View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Ghent University Academic Bibliography

American Journal of Botany 94(6): 1003–1016. 2007.

ANCIENT PARALOGY IN THE CPDNA *TRNL-F* REGION IN ANNONACEAE: IMPLICATIONS FOR PLANT MOLECULAR SYSTEMATICS[1]

MICHAEL D. PIRIE,[2,4] MARIA PAULA BALCÁZAR VARGAS,[2] MARLEEN BOTERMANS,[3] FREEK T. BAKKER,[3] AND LARS W. CHATROU[2,3]

[2]Nationaal Herbarium Nederland, Universiteit Utrecht branch, Heidelberglaan 2, 3584 CS Utrecht, Netherlands; and [3]Nationaal Herbarium Nederland, Wageningen Universiteit branch, Generaal Foulkesweg 37, 6703 BL Wageningen, Netherlands

The plastid *trnL-F* region has proved useful in molecular phylogenetic studies addressing diverse evolutionary questions from biogeographic history to character evolution in a broad range of plant groups. An important assumption for phylogenetic reconstruction is that data used in combined analyses contain the same phylogenetic signal. The *trnL-F* region is often used in combined analyses of multiple chloroplast markers. These markers are assumed to contain congruent phylogenetic signal due to lack of recombination. Here we show that *trnL-F* sequences display a phylogenetic signal conflicting with that of other chloroplast markers in Annonaceae, and we demonstrate that this conflict results from ancient paralogy. *TrnL-F* copy 2 diverged from *trnL-F* copy 1 (as used in family-wide phylogenetic analyses) in a direct ancestor of the Annonaceae. Although this divergence dates back 88 million years or more, the exons of both copies appear to be intact. In this case, assuming that (putative) chloroplast markers contain the same phylogenetic signal results in an incorrect topology and an incorrect estimate of ages. Our study demonstrates that researchers should be cautious when interpreting gene phylogenies, irrespective of the genome from which they are presumed to have been sampled.

Key words: Annonaceae; chloroplast DNA sequences; conflict; molecular dating; paralogy; phylogeny reconstruction; *trnL-F*.

The cpDNA *trnT-F* region in land plants consists of the transfer RNA genes $trnT_{ugu}$, $trnL_{uaa}$, and $trnF_{gaa}$ arranged in tandem and separated by noncoding spacer regions. The region is positioned in the large single copy region, approximately 8 kb downstream of *rbcL*. The *trnL* gene of cyanobacteria and a number of chloroplast genomes, including that of all land plants, contains a group I intron positioned between the U and the A of the UAA anticodon loop. This intron is inferred from phylogenetic analysis to have been present in the cyanobacterial ancestor of the plastid lineages of Rhodophyta, Chlorophyta, and Glaucocystophyta and to have been subsequently vertically transmitted (Besendahl et al., 2000).

The succession of conserved *trn* genes and the apparent absence of gene rearrangements in the *trnT-F* region made the design of plant universal primers possible (Taberlet et al., 1991). As a consequence, the *trnL-F* region, comprising the *trnL* intron and *trnL-F* spacer, has become one of the most widely used chloroplast markers for phylogenetic analyses in plants (Quandt et al., 2004). The accumulation of an increasingly large number of sequences of the *trn*(*T*-)*L-F* region from a wide range of plants has allowed further study of structures, functions, and evolution in different orders of flowering plants (Bakker et al., 2000), in basal angiosperms (Borsch et al., 2003), in land plants (Quandt et al., 2004), in bryophytes (Quandt and Stech, 2004), and in Gnetales (Won and Renner, 2005).

Sequences from the *trnL-F* region (excluding the *trnT-L* region and *trnL* 5′ exon) have recently been used, in combination with those from further chloroplast markers *rbcL* and *matK*, as a source of characters for phylogenetic reconstruction in the tropical flowering plant family Annonaceae Juss. These phylogenies have been used to answer questions about morphological character evolution (Doyle et al., 2000; Sauquet et al., 2003), classification (Mols et al., 2004), biogeography (Richardson et al., 2004; Pirie et al., 2006), and molecular dating (Pirie et al., 2005). These markers appeared to contain complementary phylogenetic signals, as is expected from different sequences sampled from the plastid genome (Chase and Cox, 1998), and were thus applied in combined analyses. The combined analyses yielded better resolved phylogenies, subject to higher levels of support, than those derived from individual markers.

This was not, however, the case for analyses including sequences of the Neotropical genus *Unonopsis* R.E.Fr. Phylogenetic analysis of Annonaceae *trnL-F* sequences with other Magnoliales and Laurales outgroups (Mols et al., 2004; Richardson et al., 2004; Pirie et al., 2005) suggested a monophyletic *Unonopsis* as sister group to the rest of the Annonaceae (Fig. 1). This result directly conflicted with other plastid DNA sequence data and morphology: *Unonopsis* has been grouped with two smaller South American genera, *Bocageopsis* R.E.Fr. and *Onychopetalum* R.E.Fr. (comprising four species each compared to the 38 of *Unonopsis*), on the basis of morphological similarity (Van Heusden, 1992; Van Setten and Koek-Noorman, 1992). Phylogenetic analysis of multiple chloroplast markers supports monophyly of the *Unonopsis*/*Bocageopsis*/*Onychopetalum* clade, placed with high support (100% maximum parsimony bootstrap) within

[1] Manuscript received 24 April 2006; revision accepted 8 May 2007.

The authors thank S. Renner, an anonymous reviewer, and A.I. Rees and B. Gehrke for critically reading the manuscript; T. Borsch for advice on *trnL* intron secondary structures, and J. Maas for assistance in the lab. The Molecular Systematics group of the National Herbarium of the Netherlands, and the Linder lab at the Institute for Systematic Botany, Zürich, provided useful discussion of the issues involved.

[4] Author for correspondence (e-mail: michael.pirie@systbot.unizh.ch); present address: Institute for Systematic Botany, Zollikerstrasse 107, CH 8008, Zürich, Switzerland

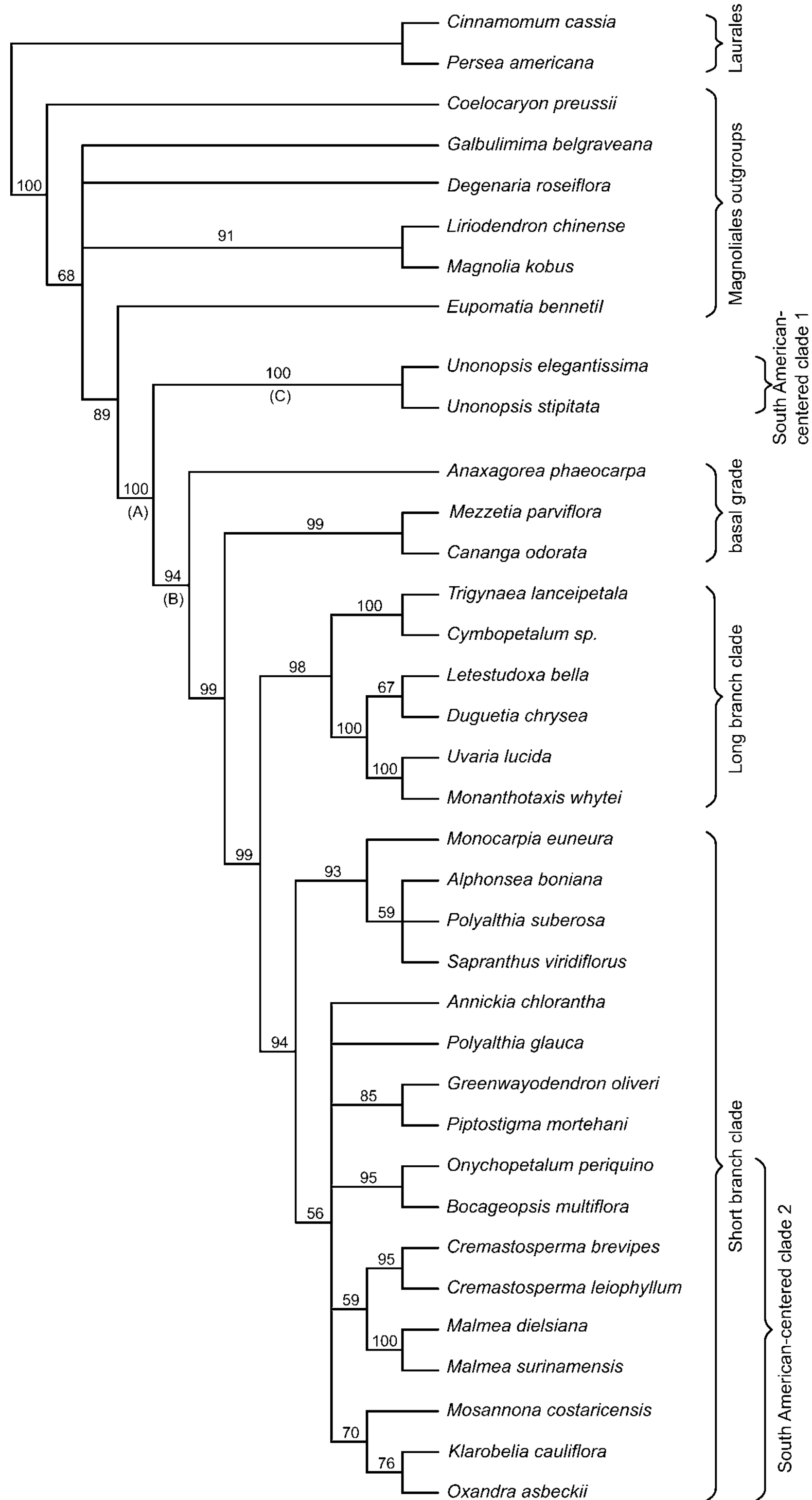


Fig. 1. Maximum parsimony 50% majority rule bootstrap consensus tree resulting from analysis of the entire *trnL-F* region for a selection of Annonaceae, other Magnoliales, and Laurales outgroups. Bootstrap support is indicated above the nodes. Nodes labeled A, B, and C are those defining *Unonopsis* as a monophyletic sister group to Annonaceae.

