, and evolution of chloroplast transfer RNA processing systems. *Mol Biol Rep* 22:147–150
- Graham SW, Reeves PA, Burns ACE, Olmstead RG (2000) Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *Int J Plant Sci* 161 (Suppl 6):S83–S96
- Gruissem W, Zurawski G (1985) Identification and mutational analysis of the promoter for a spinach chloroplast transfer RNA gene. *EMBO J* 4:1637–1644
- Gruissem W, Greenberg BM, Zurawski G, Prescott DM, Hallick RB (1983) Biosynthesis of chloroplast transfer RNA in a spinach chloroplast transcription system. *Cell* 35:815–828
- Gruissem W, Elsner-Menzel C, Latshaw S, Narita JO, Schaffer MA, Zurawski G (1986) A subpopulation of spinach chloroplast tRNA genes does not require upstream promoter elements for transcription. *Nucleic Acids Res* 14:7541–7556
- Hanley-Bowdoin L, Chua N-H (1987) Chloroplast promoters. *Trends Biochem* 12:67–70
- Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R, Sauquet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS, Chatrou LW (2003) Angiosperm phylogeny based on *matK* sequence information. *Am J Bot* 90:1758–1776
- Holschuh K, Bottomley W, Whitfield PR (1984) Structure of the spinach chloroplast genes for the D2 and 44 Kd reaction-center proteins of photosystem II and for tRNA^{Ser} (UGA). *Nucleic Acids Res* 12:8819–8834
- Huang J, Giannasi DE, Price RA (2005) Phylogenetic relationships in *Ephedra* (Ephedraceae) inferred from chloroplast and nuclear DNA sequences. *Mol Phylogenet Evol* 35:48–59
- Huelsenbeck JP, Ronquist FR (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755
- Ickert-Bond SM, Wojciechowski MF (2004) Phylogenetic relationships in *Ephedra* (Gnetales): evidence from nuclear and chloroplast DNA sequence data. *Syst Bot* 29:834–849
- Inokuchi H, Yamao F (1995) Structure and expression of prokaryotic tRNA genes. In: Söll D, RajBahndary U (eds) *tRNA: Structure, biogenesis, and function*. ASM Press, Washington, DC, pp 17–30
- Jahn D (1992) Expression of the *Chlamydomonas reinhardtii* chloroplast tRNA^{Glu} gene in a homologous *in vitro* transcription system is independent of upstream promoter elements. *Arch Biochem Biophys* 298:505–513

- Källersjö M, Farris JS, Chase MW, Bremer B, Fay MF, Humphries CJ, Petersen G, Seberg O, Bremer K (1998) Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants, and flowering plants. *Pl Syst Evol* 213:259–287
- Kanno A, Hirai A (1993) A transcription map of the chloroplast genome from rice (*Oryza sativa*). *Curr Genet* 23:166–174
- Kelchner SA (2000) The evolution of non-coding chloroplast DNA and its application in plant systematics. *Ann Mo Bot Gard* 87:482–498
- 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
- Koch MA, Dobes C, Matschinger M, Bleeker W, Vogel J, Kiefer M, Mitchell-Olds T (2005) Evolution of the *trnF*(GAA) gene in *Arabidopsis* relatives and the Brassicaceae family: monophyletic origin and subsequent diversification of a plastidic pseudogene. *Mol Biol Evol* 22:1032–1043
- Kuhnel MG, Strickland R, Palmer JD (1990) An ancient group I intron shared by Eubacteria and chloroplasts. *Science* 250:1570–1573