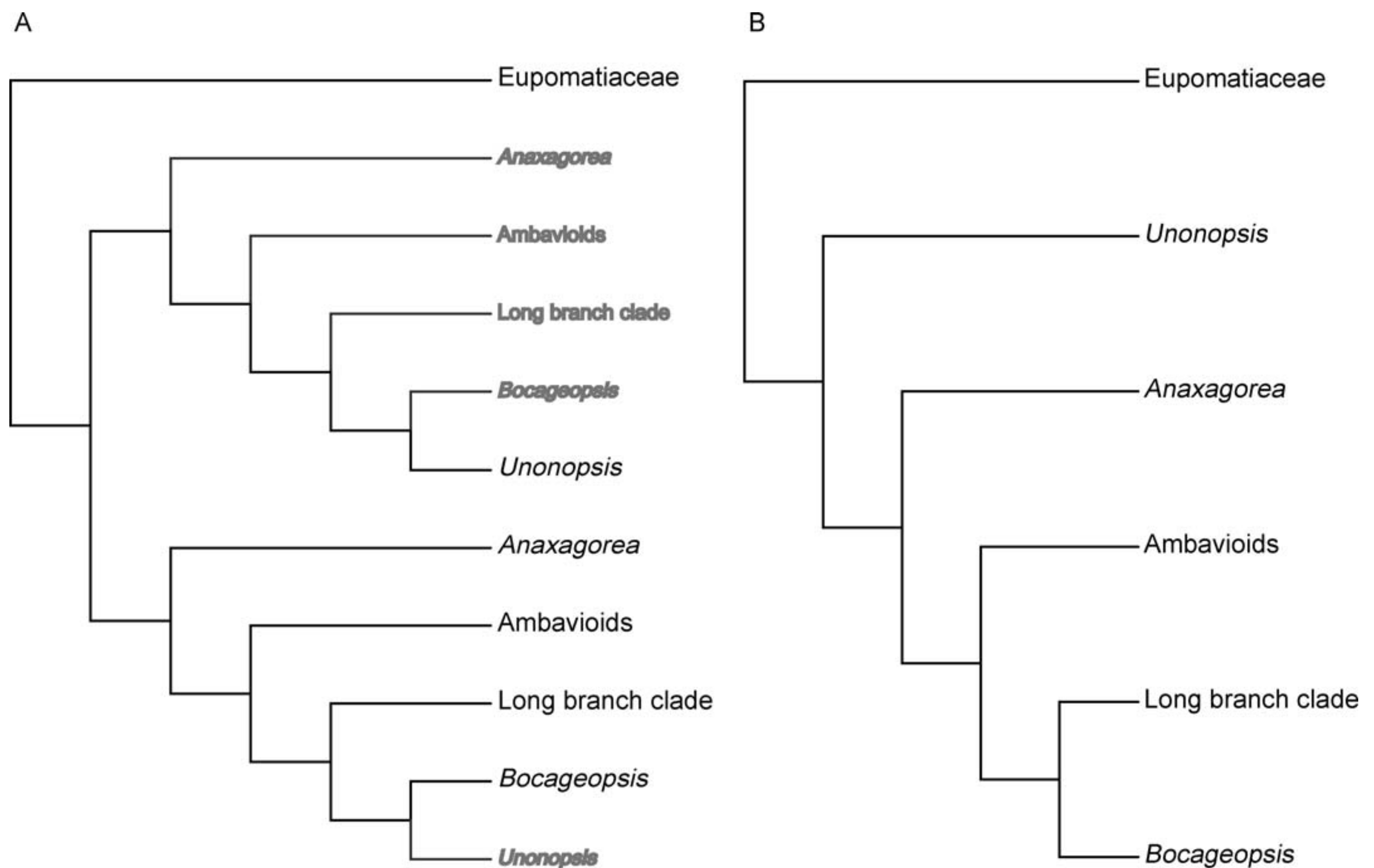


Fig. 2. Illustration of the manner in which preferential amplification and sequencing of non-homologous copies of *trnL-F* could result in conflicting gene phylogenies. In (A), the major clades in Annonaceae (and congruent relationships between them) are each represented twice, by two copies of a marker duplicated after divergence with the sister family Eupomatiaceae. One of the two copies (represented in gray) is missed during sequencing, resulting after phylogenetic analysis in the reconstructed topology (B).

the South American-centered clade, which itself is nested within the equally highly supported short branch clade (Pirie et al., 2006; for further justification of names applied to major clades in Annonaceae, see fig. 1 of Richardson et al., 2004).

Incongruent results can have a number of potential explanations. These include weak signal in the data, failings of the inference method under particular conditions, and incorrect assessment of homology during sequence alignment (e.g., Sanderson and Shaffer, 2002). No such problems were immediately obvious. The position of *Unonopsis* in the *trnL-F* phylogeny received high bootstrap support. The alignment of sequences was unambiguous. No unusually large insertion/deletions were observed, nor was the subtending branch, representing changes in the two *Unonopsis* sequences, unusually long.

Should incongruence reflect real differences in individual putative chloroplast gene trees, this might be explained by recombination, heteroplasmy, or paralogy (Wolfe and Randle, 2004). Evidence for recombination in chloroplasts is sparse, with reports limited to examples in gymnosperms (e.g., *Pinus contorta* Dougl. [Marshall et al., 2001] and *Cycas taitungensis* C.F. Shen, K.D. Hill, C.H. Tsou & C.J. Chen [Huang et al., 2001]), and heteroplasmy is generally regarded as an unstable phenomenon, observed over timescales of a few generations (Wolfe and Randle, 2004). Recombination and heteroplasmy seem unlikely given the topology presented in Fig. 1. The hypothetical donor of either an additional chloroplast genome or recombinant sequence would have to be a member of a currently unknown Magnoliales lineage descendent from before the most recent common ancestor (MRCA) of the Annonaceae. In the absence of recombination or heteroplasmy, incongruence between the phylogenetic signal of different chloroplast markers could be the result of paralogy. The node of the MRCA of putatively orthologous and paralogous sequences in the *trnL-F* (gene) phylogeny would represent the most recent possible divergence of the two paralogues. The phylogeny presented in Fig. 1 could then be reconstructed if one paralogue was amplified by PCR in *Unonopsis* and the other in the remaining Annonaceae (as illustrated in Fig. 2).

The inference of paralogy in this case leads to two predictions: first, *trnL-F* sequences with the same signal as the other chloroplast markers should be (or should have been) present in *Unonopsis*. Second, if paralogy is the result of duplication or transfer of a copy to another genomic compartment, a second copy of *trnL-F* should be (or should have been) present in all other taxa descending from the node representing the inferred duplication or transfer event. According to the preliminary results, this would include all extant Annonaceae.

In this paper we first use PCR-based and phylogenetic

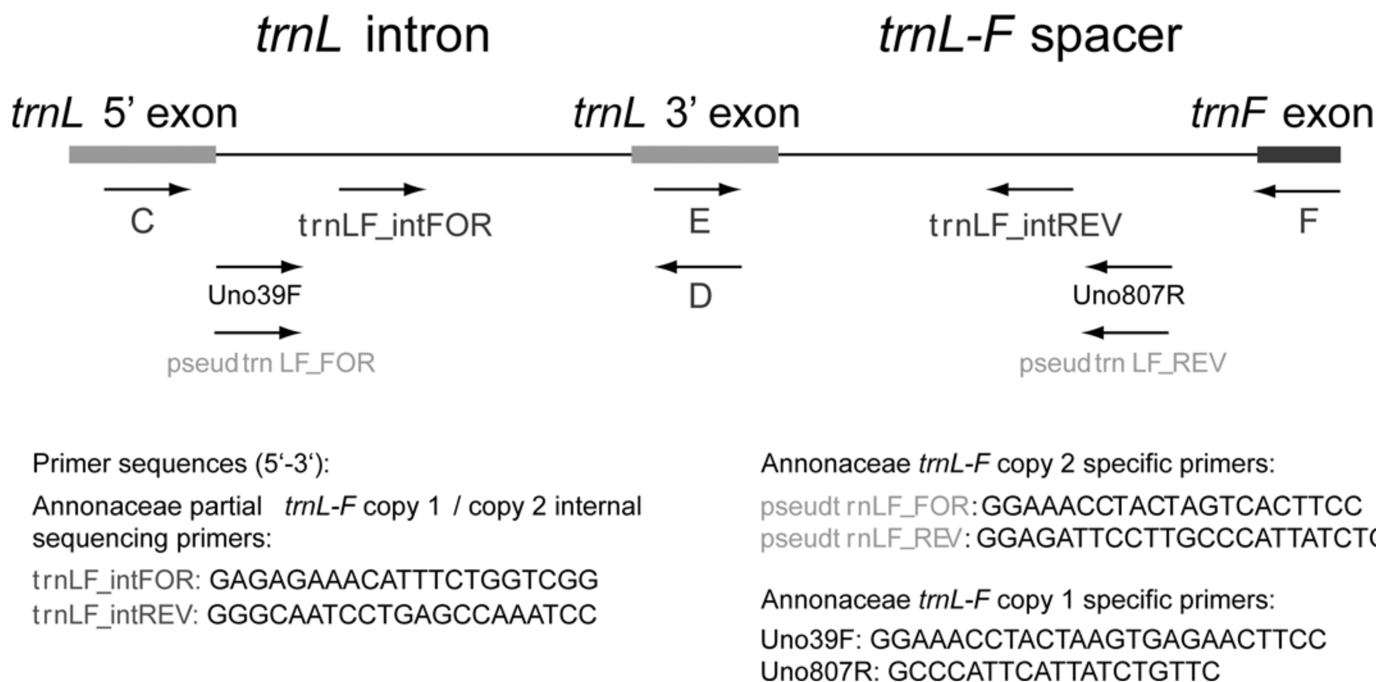


Fig. 3. Scheme of primers used to amplify and sequence *trnL-F* copies 1 and 2.

analysis techniques to test the hypothesis that the appearance of conflicting phylogenetic signals between the *trnL-F* region and other chloroplast markers in Annonaceae is the result of analysis of paralogous sequences. Having confirmed the existence of two copies of *trnL-F*, we then draw further conclusions with respect to the timing of divergence of these copies and the phylogenetic signal they contain.

Support for the paralogy hypothesis raises further issues concerning the definition of homology in Magnoliales *trnL-F* regions. To address the question of functional homology, we compare *trnL* gene and intron sequences obtained in this study with the secondary structures and corresponding functional constraints proposed for this region in studies across land plants (Borsch et al., 2003; Quandt et al., 2004). Positional homology and the precise origin of paralogues are not easily determined from sequences alone, although comparison of the rates of evolution exhibited by each copy may indicate in which genome each copy is located. We also discuss the further implications of cryptic paralogy in chloroplast markers for phylogeny reconstruction and molecular systematics in general.

MATERIALS AND METHODS

Taxon sampling—Recent improvements in both phylogenetic resolution and representation of taxa (Sauquet et al., 2003; Mols et al., 2004; Richardson et al., 2004; Pirie et al., 2006) provide a robust framework for the choice of taxa in phylogenetic reconstruction in Magnoliales. This study utilized previously unpublished sequence data, as well as published *trnL-F*, *rbcL*, *matK*, and *psbA-trnH* sequences (Kojoma et al., 2002; Sauquet et al., 2003; Mols et al., 2004; Pirie et al., 2005, 2006; L. W. Chatrou et al., unpublished data; see Appendix 1).