- Kunzmann A, Brennicke A, Marchfelder A (1998) 5' end maturation and RNA editing have to precede tRNA 3' processing in plant mitochondria. *Proc Natl Acad Sci USA* 95:108–113
- Lambowitz AM, Caprara MG, Zimmerly S, Perlman PS (1999) Group I and group II ribozymes as RNPs: clues to the past and guides to the future. In: Gesteland RF, Cech TR, Atkins JF (eds) *The RNA world*, 2nd ed. The world of modern RNA suggests a prebiotic RNA. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 451–485
- Leach DRF (1994) Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *Bioessays* 16:893–900
- Leal-Klevezas DS, Martínez-Soriano JP, Nazar RN (2000) Co-transcription of 5S rRNA-tRNA^{Arg}(ACG) from *Brassica napus* chloroplast and processing of their intergenic spacer. *Gene* 253:303–311
- Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4:203–221
- Marion-Poll A, Hibbert CS, Radebaugh CA, Hallick RB (1988) Processing of mono-, di- and tricistronic transfer RNA precursors in a spinach or pea chloroplast soluble extract. *Plant Mol Biol* 11:45–56
- Martin NC (1995) Organellar tRNAs: Biosynthesis and function. In: Söll D, RajBahndary U (eds) *tRNA: Structure, biogenesis, and function*. ASM Press, Washington, DC, pp 127–140
- Michel F, Westhof E (1990) Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* 216:585–610
- Morden CW, Wolfe KH, dePamphilis CW, Palmer JD (1991) Plastid translation and transcription genes in a non-photosynthetic plant: intact, missing and pseudo genes. *EMBO J* 10:3281–3288
- Nakamura T, Ohta M, Sugiura M, Sugita M (1999) Chloroplast ribonucleoproteins are associated with both mRNAs and intron containing precursor tRNAs. *FEBS Lett* 460:437–441
- Nasar F, Jankowski C, Nag DK (2000) Long palindromic sequences induce double-strand breaks during meiosis in yeast. *Mol Cell Biol* 20:3449–3458
- Ohme M, Kamogashira T, Sinozaki K, Sugiura M (1985) Structure and cotranscription of tobacco chloroplast genes for tRNA^{Glu}(UUC), tRNA^{Tyr}(GUA), and tRNA^{Asp}(GUC). *Nucleic Acids Res* 13:1045–1056
- Quandt D, Müller K, Stech M, Hilu KW, Frey W, Frahm J-P, Borsch T (2004) Molecular evolution of the chloroplast *trnL-F* region in land plants. *Monogr Syst Bot Mo Bot Gard* 98:13–37
- Schiffer S, Helm M, Théobald-Dietrich A, Giegé R, Marchfelder A (2001) The plant tRNA 3' processing enzyme has a broad substrate spectrum. *Biochemistry* 40:8264–8272
- Schiffer S, Rösch S, Marchfelder A (2002) Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzyme. *EMBO J* 21:2769–2777
- Schwarz Z, Jolly SO, Steinmetz AA, Bogorad L (1981) Overlapping divergent genes in the maize chloroplast chromosome and *in vitro* transcription of the gene for tRNA^{His}. *Proc Natl Acad Sci USA* 78:3423–3427
- Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am J Bot* 92:142–166
- Simon D, Fewer D, Friedl T, Bhattacharya D (2003) Phylogeny and self-splicing ability of the plastid tRNA-Leu group I intron. *J Mol Evol* 57:710–720
- Steinmetz AA, Krebbers ET, Schwarz Z, Gubbins EJ, Bogorad L (1983) Nucleotide sequences of five maize chloroplast transfer RNA genes and their flanking regions. *J Biol Chem* 258:5503–5511
- Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Pl Mol Biol* 17:1105–1109
- Tonkyn JC, Gruissem W (1993) Differential expression of the partially duplicated chloroplast S10 ribosomal protein operon. *Mol Gen Genet* 241:141–152
- van Ham RCHJ, 't Hart H, Mes THM, Sandbrink JM (1994) Molecular evolution of noncoding regions of the chloroplast genome in the Crassulaceae and related species. *Curr Genet* 25:558–566
- Vogel J, Hess WR (2001) Complete 5' and 3' end maturation of group II intron-containing tRNA precursors. *RNA* 7:285–292
- Vogel J, Börner T, Hess WR (1999) Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. *Nucleic Acids Res* 27:3866–3874
- Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc Natl Acad Sci USA* 91:9794–9798
- Wang MJ, Davis NW, Gegenheimer P (1988) Novel mechanisms for maturation of chloroplast transfer RNA precursors. *EMBO J* 7:1567–1574
- Wolfe KH, Morden CW, Ems SC, Palmer JD (1992) Rapid evolution of the plastid translation apparatus in a nonphotosynthetic plant: loss or accelerated sequence evolution of tRNA and ribosomal protein genes. *J Mol Evol* 35:304–317
- Won H, Renner SS (2003) Horizontal gene transfer from flowering plants to *Gnetum*. *Proc Natl Acad Sci USA* 100:10824–10829
- Won H, Renner SS (2005a) The internal transcribed spacer of nuclear ribosomal DNA in the gymnosperm *Gnetum*. *Mol Phylogenet Evol* 36 (in press)
- Won H, Renner SS (2005b) Bayesian relaxed clock estimates suggest that the gymnosperm *Gnetum* (Gnetales) dispersed over water between South America, Africa and Asia (submitted for publication)
- Wu C-Y, Lin C-H, Chen L-J (1997) Identification of the transcription site for the spinach chloroplast tRNA genes. *FEBS Lett* 418:157–161
- Xu MQ, Kathe SD, Goodrich-Blair H, Nierzwicki-Bauer SA, Shub DA (1990) *Science* 250:1566–1570
- Zaug AJ, McEvoy MM, Cech TR (1993) Self-splicing of the group I intron from *Anabaena* pre-tRNA: requirement for base-pairing of the exons in the anticodon stem. *Biochemistry* 32:7946–7953
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406–3415

Table A1. Species sampled, vouchers, GenBank accession numbers, and summary for length (bp) and G+C (%) content for chloroplast *trnT-trnL* IGS, *trnL* intron, and *trnL-trnF* IGS sequences

Taxon	Voucher	GenBank accession No.	Length (G+C)		
			<i>trnT-trnL</i> IGS	<i>trnL</i> intron	<i>trnL-trnF</i> IGS
Gnetaceae					
<i>G. acutum</i>	F. Markgraf Church et al. 284 (A)	AY296485	296 (31.4)	340 (37.4)	148 (43.2)
<i>G. africanum</i>	Welw. Essouma 99 (MO)	AY296486	292 (31.5)	384 (34.9)	165 (37.0)
	McPherson et al. 16261 (MO)	AY296487	292 (31.2)	384 (34.9)	165 (37.0)
<i>G. costatum</i>	K. Schum. Takeuchi et al. 15415 (MO)	AY296488	301 (34.2)	346 (37.3)	141 (39.0)
<i>G. cuspidatum</i> Bl.	Utami s.n. (BO)	AY296489	299 (31.1)	343 (37.9)	137 (42.3)*
	Won 562 (MO)	AY296490	299 (31.1)	343 (37.9)	137 (42.3)*
<i>G. diminutum</i>	F. Markgraf Postar & Won SAN 151122 (MO)	AY296491	299 (31.1)	344 (37.2)	140 (42.9)*
	Postar & Won SAN 151128 (MO)	AY296492	300 (31.0)	344 (37.2)	140 (42.9)*
<i>G. gnemon</i> L.	RBGE 19480122 (E)	AY296493	296 (31.1)	344 (36.6)	138 (36.2)