DNA extraction, PCR amplification, and sequencing—Total genomic DNA was extracted using a modified cetyl trimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1987): 50 mg silica-dried or herbarium leaf material was homogenized in 1300 μL CTAB and incubated for 20 min with 12 μL 2-mercaptoethanol at 65°C, followed by 90 min mixing at room temperature with 1 mL 24 : 1 chloroform : isoamylalcohol. After 10 min of centrifugation at 13 ,000 rpm, 300 μL supernatant was purified using the Wizard DNA purification system (Promega, Madison, Wisconsin, USA) (i.e., without isopropanol precipitation, thus avoiding the co-precipitation of oxidized material; Savolainen et al., 1995).

A standard PCR protocol was used throughout, with the addition of 1 μL 0.4% bovine serum albumin (BSA) per 25 μL reaction (which was found to increase amplification in all samples): initial denaturing of 4 min at 94°C; 35 cycles of 30 s at 94°C, 1 min at 55°–58°C, and 2 min at 72°C; and a final extension of 7 min at 72°C. The PCR products were purified using QIAquick PCR purification kits (Qiagen, Venlo, The Netherlands), sequenced with selected-PCR and specially designed sequencing primers (see below), and analyzed by electrophoresis using an automatic sequencer ABI 3730XL (Applied Biosystems, Foster City, California, USA).

BLAST searching (Altschul et al., 1997) was employed using the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/BLAST/) to compare the *Unonopsis trnL-F*s with published sequences. To avoid confusion, the *trnL-F* copy homologous with those previously sequenced in Annonaceae will be referred to as *trnL-F* copy 1 and that homologous with those first sequenced only in *Unonopsis* will be referred to as *trnL-F* copy 2.

For taxa other than *Unonopsis* the *trnL-F* region was amplified and sequenced using plant universal primers of Taberlet et al. (1991) in combination C/F or C/D and E/F. (The positions and sequences of all primers used to amplify and sequence both copies of the *trnL-F* region are presented in Fig. 3). Other primers were designed to specifically amplify and sequence the different *trnL-F* copies. In *Unonopsis*, *trnL-F* copy 1 was amplified using primers Uno39F and Uno807R. In other Annonaceae (and in *Unonopsis* samples where sequencing using primers C/D, E/F failed), *trnL-F* copy 2 was amplified using primers pseudtrnLF_FOR and pseudtrnLF_REV. Uno39F and pseudtrnLF_FOR were designed to anneal to the same region near the beginning of the *trnL* intron, where length differences were specific to the different copies; the same was true of Uno807 and pseudtrnLF_REV, located within the *trnL-F* intergenic spacer (Fig. 4). The higher annealing temperature of 58°C was employed to reduce the chances of noncopy-specific annealing. The PCR products amplified using both sets of copy-specific primers were sequenced using primers trnLF_intFOR and trnLF_intREV, which were designed to anneal within the amplified fragments.

```
trnL  3' exon<=**************............................ trnL intron=>
Persea americana              ????????????GGTATGGAAACCTACTAAGTGATAACTTCCAAATT
Coelocaryon preussii          ????????????GGTATGGAAACCTACTAAGTGGTAACTTCCAAATT
Magnolia kobus                ????????????GGTAGGGAAACCTAcTAAGTGGTACCTTCCAAATT
Eupomatia bennetii            ????????????GGTATGGAAACCTACCAAGTGGTAACTTCCAAATT
Malmea dielsiana              ????????????GGTATGGAAACCTACTAAGTGAGAACTTCCAAATT
Cremastosperma brevipes       GGATTGAGCCTTGGTATGGAAACCTACTAAGTGAGAACTTCCAAATT
Unonopsis elegantissima       GGATTGAGCCTTGGTATGGAAACCTACTA-GTCA---CTTCCAAATT
Unonopsis stipitata           GGATTGAGCCTTGGTATGGAAACCTACTA-GTCA---CTTCCAATCT
pseudtrnLF_FOR                                  GGAAACCTACTA-GTCA---CTTCC=>
Uno39F                                          GGAAACCTACTAAGTGAGAACTTCC=>

                              <=......... trnL-F intergenic spacer..........=>
Persea americana              TACAAATGAACATAT-AT----AGGCAAGGAATTTCCATTATTAAAT
Coelocaryon preussii          CACAAATGAACAGATTAT----GGGCAAGGAATCCCCATTATTGAAT
Magnolia kobus                TACAAATGCCCATATTAT----GGGCAAGGAATCTCCATTATTGAAT
Eupomatia bennetii            TACAAATGAACATATTAT----GGGCAAGGAATCTCCATTATTGAAT
Malmea dielsiana              TACAAATGAACAGATAATGAATGGGC---GAATCTCCACTATTGAAT
Cremastosperma brevipes       TACAAATGAACAGATAATGAATGGG----GAATCTCCACTATTGAAT
Unonopsis elegantissima       TACAAATGAACAGATAAT----GGGCAAGGAATCTCCATTATTGAAT
Unonopsis stipitata           TACAAATGAACAGATAAT----GGGCAAGGAATCTCCATTATTGAAT
pseudtrnLF_REV                        <=CAGATAAT----GGGCAAGGAATCTCC
Uno807R                            <=GAACAGATAATGAATGGGC
```

Fig. 4. Copy-specific primer sequences, illustrating length differences in *trnL-F* intron and spacer.

Analysis—Two matrices were constructed: (1) a *trnL-F* matrix including both *trnL-F* copies (thus with some taxa present twice) and (2) a multigene matrix also including *rbcL*, *matK*, and *psbA-trnH*, where *trnL-F* copies 1 and 2 were treated as separate partitions (and where taxa were thus present only once). In the latter case, sequences of *trnL-F* copies 1 and 2 (as identified from analysis of the first matrix) were both aligned with non-Annonaceae outgroup *trnL-F* sequences (which were thus present in both partitions), and Annonaceae taxa for which *trnL-F* copy 2 sequences were unavailable (i.e., non-South American centered clade taxa) were excluded.

DNA sequences were edited in SeqMan 4.0 (DNAStar, Madison, Wisconsin, USA) and aligned by eye. All areas of the alignment in which the assessment of homology was ambiguous were excluded from the analyses. In the analyses in which sequences of *trnL-F* copy 1 (in *Unonopsis*) or copy 2 (in other Annonaceae) were incomplete (because of the position of the copy-specific primers), the corresponding stretches of the alignment at both ends were excluded to facilitate direct comparison of information content and rate of change independent of the influence of missing data biased toward copy 2.

Maximum parsimony (MP) analysis—Data were analyzed using the parsimony algorithm of the software package PAUP* version 4.0b10 (Swofford, 2000), assuming unordered character state transformation (Fitch parsimony; Fitch, 1971) and equal weights. The lengths of the shortest trees were estimated with "full" heuristic searches of 1000 random addition sequences (RAS), with tree bisection and reconnection (TBR), saving a maximum of 100 trees in each RAS. Support was estimated using bootstrap analyses of 500 replicates with full heuristic searches of 100 RAS, with TBR, saving a maximum of 50 trees each RAS. Bootstrap percentages were interpreted following Richardson et al. (2004): 50–74% represent weak support, 75–84% moderate support, and 85–100% strong support. For the multigene matrix, support was estimated for the markers independently and, where no supported conflict was observed, in combined analysis (Mason-Gamer and Kellogg, 1996).

Selecting the best-fitting DNA substitution model—ModelTest 3.06 (Posada and Crandall, 1998) was used to select the substitution model that best fit each sequence data partition for each matrix using a most parsimonious tree topology. For matrix 1, ModelTest was run both with and without non-Annonaceae sequences to check whether omitting the outgroups (and their relatively long branches) could have resulted in different models and parameters.

Bayesian analysis—Bayesian inference was applied as implemented in MrBayes version 3.0 (Huelsenbeck, 2000). The use of Markov chain Monte Carlo analyses (MCMC; Geyer, 1991) in Bayesian inference facilitates heuristic searching of parameter value space for maximum likelihood models of DNA substitution in phylogeny reconstruction (Huelsenbeck et al., 2001). The prior model for DNA substitution was determined using ModelTest. Prior probabilities for all topologies were assumed to be equal. *Persea americana* (Lauraceae) was chosen as the single outgroup taxon permitted by MrBayes for the *trnL-F* matrix (1), and *Coelocaryon preussii* (Myristicaceae, sister group to the rest of Magnoliales; Sauquet et al., 2003) was chosen as the single outgroup taxon for the multigene matrix (2). In the multigene matrix the data were partitioned according to the separate markers, and both rates and substitution models were allowed to vary across the partitions. The MCMC analyses were run for 5 000 000 generations with four simultaneous MCMC chains to calculate posterior probabilities (PP), saving one tree per 100 generations. The burn-in values were determined empirically from the log-likelihood values, and 50% majority rule consensus trees were calculated together with approximations of the PP for the observed bipartitions. The PP values of 95% and above were considered to represent significant support.

r8s analyses—Three different data partitions were used to estimate the ages and rates of particular nodes: (1) *trnL-F*, including both copies aligned together; (2) *trnL-F* copy 1; and (3) the combined *trnL-F* copy 1, *matK*, *rbcL*, and *psbA-trnH*. In (1), those Annonaceae taxa for which both copies of *trnLF* were available were included, plus outgroups, leaving a total of 33 sequences. All

missing data were excluded, leaving 635 characters. A single MP topology was selected from the three resulting from a heuristic search of the *trnL-F* matrix (1), having constrained the relationships between Magnoliales outgroups to conform to those demonstrated by Sauquet et al. (2003). These would otherwise have remained unresolved because of the much smaller number of characters being analyzed.

In (2) and (3), non-short branch clade Annonaceae taxa were added, bringing the number of taxa to 20. *Unonopsis* species were represented either by (a) homologous (copy 1) sequences or (b) paralogous (copy 2) sequences to investigate the effect of incorrect homology assessment on age and rate estimations. When including paralogous *Unonopsis trnL-F* copy 2 sequences instead of homologous copy 1, the topology was constrained to force *Unonopsis* into the position it occupies when homologous copy 1 sequences are analyzed. For (3b), this constraint was unnecessary because the signal of *trnL-F* copy 2 was overridden by the other markers.

A likelihood ratio test was performed on the selected topologies: likelihoods of the data with and without constraint of a molecular clock were calculated, and the likelihood ratio statistic compared with χ^2 critical value with 31 degrees of freedom (i.e., number of sequences − 2). The ML branch lengths were then calculated using the substitution model calculated as above with (a) the original matrix and (b) 100 bootstrap resampled matrices. Thereafter, the penalized likelihood (PL) method of Sanderson (2002a) was applied using the program r8s (Sanderson, 2002b) to estimate rates and divergence times. Application of the cross-validation procedure determined the optimal smoothing parameter to be 31.62. This value was applied in r8s analyses using the ML phylogram based on the original data (for the point estimates), and using the 100 phylograms based on the bootstrap resampled matrices (to derive SD for these values). To calibrate the rate-smoothed tree in absolute time, the age of the fossil *Archaeanthus* (Dilcher and Crane, 1984) was used to impose a minimum age of 98 million years for the Magnoliaceae stem node (following Doyle et al., 2004; Richardson et al., 2004; Pirie et al., 2006).