	Suksathan s.n. (QSB)	AY296494	296 (31.1)	345 (37.1)	138 (36.2)
	Won 514 (MO)	AY296495	307 (32.9)	344 (37.2)	140 (38.6)
<i>G. gnemonoides</i>	Brongn. Postar & Won SAN151121 (MO)	AY296496	311 (32.8)	351 (38.7)	166 (33.1)
	Utami s.n. (BO)	AY296497	311 (33.1)	329 (38.9)*	162 (34.0)
<i>G. hainanense</i> s.l.	RBGE 19550226 (E)	AY296498	320 (31.9)	338 (39.3)	137 (43.8)
	Won 555 (MO)	AY296499	320 (31.9)	338 (38.6)	137 (43.8)
	Won & Maxwell 580 (MO)	AY296500	320 (31.6)	338 (38.5)	138 (43.5)
	Won & Gang 600 (MO)	AY296501	320 (31.9)	340 (38.5)	138 (44.9)
	Jian s.n. (MO)	AY296502	320 (31.9)	338 (38.5)	137 (43.8)
<i>G. klossii</i>	Merrill ex F. Markgraf Postar & Won SAN 151117 (MO)	AY296503	309 (30.4)	334 (38.3)*	140 (41.4)
	Postar & Won SAN 151124 (MO)	AY296504	309 (30.4)	334 (38.3)*	140 (41.4)
<i>G. latifolium</i> Bl. s.l.	Won 524 (MO)	AY296505	320 (31.6)	338 (38.2)	142 (42.3)
	Won et al. 575 (MO)	AY296506	319 (31.7)	338 (38.2)	142 (41.5)
	Beaman et al. 9408 (MO)	AY296509	340 (31.2)	337 (37.7)	149 (41.6)
<i>G. aff. latifolium</i>	Won 545 (MO)	AY296507	320 (31.3)	328 (38.1)*	144 (41.7)
<i>G. aff. latifolium</i>	Postar & Won SAN 151116 (MO)	AY296508	320 (31.6)	339 (38.1)	145 (42.1)
<i>G. macrostachyum</i>	Hook. f. Won 535 (MO)	AY296511	300 (31.7)	338 (38.2)*	144 (43.1)
<i>G. aff. Microcarpum</i> Bl.	Won 526 (MO)	AY296512	300 (31.0)	347 (38.0)	147 (42.9)
	Won 533 (MO)	AY296513	299 (31.1)	346 (37.3)	147 (43.5)
<i>G. microstachyum</i> Benth. ex Tul.	Renske s.n. (U)	AY296514	304 (32.6)	329 (35.6)	132 (43.9)
<i>G. neglectum</i> Bl.	Postar & Won SAN151111 (MO)	AY296510	320 (32.5)	331 (39.3)*	147 (42.2)
<i>G. nodiflorum</i> Brongn.	Gillespie 4246 (US)	AY296515	301 (32.6)	347 (37.5)	158 (36.7)
<i>G. paniculatum</i> Spruce ex Benth.	Aymard et al. 12208 (MO)	AY296516	301 (32.2)	379 (34.8)	154 (38.3)
<i>G. parvifolium</i> (Warb.)	Cheng ex Chun Won & Gang 601 (MO)	AY296517	320 (31.6)	336 (37.5)	140 (45.7)
<i>G. raya</i>	F. Markgraf Kessler et al. 624 (MO)	AY296518	302 (32.1)	342 (39.2)	160 (39.4)
<i>G. aff. Schwackeanum</i> Taub. ex F. Markgraf	Won 512 (MO)	AY296519	301 (31.9)	319 (35.1)*	169 (34.3)
<i>G. tenuifolium</i> Ridl.	Carlquist 8087 (MO)	AY296520	300 (31.3)	343 (36.4)	141 (41.8)
<i>G. ula</i> Brongn.	Maheshwari s.n. (UCB)	AY296521	300 (32.0)	334 (38.0)	145 (42.1)
<i>G. urens</i> (Aubl.) Bl.	Prévost 4094 (MO)	AY296522	300 (32.3)	329 (35.6)	147 (39.5)
<i>G. woodsonianum</i> (F. Markgraf) H. Won stat. nov.	Won 513 (MO)	AY296523	300 (32.0)	331 (35.3)	154 (37.0)
<i>G. sp.</i> Harder et al. 5621	Harder et al. 5621 (MO)	AY296524	300 (32.7)	340 (38.5)	133 (43.6)
<i>G. sp.</i> (<i>Vinkella</i> nom. nud.)	Takeuchi et al. 7049 (MO)	AY296525	300 (32.0)	332 (37.4)	140 (40.0)
Welwitschiaceae					
<i>Welwitschia</i>	Olson s.n. (MO)	AY513733	313 (30.4)	317 (37.4)	250 (33.2)
Ephedraceae					
<i>Ephedra trifurca</i> Torr.	Ickert-Bond SB753 (ASU)	AY513734	156 (28.8)	290 (26.6)	–
<i>Ephedra torreyana</i> Coville	Ickert-Bond SB941 (ASU)	AY513736	156 (28.8)	290 (36.9)	–
<i>Ephedra viridis</i> S. Wats.	Ickert-Bond SB954 (ASU)	AY513735	156 (28.8)	290 (37.2)	–

*Indicates incomplete sequence.