Secondary structure of the trnL gene and intron—The secondary structures of the *trnL* gene and intron were calculated for representatives of both *trnL-F* copies. Stem-loop regions were identified by comparison to the structure proposed by Borsch et al. (2003) for *Nymphaea odorata*, with further reference to the conserved sequence motifs reported across land plants by Quandt et al. (2004). Secondary structures of these regions were then estimated individually using Mfold (Zuker, 2003), except in the case of the more variable P8 region. The precise structure of the P8 region in angiosperms is not fully understood (D. Quandt, Technische Universität, Dresden, personal communication), and because no conserved regions within P8 have been identified, sequences produced in this study could not be meaningfully compared. This region was therefore not considered further.

RESULTS AND DISCUSSION

Robustness of the position of Unonopsis in the trnL-F phylogeny—BLAST search (Altschul et al., 1997) identified chloroplast *trnL-F* regions derived from species of Magnoliaceae—a family of the same order as Annonaceae, the Magnoliales (APG II, 2003; Sauquet et al., 2003)—to be most similar to the *Unonopsis trnL-F* sequences. BLAST-based methods assume identical divergence rates and are therefore not suitable for inferring relatedness of the sequences (Thornton and DeSalle, 2000). However, this result would appear to exclude the possibility of the *Unonopsis trnL-F*s being related either to chloroplast sequences from outside the taxonomic scope of our analyses or to tRNA genes of other genomic compartments inherited from an ancestor in common with closely studied taxa, such as *Arabidopsis* (Arabidopsis Genome Initiative, 2000).

The alignment of the *trnL-F* matrix was 1222 characters long, of which 139 were excluded from the analyses. Of the remaining characters, 359 were variable and 208 parsimony informative. The length of the shortest tree was 561, CI = 0.772, RI = 0.881 (based on an arbitrary MP topology). The best-fitting model was K81uf+G. Reconstruction of the phylogeny of the entire *trnL-F* region (including complete sequences of both copies initially generated using universal primers) using both Bayesian inference and MP resulted in topologies congruent with that presented in Fig. 1. The nodes defining the position of the *Unonopsis trnL-F* sequences as monophyletic sister group to the rest of the Annonaceae (A, B, and C in Fig. 1) were subject in all cases to strong BS and 100% PP.

It appears unlikely that the incongruent position of *Unonopsis* sequences in the *trnL-F* phylogeny can be explained by errors in the analyses. To test possible sensitivity of the result to alignment, ClustalX (Thompson et al., 1997) was applied with default multiple alignment parameters. The entire, unedited, resulting alignment was analyzed using MP and Bayesian inference, which recovered nodes A, B, and C with moderate to strong BS and >95% PP (data not shown). Long-branch attraction seemed to be an unlikely explanation of the result, as no potentially attracting long branches (Siddall and Whiting, 1999) are evident in this part of the topology. The rest of the topology is consistent with results derived from other data (e.g., Richardson et al., 2004). Finally, applying ModelTest with and without non-Annonaceae sequences resulted in the same best-fitting substitution model (K81uf+G).

Paralogy in Annonaceae trnL-F sequences—The PCR-based approach employed here resulted in the amplification and sequencing of *trnL-F* copy 1 in *Unonopsis* and of *trnL-F* copy 2 in accessions of *Bocageopsis*, *Cremastosperma*, *Malmea* R.E.Fr., *Onychopetalum*, and *Oxandra* A.Rich. The *Unonopsis* copy 1 sequences formed a monophyletic group with *Bocageopsis* and *Onychopetalum* copy 1 sequences in the short branch clade. The latter taxa belong to the same subclade as *Bocageopsis* and *Onychopetalum*, within the short branch clade. Their copy 2 sequences formed a monophyletic group with the *Unonopsis trnL-F* copy 2 sequences (Fig. 5). Copy-specific amplification was not always successful: in some taxa of the South American-centered clade (e.g., *Pseudoxandra* R.E. Fr.) and accessions of further short branch clade taxa (such as *Annickia* Setten & Maas and *Polyalthia* Blume), the *trnL-F* copy 2 specific primers instead amplified *trnL-F* copy 1. When applied to accessions of long-branch clade or basal grade taxa, no amplified product was produced. It is possible that the small length differences (indels) used as targets for the copy-specific primers are only present in the South American-centered clade, either representing synapomorphies or symplesiomorphies secondarily lost in the other clades sampled. This result therefore offers no direct evidence for the presence or absence of *trnL-F* copy 2 in other clades in Annonaceae.

Where two paralogues represented in a gene tree share a MRCA, the age of that MRCA represents the most recent possible age of the divergence of the paralogues (Thornton and DeSalle, 2000). In the case of the divergence of *trnL-F* copies, this age is younger than the MRCA of Annonaceae and Eupomatiaceae, but older than the generally accepted MRCA of Annonaceae as represented by the stem lineage of *Anaxagorea* (Fig. 1, node B). This latter node has been estimated at 57 or 69 mya (Doyle et al., 2004) and 65 or 77 mya (Richardson et al., 2004), based on molecular dating with different fossil calibrations. Similar analyses were performed in this study, primarily to estimate the rates of evolution of the two copies of *trnL-F* (see below). When using the *trnL-F* gene phylogeny calibrated using the fossil *Archaeanthus*, these

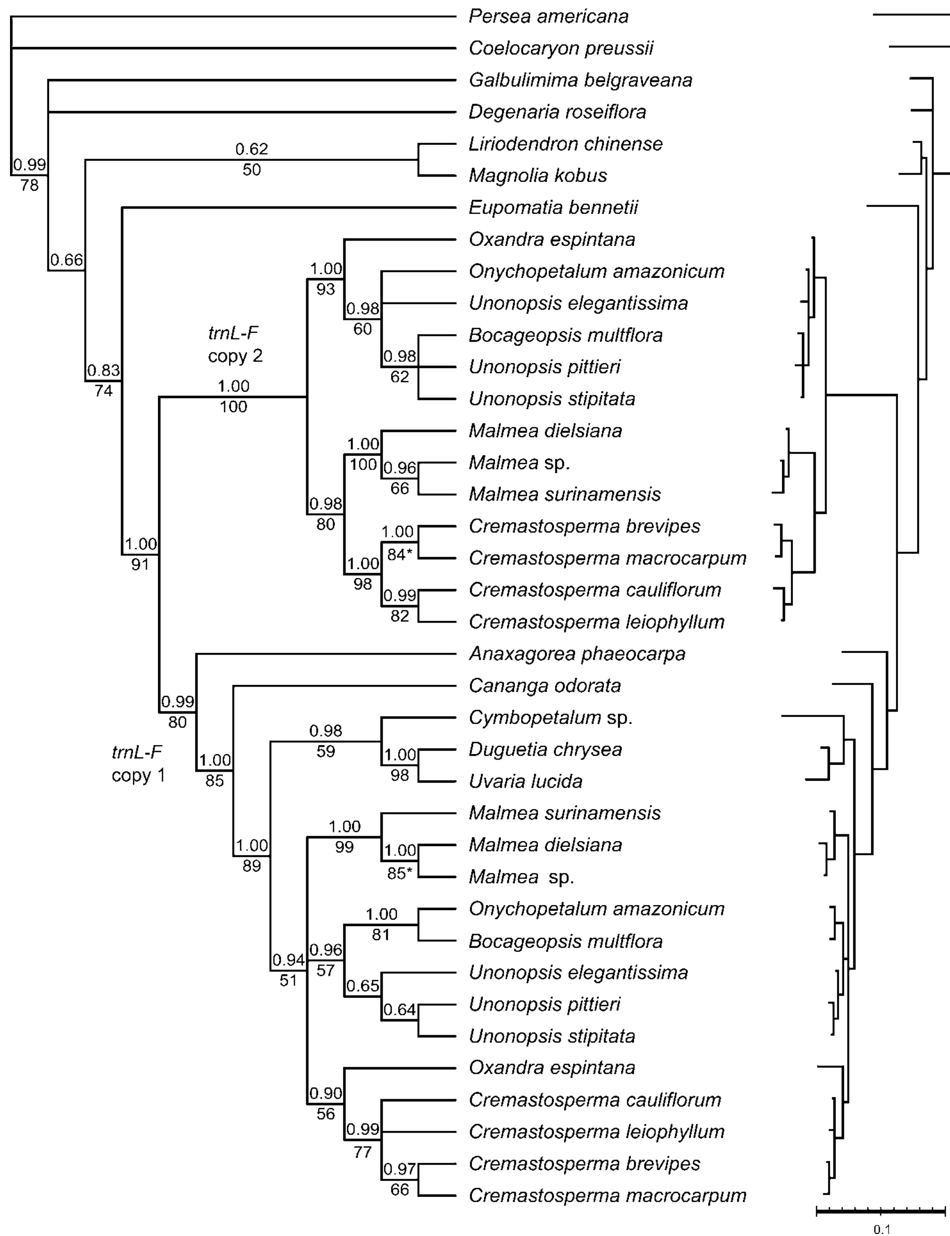


Fig. 5. Bayesian 50% majority rule consensus tree and phylogram resulting from analysis of partial sequences of *trnL-F* copies 1 and 2. Posterior probabilities are indicated above the nodes, maximum parsimony bootstrap support values (where >50%) appear below the nodes. Bootstrap support values for conflicting nodes referred to in the text are labeled with an asterisk (*).

resulted in a minimum age of 81 mya (SD = 7) for the divergence of the two *trnL-F* copies (Fig. 1, node A).

The age of the South American-centered clade crown group was also estimated, using different partitions of the data. In the *trnL-F* gene phylogeny as above, the MRCA of the South American-centered clade is represented by two different nodes: (1) the MRCA of all included Annonaceae *trnL-F* copy 1 sequences and (2) the MRCA of the corresponding *trnL-F* copy 2 sequences. The age estimations for these nodes were not significantly different (Table 1). Age estimations derived from different data partitions (i.e., *trnL-F* copy 1 alone compared with the combined matrix) differed significantly. This may be due to the generally low level of taxon sampling, combined with the differences in numbers of taxa sampled resulting from exclusion of "taxa" represented by second copies of *trnL-F*. However, for each data partition, inclusion of paralogous sequences resulted in older age estimations (Table 1).

Phylogenetic utility of Annonaceae trnL-F copies—Duplicated genes have proved useful in rooting phylogenetic analyses, in particular those in which there are no extant outgroups (an extreme example being that of the tree of life) or where outgroups are too distant (such as in angiosperms; Donoghue and Mathews, 1998; Mathews and Donoghue,

TABLE 1. r8s analyses (A–C). Ages and evolutionary rates of defined nodes, estimated from different partitions of the sequence data using the penalized likelihood method. Where paralogous sequences have been analyzed, and/or the topology constrained (to conform to that recovered when homologous sequences are analyzed), this is indicated. The units for ages are millions of years, those for rates are changes per site per million years. Standard deviations for the estimates (where available) are presented between parentheses.

Analysis	Data on which estimate is based	Paralogy or topological constraint	Age: divergence of *trnL-F* paralogues	Rate: divergence of *trnL-F* paralogues
A) *trnL-F* gene phylogeny	*trnL-F* copies 1 and 2	None	81 (7)	0.000612 (0.000085)
			Age: MRCA of South American-centered clade	Rate: MRCA of South American-centered clade
B) *trnL-F* gene phylogeny	*trnL-F* copy 1	None	34 (8)	0.000461 (0.000106)
	trnL-F copy 2	None	32 (6)	0.000744 (0.00014)
trnL-F copy 1 phylogeny	*trnL-F* copy 1	None	29	
	trnL-F copy 1/*Unonopsis trnL-F* copy 2	Paralogy–topology constrained	55	
Combined matrix	*trnL-F* copy 1, *rbcL*, *matK*, *psbA-trnH*	None	23 (3)	0.000511 (0.000061)
	trnL-F copy 1/*Unonopsis trnL-F* copy 2, *rbcL*, *matK*, *psbA-trnH*	Paralogy	29 (4)	0.000511 (0.000053)
			Age: MRCA of the *Unonopsis* clade	Rate: MRCA of the *Unonopsis* clade
C) Combined matrix	*trnL-F* copy 1, *rbcL*, *matK*, *psbA-trnH*	None	10 (2)	0.000478 (0.000052)
	trnL-F copy 1/*Unonopsis trnL-F* copy 2, *rbcL*, *matK*, *psbA-trnH*	Paralogy	21 (3)	0.000501 (0.000063)

1999). Character polarization in Annonaceae is uncontroversial: *Anaxagorea* is supported as the first branching lineage in cladistic analyses of both morphological and of molecular data (Doyle and Le Thomas, 1996; Doyle et al., 2000; Sauquet et al., 2003). However, if this rooting of the Annonaceae were incorrect, this might have explained the increased rate of evolution of cpDNA markers apparent in the Long Branch Clade. No such conclusion can be drawn from analyses performed here: phylogenetic analysis of *trnL-F* sequences resulted in a gene tree in which the *Anaxagorea* rooting is not disrupted (see Figs. 1 and 5). Relationships between taxa as represented by the two *trnL-F* gene trees were mostly consistent, with only one conflicting node subject to PP > 95% (relationships between species of *Malmea*) and none subject to BS > 75% (Fig. 5). Separate analysis of *rbcL*, *matK*, and *psbA-trnH* sequences revealed no supported conflict. These three partitions were therefore included in a combined analysis, which resulted in a topology consistent with those of the *trnL-F* copies, with two apparent exceptions (Fig. 6A, compare with Fig. 5): (a) relationships between species of *Malmea*, which conflicted with those revealed by *trnL-F* copy 1 and (b) those between species of *Cremastosperma*, which conflicted with those revealed by *trnL-F* copy 2 (BS values of corresponding nodes are in boldface type and are labeled with asterisks in Figs. 5 and 6).

Combined analysis of all five partitions resulted in generally higher support for a more-resolved topology (Fig. 6B), with the exception of lower support for relationships between species of *Malmea*.

Proportions of variable and parsimony-informative characters and CI/RI (based on the topology derived from combined analysis) are presented in Table 2. Of the five markers compared, the roughly 600-bp long fragment of *trnL-F* copy 2, amplified using the copy-specific primers, provided the highest proportion of parsimony informative characters, and more characters in total than *rbcL*, which is more than twice as long and has to be amplified and sequenced in two pieces. The phylogenetic utility of *trnL-F* copy 2 would also appear clear when comparing the resolution and support values, particularly within *Cremastosperma* (Fig. 5). The limited conflict apparent in the topologies (see Figs. 5 and 6) is not consistently between one partition and the others. It may be a result of either the small numbers of informative characters involved, the very limited taxon sampling, or both.

TABLE 2. Comparison of the length, total variability, numbers and proportions of parsimony informative characters, and consistency and retention indices of markers used in the combined analyses.

Marker	Approx. length (bp)	Variable characters No.	Variable characters %	Pars. inf. char. No.	Pars. inf. char. %	*CI*	*RI*
(Partial) *trnL-F* copy 1	600	74	12.3	27	4.5	0.94	0.88
(Partial) *trnL-F* copy 2	600	89	14.8	50	8.3	0.92	0.92
rbcL	1500	122	8.1	44	2.9	0.83	0.70
matK	850	136	16	60	7.1	0.95	0.92
psbA-trnH	430	61	14.2	10	2.3	0.93	0.79
Combined	3980	482	12.1	191	4.8	0.91	0.86

Note: Consistency index/retention index (CI/RI) calculated using single most parsimonious topology of combined analysis. Pars. inf. char. = parsimony-informative characters

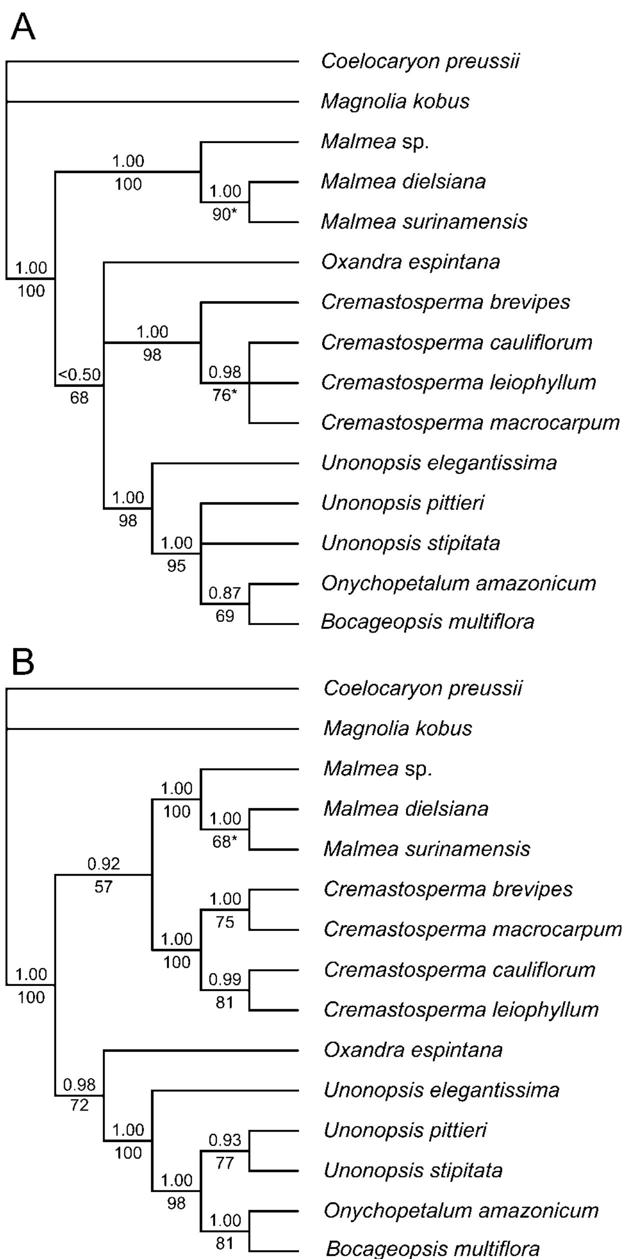


Fig. 6. Bayesian/maximum parsimony bootstrap consensus trees of (A) combined chloroplast DNA markers *rbcL*, *matK*, and *psbA-trnH* and (B) *rbcL*, *matK*, *psbA-trnH*, and *trnL-F* copies 1 and 2. Posterior probabilities are indicated above the nodes, maximum parsimony bootstrap support values (where >50%) appear below the nodes. Bootstrap support values for conflicting nodes referred to in the text are labeled with an asterisk (*).

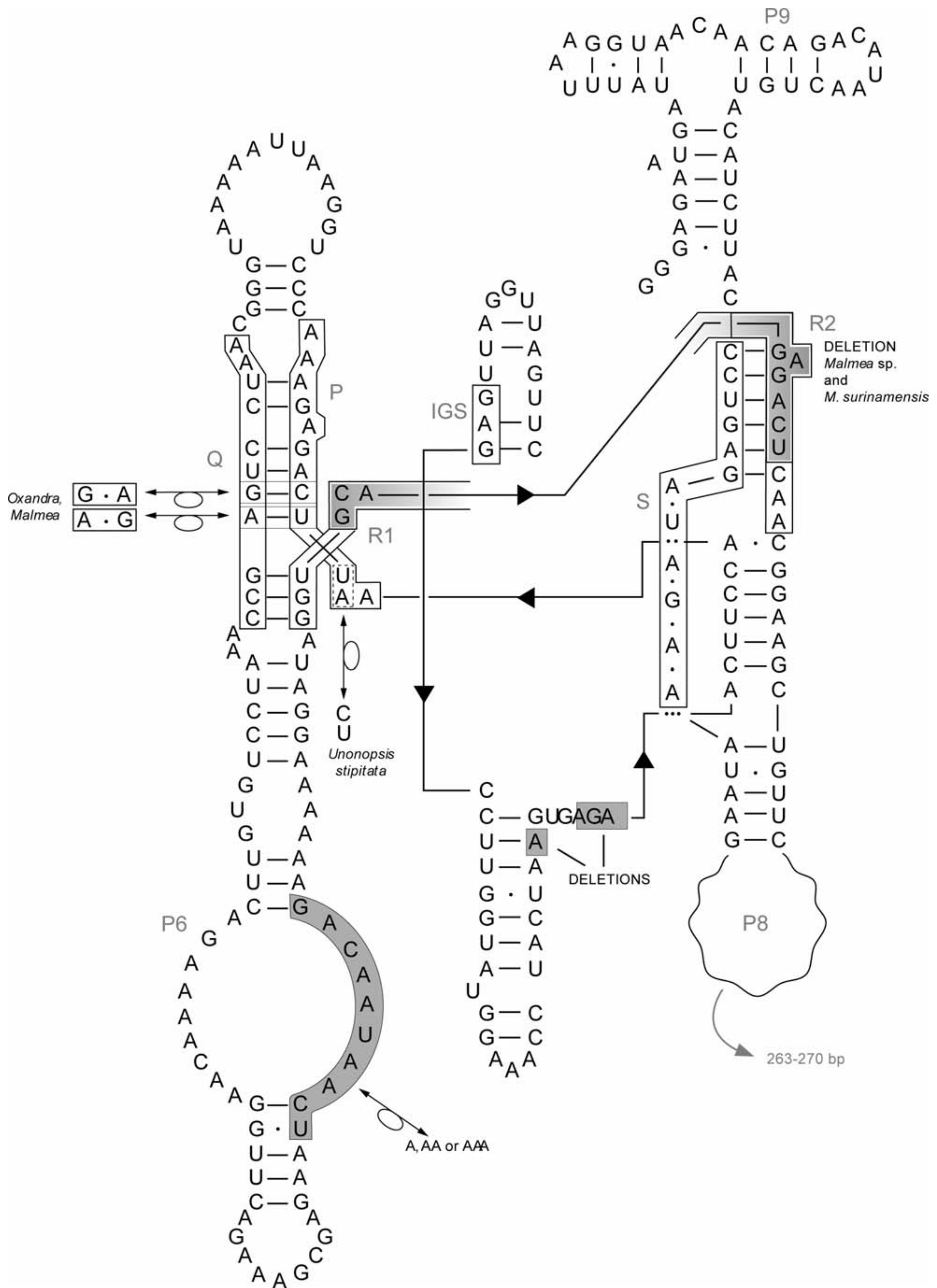
However, in comparing the two *trnL-F* paralogues with the single corresponding copy found in other plants, two further levels of homology can be considered: functional and positional. These may have implications for the patterns of evolution displayed, which might lead us to treat this new potential marker with some caution.