Table A2. List of informative gaps in the *Gnetum* chloroplast tRNA^{Leu} intron and adjacent IGS regions: Base positions 142–181, 246–260, 297–301, 576–606, 706–745, 810–853, 976–993, and 1012–1070, which include these gaps, were excluded from the analyses

Code	Location (bp)	Region	Distribution of indel
A	142–161	<i>trnT-trnL</i> IGS	<i>G. hainanense s.l.</i> , <i>G. parvifolium</i> , <i>G. latifolium s.l.</i> , <i>G. neglectum</i>
B	267–271	<i>trnT-trnL</i> IGS	<i>G. africanum</i> , S. American spp.
C	297–301	<i>trnT-trnL</i> IGS	<i>G. gnemonoides</i> , <i>G. raya</i>
D	364	<i>trnT-trnL</i> IGS	S. American spp.
E	396–402	<i>trnT-trnL</i> IGS	<i>G. africanum</i> , SE Asia clade I spp., S. American spp.
F	662–670	<i>trnL</i> intron	<i>G. microstachyum</i> , <i>G. schwackeanum</i> , <i>G. urens</i> , <i>G. woodsonianum</i>
G	741–744	<i>trnL</i> intron	<i>G. acutum</i> , <i>G. cuspidatum</i> , <i>G. diminutum</i> , <i>G. klossii</i> , <i>G. macrostachyum</i> , <i>G. microcarpum</i> , <i>G. tenuifolium</i> , <i>G. gnemon</i> , <i>G. costatum</i>
H	749	<i>trnL</i> intron	SE Asia clade I spp.
I	1056–1062	<i>trnL-trnF</i> IGS	S. American spp.
J	1069–1070	<i>trnL-trnF</i> IGS	S. American spp.
K	1106–1113	<i>trnL-trnF</i> IGS	<i>G. hainanense s.l.</i> , <i>G. parvifolium</i>

Table A3. Sequence divergences (K-2-P) for the cp *trnL* intron and adjacent IGS regions

	S. America	Africa	Asia I	Asia II
<i>trnL</i> Intron				
S. America	0–0.022			
Africa	0.030–0.040	0		
Asia I	0.026–0.057	0.024–0.045	0–0.032	
Asia II	0.022–0.062	0.024–0.040	0.018–0.045	0–0.025
<i>trnT-trnL</i> IGS				
S. America	0–0.017			
Africa	0.108–0.120	0.003		
Asia I	0.052–0.100	0.083–0.124	0–0.076	
Asia II	0.046–0.091	0.078–0.123	0.014–0.092	0–0.046
<i>trnL-trnF</i> IGS				
S. America	0–0.061			
Africa	0.124–0.286	0		
Asia I	0.069–0.288	0.082–0.257	0–0.225	
Asia II	0.048–0.114	0.059–0.088	0.024–0.137	0–0.040
Total				
S. America	0–0.023			
Africa	0.077–0.112	0.001		
Asia I	0.047–0.106	0.060–0.106	0.003–0.064	
Asia II	0.043–0.066	0.052–0.072	0.023–0.070	0–0.027