Functional homology of Annonaceae trnL-F copies—Examples of paralogues of chloroplast genes have been documented where function in one copy (often located in a different genomic compartment) appears to have been lost (see below). In the case of protein coding genes, loss of function can often be diagnosed by the incidence of mutations resulting in disruption of the reading frame or the appearance of stop codons. The function of the transfer RNA for which the *trnL* gene codes is related to its secondary structure and that of the intron within it. We therefore attempted to assess the functionality of copies of *trnL-F* in Annonaceae by comparing them with plant *trnL* introns and 3′ exons for which secondary structures have been proposed (by Borsch et al., 2003; Quandt and Stech, 2004; Quandt et al., 2004).

Borsch et al. (2003) demonstrated that the secondary structure of the *trnL* intron is highly conserved in basal angiosperms. Only 20% of the 95 positions corresponding to proposed stem structures were variable across their study group. The structure presented in Fig. 7 (following Cech et al. [1994], based on the secondary structure of group I introns modeled by Michel and Westhof [1990]), is that of the *trnL* copy 1 intron sequence of *Cremastosperma brevipes* (see Table 1). The conserved sequence motifs, as described by Quandt et al. (2004), and selected differences between these motifs in *C. brevipes* copy 1 and those in *trnL-F* copy 2 sequences are indicated.

The greater number of parsimony-informative characters in *trnL-F* copy 2 sequences than in *trnL-F* copy 1 sequences is reflected in a higher number of changes occurring in regions inferred to represent stem structures as opposed to loops in *Nymphaea*. Eleven single-base changes are observed in *trnL-F* copy 2 putative stem regions, in comparison to five in the same taxa for *trnL-F* copy 1 (excluding the P6 and P8 regions, for which secondary structures in both copies appeared to differ significantly from that of *Nymphaea*). No compensatory changes were apparent in sequences of either copies. The few base changes apparent in the 3′ exon relative to that of *Nymphaea* are limited to within loop structures in both copies. More significantly, a 9-bp long deletion is observed in the *trnL-F* copy 2 sequences of two of the *Malmea* accessions (representing a putative synapomorphy for a clade also supported by *rbcL*, *psbA-trnH*, and *matK* data but apparently contradicted by *trnL-F* copy 1). This deletion includes most of the conserved R1 and R2 sequence motif (as described by Quandt et al., 2004; see Fig. 6) and thus seems likely to represent a significant change in the secondary structure. This suggests that copy 1, rather than copy 2, is the functional homologue of *trnL-F* in other plants.

Rate of evolution of Annonaceae trnL-F copies—The role of selection can be investigated by estimating relative and absolute rates of sequence divergence for different branches in the gene family tree. If mutation rate is constant, differences in divergence rates represent strength of selection (Thornton and DeSalle, 2000). Under the best-fitting substitution model (TIM + G), the likelihood of the partial *trnL-F* data given one of the MP topologies was 2022.27. Enforcing a molecular clock resulted in a significantly different likelihood of 2074.00 ($P < 0.01$, chi-squared test, 31 degrees of freedom), and the clock hypothesis was thus rejected. We therefore used the penalized likelihood method of Sanderson (2002a) to estimate rates of evolution in different branches of the *trnL-F* gene tree, applying a bootstrapping technique to assess error according to character sampling. The rate at the South American-centered clade crown node was estimated to be 0.000744 (SD =


P9
P
Q
IGS
R2
DELETION
Malmea sp.
and
M. surinamensis
S
R1
Oxandra,
Malmea
Unonopsis
stipitata
DELETIONS
P6
P8
263-270 bp
A, AA or AAA

0.000146) changes per position per million years for *trnL-F* copy 2, significantly higher than that estimated for *trnL-F* copy 1: 0.000461 (SD = 0.000106).

If we were comparing two chloroplast-encoded markers, the difference observed might thus be interpreted to suggest less stringent selection acting on *trnL-F* copy 2. However, although one of the two copies of *trnL-F* in Annonaceae is presumably located in the expected position in the chloroplast genome, the position of the other is unknown. Relative rate differences can be attributable to other evolutionary or population genetic phenomena (Small et al., 1998), some of which, such as background mutational processes, drift, and rates of recombination, differ across genomic compartments (Wolfe and Randle, 2004). The positional homology of both copies is thus critical to interpreting this higher rate of change.

Positional homology in Annonaceae trnL-F copies—There are numerous examples of duplicated chloroplast genes in the literature, the position of which has been demonstrated in many cases. Partial or complete pseudogenes of *trnF* have been observed as insertions in the *trnL-F* spacer in *Microseris* (Asteraceae; Vijverberg and Bachmann, 1999) and particular lineages of Brassicaceae (Dobes et al., 2004; Koch et al., 2005). These examples, however, represent relatively small fragments compared to that reported here. Ayliffe et al. (1998) and Millen et al. (2001) identified sequences of plastid homology in the nuclear genome of various angiosperms, demonstrating transfers of *inf*A from the chloroplast to the nucleus. Gene content of the mitochondrial genome is considered particularly dynamic and flexible (Nakazono and Hirai, 1993). Cummings et al. (2003) reported hundreds of successful transfers of *rbcL* from the chloroplast to the mitochondrion in flowering plants: of the five transferred sequences examined, all had disrupted reading frames. Nakazono and Hirai (1993) discovered nine intact and three defunct chloroplast tRNA genes in the rice mitochondrion, including the 3′ *trnL* exon and *trnF*. With the publication of the complete rice mitochondrial genome, Notsu et al. (2002) discovered a total of 17 tRNA genes and five pseudo-tRNA sequences of chloroplast origin. They additionally identified nuclear sequences of chloroplast origin positioned adjacent to sequences of mitochondrial origin, suggesting that secondary transfers occurred from the mitochondrion.

Positional homology is critical for the purposes of phylogenetic reconstruction, as it largely determines the mode of inheritance. Although the general rule for both chloroplast and mitochondrial genomes in angiosperms is maternal transmission (Corriveau and Coleman, 1988), the occasional occurrence of biparental inheritance can be expected for many, if not all, species (Milligan, 1992). For example, paternal inheritance has been reported for the chloroplasts of *Actinidia* Lindl. (Testolin and Cipriani, 1997) and for the mitochondria of *Brassica napus* L. (Erickson and Kemble, 1990) and *Musa acuminata* Colla (Fauré et al., 1994). Therefore, although mitochondrial copies of chloroplast genes would be more likely to contain phylogenetic signal congruent with that of genuine chloroplast markers than chloroplast genes that have been transferred to the biparentally inherited nucleus, in both cases hybridization events could result in conflicting gene trees.

Conclusions—Two copies of the widely used chloroplast marker *trnL-F* are found in Annonaceae. Either copy can be discovered using standard experimental techniques, and at first sight, there is no obvious way to distinguish one from the other. However, when they are incorrectly assumed to be homologous and aligned together as characters in phylogenetic analysis, the resulting phylogeny is both misleading and very different to that supported by other data.

The occasional incidences of indels and substitutions within the structural regions of the *trnL-F* copy 2 intron suggest that copy 1 is the functional homologue of *trnL-F* sequences found in other plants. Copy 1 is therefore also most likely to be the positional homologue of *trnL-F* sequences found in other plants. Copy 2 sequences appear to be subject to an increased rate of evolution, which may reflect relaxed selectional constraint and/or a higher ambient rate of change. If the ambient rate of change is higher, the most likely position for copy 2 would be in the nucleus rather than the chloroplast or the more slowly evolving mitochondrion. This conclusion implies copy 2 is biparentally inherited, rather than maternally inherited as is likely of copy 1. This means that, even if the two copies are correctly identified as separately evolving units, each could have a distinct evolutionary history. The data presented here do not conclusively show conflicting signal between *trnL-F* copy 2 and chloroplast markers. Greater sampling of taxa for this marker may yet reveal such conflict, but given this caveat, its high variability relative to chloroplast markers may make it a useful tool in phylogenetic reconstruction in Annonaceae.

The example presented here confirms that data sets dominated by characters from one marker should be interpreted cautiously. Unnoticed paralogy may not be a problem limited to the increasingly rare phylogenetic studies applying only single markers. Because of their known lack of recombination, chloroplast sequences are often assumed a priori to contain complementary phylogenetic signal, and increased resolution resulting from their inclusion in combined analyses is interpreted as support for this assumption a posteriori. In Annonaceae, single chloroplast markers provide insufficient total numbers of informative characters to arrive at fully resolved topologies and thus do not reveal all possible conflicting nodes. This low variability is not uncommon, particularly at lower taxonomic levels, and can also pose problems for the application of alternative methods to assess conflict between data partitions, such as the incongruence length difference test (Reeves et al., 2001; Yoder et al., 2001; Darlu and Lecointre, 2002).

In recent years, increasing emphasis has been placed on the transfer of genes between genomic compartments within the same organisms (Ayliffe et al., 1998; Millen et al., 2001; Cummings et al., 2003) and even horizontal transfer of genes between organisms (e.g., Bergthorsson et al., 2003). One of the results of such processes is that a proportion of DNA fragments amplified using PCR are in fact paralogues of the marker in

←

Fig. 7. Secondary structure of the *trnL* copy 1 intron sequence of *Cremastosperma brevipes*. Conserved sequence motifs are outlined and labeled following Cech et al. (1994). Selected differences between this representative *trnL-F* copy 1 sequence and specified *trnL-F* copy 2 sequences are outlined and labeled, with deletions shaded in gray.

question. Investigation of these cases can give further insight into the processes of evolution. However, unwittingly combining markers with conflicting signals in phylogenetic analyses violates the assumptions behind the method. This may affect both the topology reconstructed and the branch lengths optimized onto that topology. This in turn may have an impact on estimations of ages and of rates of molecular change. The results presented here indicate that this might occur more often than we assume or recognize.

LITERATURE CITED

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. ZHANG, AND Z. ZHANG. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389–3402.

APG II. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141: 399–436.

ARABIDOPSIS GENOME INITIATIVE. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.

AYLIFFE, M. A., N. S. SCOTT, AND J. N. TIMMIS. 1998. Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants. *Molecular Biology and Evolution* 15: 738–745.

BAKKER, F. T., A. CULHAM, R. GOMEZ-MARTINEZ, J. CARVALHO, J. COMPTON, R. DAWTREY, AND M. GIBBY. 2000. Patterns of nucleotide substitution in angiosperm cpDNA *trnL* (UAA)-*trnF* (GAA) regions. *Molecular Biology and Evolution* 17: 1146–1155.

BERGTHORSSON, U., K. L. ADAMS, B. THOMASON, AND J. D. PALMER. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424: 197–201.

BESENDAHL, A., Y.-L. QIU, J. LEE, J. D. PALMER, AND D. BHATTACHARYA. 2000. The cyanobacterial origin and vertical transmission of the plastid tRNALEU group-I intron. *Current Genetics* 37: 12–23.

BORSCH, T., K. W. HILU, D. QUANDT, V. WILDE, C. NEINHUIS, AND W. BARTHLOTT. 2003. Noncoding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* 16: 558–576.

CECH, T. R., S. H. DAMBERGER, AND R. GUTELL. 1994. Representation of the secondary and tertiary structure of group I introns. *Nature Structural Biology* 1: 273–280.

CHASE, M. W., AND A. V. COX. 1998. Gene sequences, collaboration and analysis of large data sets. *Australian Systematic Botany* 11: 215–229.

CORRIVEAU, J. L., AND A. W. COLEMAN. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany* 75: 1443–1458.

CUMMINGS, M. P., J. M. NUGENT, R. G. OLMSTEAD, AND J. D. PALMER. 2003. Phylogenetic analysis reveals five independent transfers of the chloroplast gene *rbcL* to the mitochondrial genome in angiosperms. *Current Genetics* 43: 131–138.

DARLU, P., AND G. LECOINTRE. 2002. When does the incongruence length difference test fail? *Molecular Biology and Evolution* 19: 432–437.

DILCHER, D. L., AND P. R. CRANE. 1984. *Archaeanthus*: an early angiosperm from the Cenomanian of the Western Interior of North America. *Annals of the Missouri Botanical Garden* 71: 351–383.

DOBES, C. H., T. MITCHELL-OLDS, AND M. A. KOCH. 2004. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii* × *divaricarpa*, and *A. holboelli* (Brassicaceae). *Molecular Ecology* 13: 349–370.

DONOGHUE, M. J., AND S. MATHEWS. 1998. Duplicate genes and the root of angiosperms, with an example using phytochrome sequences. *Molecular Phylogenetics and Evolution* 9: 489–500.

DOYLE, J. A., P. BYGRAVE, AND A. LE THOMAS. 2000. Implications of molecular data for pollen evolution in Annonaceae. *In* M. M. Harley, C. M. Morton, and S. Blackmore [eds.], Pollen and spores: morphology and biology, 259–284. Royal Botanic Garden, Kew, UK.

DOYLE, J. A., AND A. LE THOMAS. 1996. Phylogenetic analysis and character evolution in Annonaceae. *Adansonia* 18: 279–334.

DOYLE, J. A., H. SAUQUET, T. SCHARASCHKIN, AND A. LE THOMAS. 2004. Phylogeny, molecular and fossil dating, and biogeographic history of Annonaceae and Myristicaceae (Magnoliales). *International Journal of Plant Sciences* 165 (Supplement): S55–S67.

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

ERICKSON, L., AND R. KEMBLE. 1990. Paternal inheritance of mitochondria in rapeseed (*Brassica napus*). *Molecular and General Genetics* 222: 135–139.

FAURÉ, S., J. NOYER, F. CARREEL, J. HORRY, F. BAKRY, AND C. LANAUD. 1994. Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). *Current Genetics* 25: 265–269.

FITCH, W. M. 1971. Toward defining the course of evolution: minimum change for a specified tree topology. *Systematic Zoology* 20: 406–416.

FULTON, T. M., R. VAN DER HOEVEN, N. T. EANNETTA, AND S. D. TANKSLEY. 2002. Identification, analysis, and utilization of conserved otholog set markers for comparative genomics in higher plants. *Plant Cell* 14: 1457–1467.

GEYER, C. J. 1991. Markov chain Monte Carlo maximum likelihood. *In* M. Keramidas [ed.], Computing science and statistics: proceedings of the 23rd symposium on the interface, 156–163. Interface Foundation of North America, Fairfax, Virginia, USA.

HUANG, S., Y. C. CHIANG, B. A. SCHAAL, C. H. CHOU, AND T. Y. CHIANG. 2001. Organelle DNA phylogeography of *Cycas taitungensis*, a relict species in Taiwan. *Molecular Ecology* 10: 2669–2681.

HUELSENBECK, J. P. 2000. MrBayes: Bayesian inference of phylogeny. Distributed by the author. Department of Biology, University of Rochester, Rochester, New York, USA. Available from http://mrbayes.sourceforge.net/.

HUELSENBECK, J. P., F. RONQUIST, R. NIELSEN, AND J. P. BOLLBACK. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294: 2310–2314.

KOCH, M. A., C. DOBES, M. MATSCHINGER, W. BLEEKER, J. VOGEL, M. KIEFER, AND T. MITCHELL-OLDS. 2005. Evolution of the *trnF*(GAA) gene in Arabidopsis relatives and the Brassicaceae family: monophyletic origin and subsequent diversification of a plastidic pseudogene. *Molecular Biology and Evolution* 22: 1032–1043.

KOJOMA, M., K. KURIHARA, K. YAMADA, S. SEKITA, M. SATAKE, AND O. IIDA. 2002. Genetic identification of cinnamon (*Cinnamomum* spp.) based on *trnL-trnF* chloroplast DNA. *Planta Medica* 68: 94–96.

MARSHALL, H. D., C. NEWTON, AND K. RITLAND. 2001. Sequence-repeat polymorphisms exhibit the signature of recombination in lodgepole pine chloroplast DNA. *Molecular Biology and Evolution* 18: 2136–2138.

MASON-GAMER, R. J., AND E. A. KELLOGG. 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Systematic Biology* 45: 524–545.

MATHEWS, S., AND M. J. DONOGHUE. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* 286: 947–950.

MICHEL, F., AND E. WESTHOF. 1990. Modeling of the three-dimensional structure of group I catalytic introns based on comparative sequence analysis. *Journal of Molecular Biology* 216: 585–610.

MILLEN, R. S., R. G. OLMSTEAD, K. L. ADAMS, J. D. PALMER, N. T. LAO, L. HEGGIE, T. A. KAVANAGH, J. M. HIBBERD, J. C. GRAY, C. W. MORDEN, P. J. CALIE, L. S. JERMIIN, AND K. H. WOLFE. 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* 13: 645–658.

MILLIGAN, B. G. 1992. Is organelle DNA strictly maternally inherited? Power analysis of a binomial distribution. *American Journal of Botany* 79: 1325–1328.

MOLS, J. B., B. GRAVENDEEL, L. W. CHATROU, M. D. PIRIE, P. C. BYGRAVE, M. W. CHASE, AND P. J. A. KESSLER. 2004. Identifying clades in Asian Annonaceae: monophyletic genera in the polyphyletic Miliuseae. *American Journal of Botany* 91: 590–600.

NAKAZONO, M., AND A. HIRAI. 1993. Identification of the entire set of transferred chloroplast DNA sequences in the mitochondrial genome of rice. *Molecular Genetics and Genomics* 236: 341–346.

NOTSU, Y., S. MASOOD, T. NISHIKAWA, N. KUBO, N. AKIDUKI, M. NAKAZONO, A. HIRAI, AND K. KADOWAKI. 2002. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Molecular Genetics and Genomics* 268: 434–445.

PIRIE, M. D., L. W. CHATROU, R. H. J. ERKENS, J. W. MAAS, T. VAN DER NIET, J. B. MOLS, AND J. E. RICHARDSON. 2005. Phylogeny reconstruction and molecular dating in four neotropical genera of Annonaceae: the effect of taxon sampling in age estimations. *In* F. T. Bakker, L. W. Chatrou, B. Gravendeel, and P. B. Pelser [eds.], Plant species-level systematics: new perspectives on pattern and process, 149–174. Regnum Vegetabile 143. A. R. G. Gantner Verlag, Vaduz, Liechtenstein.

PIRIE, M. D., L. W. CHATROU, J. B. MOLS, R. H. J. ERKENS, AND J. OOSTERHOF. 2006. 'Andean-centred' genera in the Short Branch Clade of Annonaceae: testing biogeographic hypotheses using phylogeny reconstruction and molecular dating. *Journal of Biogeography* 33: 31–46.

POSADA, D., AND K. A. CRANDALL. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.

QUANDT, D., K. MULLER, M. STECH, J.-P. FRAHM, W. FREY, K. HILU, AND T. BORSCH. 2004. Molecular evolution of the chloroplast *trn*L-F region in land plants. *In* B. Goffinet, V. Hollowell, and R. Magill [eds.], Molecular systematics of bryophytes, 13–37. Monographs in Systematic Botany 98. Missouri Botanical Garden Press, St. Louis, Missouri, USA.

QUANDT, D., AND M. STECH. 2004. Molecular evolution of the *trn*T_{UGU}-*trn*F_{GAA} region in bryophytes. *Plant Biology* 6: 545–554.

REEVES, G., M. W. CHASE, P. GOLDBLATT, T. DE CHIES, B. LEJEUNE, M. F. FAY, A. V. COX, AND P. J. RUDALL. 2001. Molecular systematics of Iridaceae: evidence from four plastid DNA regions. *American Journal of Botany* 88: 2074–2087.

RICHARDSON, J. E., L. W. CHATROU, J. B. MOLS, R. H. J. ERKENS, AND M. D. PIRIE. 2004. Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 359: 1495–1508.

SANDERSON, M. J. 2002a. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* 19: 101–109.

SANDERSON, M. J. 2002b. r8s, version 1.50. Distributed by the author, Section of Evolution and Ecology, University of California, Davis, Davis, California, USA. Available from http://loco.biosci.arizona.edu/r8s/.

SANDERSON, M. J., AND H. B. SHAFFER. 2002. Troubleshooting molecular phylogenetic analyses. *Annual Review of Ecology and Systematics* 33: 49–72.

SAUQUET, H., J. A. DOYLE, T. SCHARASCHKIN, T. BORSCH, K. HILU, L. W. CHATROU, AND A. LE THOMAS. 2003. Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: implications for character evolution. *Botanical Journal of the Linnean Society* 142: 125–186.

SAVOLAINEN, V., P. CUÉNOUD, R. SPICHIGER, M. D. P. MARTINEZ, M. CRÈVECOEUR, AND J.-F. MANEN. 1995. The use of herbarium specimens in DNA phylogenetics: evaluation and improvement. *Plant Systematics and Evolution* 197: 87–98.

SIDDALL, M. E., AND M. F. WHITING. 1999. Long-branch abstractions. *Cladistics* 15: 9–24.

SMALL, R. L., J. A. RYBURN, R. C. CRONN, T. SEELANAN, AND J. F. WENDEL. 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany* 85: 1301–1315.

SWOFFORD, D. L. 2000. PAUP*: phylogenetic analysis using parsimony (* and other methods). Sinauer, Sunderland, Massachusetts, USA.

TABERLET, P., L. GIELLY, G. PAUTOU, AND J. BOUVET. 1991. Universal primers for amplification of the three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.

TESTOLIN, R., AND G. CIPRIANI. 1997. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in the genus *Actinidia*. *Theoretical and Applied Genetics* 94: 897–903.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876–4882.

THORNTON, J. W., AND R. DESALLE. 2000. Gene family evolution and homology: genomics meets phylogenetics. *Annual Review of Genomics and Human Genetics* 1: 41–73.

VAN HEUSDEN, E. C. H. 1992. Flowers of Annonaceae: morphology, classification, and evolution. *Blumea Supplement* 7: 1–218.

VAN SETTEN, A. K., AND J. KOEK-NOORMAN. 1992. Studies in Annonaceae. XVII. Fruits and seeds of Annonaceae: morphology and its significance for classification. *Bibliotheca Botanica* 142: 1–101.

VIJVERBERG, K., AND K. BACHMANN. 1999. Molecular evolution of a tandemly repeated *trnF* (GAA) gene in the chloroplast genomes of *Microseris* (Asteraceae) and the use of structural mutations in phylogenetic analyses. *Molecular Biology and Evolution* 16: 1329–1340.

WOLFE, A. D., AND C. P. RANDLE. 2004. Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: implications for plant molecular systematics. *Systematic Botany* 29: 1011–1020.

WON, H., AND S. S. RENNER. 2005. The chloroplast trnT-trnF region in the seed plant lineage Gnetales. *Journal of Molecular Evolution* 61: 425–436.

YODER, A. D., J. A. IRWI, AND B. A. PAYSEUR. 2001. Failure of the ILD test to determine data combinability for slow loris phylogeny. *Systematic Biology* 50: 408–424.

ZUKER, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31: 3406–3415.

APPENDIX. Voucher information and GenBank accession numbers for taxa used in this study. Original sequences were generated for this study, and published sequences were taken from GenBank and Mols et al. (2004), Sauquet et al. (2003), Pirie et al. (2005, 2006), and Kojoma et al. (2002). A dash indicates the region was not sampled. Voucher specimens were deposited in the following herbaria: U = National Herbarium of the Netherlands, Universiteit Utrecht branch; WAG = National Herbarium of the Netherlands, Wageningen Universiteit branch; SUVA = University of the South Pacific, Fiji; NCU = University of North Carolina.

***Taxon*—**GenBank accession nos.: *trnL-trnF* copy 1, *trnL-trnF* copy 2, *rbcL, matK, psbA-trnH*; Source; *Voucher specimen*.

Alphonsea boniana Finet & Gagnep.—AY319077, —, —, —, —; Vietnam; *Kessler, P.J.A. 3116 (L).*

Anaxagorea phaeocarpa Mart.—AY231284, AY238944, —, —, —; Ecuador, Napo; *Maas, P.J.M. et al. 8592 (U).* ***Annickia chlorantha*** (Oliv.) Setten & Maas—AY841671, —, —, —, —; Gabon; *Sosef, M.S. M. 1877 (WAG).* ***Bocageopsis multiflora*** (Mart.) R.E.Fr.—AY841678, DQ018199, AY841600, DQ018262, AY841445; Guyana; *Jansen-Jacobs, M.J. et al. 5789 (U).* ***Cananga odorata*** (Lam.) Hook.f & Thomson—AY841680, —, —, —, —; Costa Rica, Limón; *Chatrou, L.W. et al. 93 (U).* ***Cinnamomum cassia*** Blume—AB054241, AB054233, —, —, —; origin unknown; *Izu experimental station for medicinal plants 21.* ***Coelocaryon preussii*** Warb.—AY743456, —, AY743437, AY743475, —; Gabon; *Wieringa, J.J. et al. 3640 (WAG).* ***Cremastosperma brevipes*** (DC.) R.E.Fr.—AY743573, DQ018191, AY743527, AY743550, AY841447; French Guiana; *Scharf, U. 76 (U).* ***Cremastosperma cauliflorum*** R.E.Fr.—AY743565, DQ018192,

AY743519, AY743542, AY841448; Peru, Loreto; *Chatrou, L.W. et al. 224 (U)*. ***Cremastosperma leiophyllum*** R.E.Fr.—AY743569, DQ018193, AY743523, AY743546, AY841449; Bolivia, Santa Cruz; *Pirie, M.D. et al. 2 (U)*. ***Cremastosperma macrocarpum*** Maas—AY743574, DQ018194, AY743528, AY743551, AY841450; Venezuela, Falcón; *Wingfield, R. 6751 (U)*. ***Cymbopetalum* sp.**—AY841537, —, —, —, —; *Costa Rica*; Chatrou, L.W. et al. 44 (U). ***Degeneria roseiflora*** J.M. Milller—AY220414, AY220361, —, —, —; origin unknown; *J.M. Miller 1189 (SUVA)*. ***Duguetia chrysea*** Maas—AY841691, —, —, —, —; *Brazil – Amazonas*; *Maas, P.J.M. et al. 8052 (U)*. ***Eupomatia bennettii*** F.Muell.—Prov. 50/51 —, —, —, —; origin unknown; *voucher unknown*. ***Galbulimima belgraveana*** (F. Muell.) Sprague—AY220415, —, —, —, —; origin unknown; *Qiu, Y.-L. 90034 (NCU)*. ***Greenwayodendron oliveri*** (Engl.) Verdc.—AY743470, —, —, —, —; Ghana; *Jongkind, C.C.H. et al. 1795 (WAG)*. ***Klarobelia cauliflora*** Chatrou—AY841705, —, —, —, —; Peru, Loreto; *Chatrou, L.W. et al. 161 (U)*. ***Letestudoxa bella*** Pellegr.—AY841707, —, —, —, —; Gabon; *Wieringa, J.J. & T. Nzabi 2797 (WAG)*. ***Liriodendron chinense*** Sargent—AY841670, —, —, —, —; China[a]; *Chatrou, L.W. et al. 279 (U)*. ***Magnolia kobus*** DC.—AY743457, —, AY743438, AY743476, AY841425; Japan[a]; *Chatrou, L.W. et al. 278 (U)*. ***Malmea dielsiana*** R.E.Fr.—AY231288 AY238948, DQ018195, AY238955, AY238964, AY841473; Peru – Madre de Dios; *Chatrou, L.W. et al. 122 (U)*. ***Malmea* sp.**—AY841541, DQ018196, AY841527, AY841397, AY841475; Peru, Loreto; *Chatrou, L.W. et al. 8 (U)*. ***Malmea surinamensis*** Chatrou—AY743472, DQ018197, AY743453, AY743491, AY841476; Suriname; *Jansen-Jacobs, M.J. et al. 6207 (U)*. ***Mezzettia parviflora*** Becc.—AY319095, —, —, —, —; Indonesia; *Okada 3388 (L)*. ***Monanthotaxis whytei*** (Stapf) Verdc.—AY841713, —, —, —, —; Tropical Africa[a] *Chatrou, L.W. 475 (U)*. ***Monocarpia euneura*** Miq.—AY319111, —, —, —, —; Indonesia; *Slik, F. 2931 (L)*. ***Mosannona costaricensis*** R.E.Fr.—AY743496, —, —, —, —; Costa Rica, Limón; *Chatrou, L.W. et al. 90 (U)*. ***Onychopetalum amazonicum*** R.E.Fr.—DQ018175, DQ018198, DQ018222, DQ018261, DQ018237; Brazil, Para; *Sperling, C.R. et al. 5925*. ***Oxandra asbeckii*** (Pulle) R.E.Fr.—AY841717, —, —, —, —; Guyana; *UG-NB-55 (U)*. ***Oxandra espintana*** (Spruce ex Benth.) Baill.—AY319180, DQ018189, AY319066, DQ018260, AY841487; Peru, Madre de Dios; *Chatrou, L.W. et al. 133 (U)*. ***Persea americana*** Mill. cv. accnum—AY841669, —, —, —, —; neotropical[a]; *Chatrou, L.W. 479 (U)*. ***Piptostigma mortehani*** De Wild.—AY743473, —, —, —, —; Gabon; *Wieringa, J. J. et al. 2779 (WAG)*. ***Polyalthia glauca*** (Hassk.) Boerl.—AY319137, —, —, —, —; Indonesia; *Mols, J.B. 20 (L)*. ***Polyalthia suberosa*** (Roxb.) Thwait.—AY231289, AY238949, —, —, —; India[a]; *Chatrou, L.W. 480 (U)*. ***Sapranthus viridiflorus*** G.E.Schatz—AY319165, —, —, —, —; Costa Rica, La Selva; *Chatrou, L.W. et al. 55 (U)*. ***Trigynaea lanceipetala*** D.M.Johnson & N.A.Murray—AY743468, —, —, —, —; Peru, Loreto; *Chatrou, L.W. et al. 234 (U)*. ***Unonopsis elegantissima*** R.E.Fr.—DQ018176, DQ018200, DQ018223, DQ018263, DQ018238; Peru, Loreto; *Chatrou, L.W. et al. 250 (U)*. ***Unonopsis pittieri*** Saff.—AY841739, DQ018201, AY841661, DQ018264, AY841517; Costa Rica, Braulio Carillo; *Chatrou, L.W. et al. 68 (U)*. ***Unonopsis stipitata*** Diels—AY841740, DQ018202, AY841662, AY841400, AY841519; Peru, Loreto; Chatrou, L.W. et al. 253 (U). ***Uvaria lucida*** Benth. **subsp. *virens*** (N.E.Br.) Verdc.—AY231290, AY238950, —, —, —; tropical West Africa; *Botanische Tuinen 84GR00334 (U)*.

[a] Cultivated in Utrecht University Botanic Garden.