Northern Illinois University

Huskie Commons

Faculty Peer-Reviewed Publications | Faculty Research, Artistry, & Scholarship

1-1-2008

# Phylogenetic Analyses of Two Mitochondrial Metabolic Genes Sampled in Parallel from Angiosperms Find Fundamental Interlocus Incongruence

Melvin R. Duvall

J.W. Robinson

J.G. Mattson

A. Moore

Follow this and additional works at: https://huskiecommons.lib.niu.edu/allfaculty-peerpub

## Original Citation

Duvall, M., J. W. Robinson, J. G. Mattson and A. Moore. 2008. Phylogenetic analyses of two mitochondrial metabolic genes sampled in parallel from angiosperms find fundamental interlocus incongruence. American Journal of Botany 95: 871-884.

This Article is brought to you for free and open access by the Faculty Research, Artistry, & Scholarship at Huskie Commons. It has been accepted for inclusion in Faculty Peer-Reviewed Publications by an authorized administrator of Huskie Commons. For more information, please contact jschumacher@niu.edu.

American Journal of Botany 95(7): 871–884. 2008.

# Phylogenetic analyses of two mitochondrial metabolic genes sampled in parallel from angiosperms find fundamental interlocus incongruence[1]

Melvin R. Duvall,[2] Jace W. Robinson, Jeremy G. Mattson, and Anni Moore

Biological Sciences, Northern Illinois University, DeKalb, Illinois USA 60115-2861

Plant molecular phylogeneticists have supported an analytical approach of combining loci from different genomes, but the combination of mitochondrial sequences with chloroplast and nuclear sequences is potentially problematic. Low substitution rates in mitochondrial genes should decrease saturation, which is especially useful for the study of deep divergences. However, individual mitochondrial loci are insufficiently informative, so that combining congruent loci is necessary. For this study *atp1* and *cox1* were selected, which are of similar lengths, encode components of the respiratory pathway, and generally lack introns. Thus, these genes might be expected to have similar functional constraints, selection pressures, and evolutionary histories. Strictly parallel sampling of 52 species was achieved as well as six additional composite terminals with representatives from the major angiosperm clades. However, analyses of the separate loci produced strongly incongruent topologies. The source of the incongruence was investigated by validating sequences with questionable affinities, excluding RNA-edited nucleotides, deleting taxa with unexpected phylogenetic associations, and comparing different phylogenetic methods. However, even after potential artifacts were addressed and sites and taxa putatively associated with conflict were excluded, the resulting gene trees for the two mitochondrial loci were still substantially incongruent by all measures examined. Therefore, combining these loci in phylogenetic analysis may be counterproductive to the goal of fully resolving the angiosperm phylogeny.

**Key words:** angiosperms; *atp1*; *cox1*; incongruence; mitochondrion; phylogenetics.

Molecular phylogenetic relationships in angiosperms are confidently inferred when independent lines of evidence support the same phylogenetic topologies. Molecular characters may be drawn from the chloroplast, mitochondrion, and nucleus in what have sometimes been called trigenomic analyses (summarized in Savolainen and Chase, 2003). However, the trigenomic approach has not retrieved a fully resolved, well-supported phylogeny—especially for early nodes such as those defining the associations between magnoliids, Chloranthaceae, Ceratophyllaceae, eudicots, and monocots, which largely remain unresolved or weakly supported in these studies (Parkinson et al., 1999; Soltis et al., 1999, 2000; Duvall, 2000; Zanis et al., 2003; Duvall et al., 2006). Continued failure to resolve early nodes after even more loci from the three genomes were added suggested that the underlying problem is not simply one of insufficient phylogenetic information. Phylogenetic efforts are likely stymied by a synergy between several factors, possibly including selection, functional constraints, rapid radiation, and interlocus historical conflicts.

Single-gene phylogenies inferred from loci drawn from different genomes may be studied by comparative separate and combined analyses. For example, Duvall and Ervin (2004) compared the widely used 18S rDNA locus against plastid and mitochondrial loci. Substantial conflicting phylogenetic signal was found, particularly regarding the monophyly and placement of the monocots. Residual effects of lineage sorting among 18S loci were hypothesized to be responsible for the inability to resolve the monocots and to prevent robust phylogenetic placement of the monocots and Ceratophyllaceae. This conflict also may have lowered support at other deep nodes placing Chloranthaceae, eudicots, and magnoliids in the angiosperm phylogeny. As expected, extensive further sampling did not solve the problem. In a nine-gene study including 18S sequences of 100 species, Qiu et al. (2005) found that monocot monophyly was not supported by the nuclear data. Conflicts involving the 26S nuclear ribosomal gene were also observed. Thus, while phylogenetic studies of nuclear ribosomal loci have had utility in broad phylogenetic research, combining these sequences with others in multigene studies of deep angiosperm phylogenetics has not been particularly productive for resolving specific deep nodes. Duvall and Ervin (2004) demonstrated that the *PHYC* locus was a more useful nuclear marker that had greater congruence with gene trees from other genomes for the sampled taxa. However, substituting *PHYC* for 18S sequences in trigenomic data sets did not by itself resolve the deep angiosperm phylogeny (Duvall et al., 2006).

Mitochondrial genes, with very low substitution rates (Wolfe et al., 1987), are potentially useful tools for exploring deep evolutionary divergences among angiosperms and other green plants. For example, Parkinson et al. (1999) elected to emphasize mitochondrial genes for a phylogenetic study of early diverging angiopsperms. Included among these genes were mitochondrial small subunit (SSU) rDNA, *cox1* and *rps2*. Bowe et al. (2000) used mitochondrial *atp1* and *cox1* among other genes to explore the relationships among seed plants. Hajibabaei et al. (2006) compared these same mitochondrial gene phylogenies with those of chloroplast and nuclear genes in phylogenetic studies of gnetophytes and other gymnosperms. Qiu et al. (2005) included four mitochondrial genes, *atp1*, *matR*, mtSSU, and mitochondrial large subunit (mtLSU) in a

[1] Manuscript received 28 September 2007; revision accepted 14 May 2008.

The authors thank Y.-L. Qiu, University of Michigan, Ann Arbor and D. Les, University of Connecticut, Storrs for donating DNA extracts. The authors thank A. Hackman and T. Russell, Northern Illinois University, for technical assistance and Mark Simmons and three anonymous reviewers for valuable comments on the paper. This work was supported in part by the Plant Molecular Biology Center, Northern Illinois University.

[2] Author for correspondence (e-mail: t80mrd1@wpo.cso.niu.edu), phone: 815-753-7806, fax: 815-753-7855

doi:10.3732/ajb.2007310

study of basal angiosperms. Davis et al. (2004) analyzed mitochondrial *atp1* and *cob* sequences among monocots for comparisons against chloroplast gene trees. In these and other studies, relatively low signals in individual mitochondrial genes necessitated the combination of two or more with the resulting potential for intragenomic conflict, similar to that between 18S and *PHYC*.

Recent di- and trigenomic studies (Duvall et al., 2006; Davis et al., 2004) resolved the monocots to be the sister group to the magnoliids, a clade comprised of Canellales, Piperales, Magnoliales, and Laurales (i.e., the "eumagnoliids" of Qiu et al., 2000). Notably, these studies included *atp1* sequences from the mitochondrial genome, in combination with other sequences (*rbcL*, *ndhF* and *PHYC* in Duvall et al., 2006; *rbcL* in Davis et al., 2004). A troubling aspect of these results is that the phylogenetic signal in *atp*1 has a disproportionate impact in combined analyses on the resolution of the deep topology, where there is less phylogenetic information contributed by the other genes, possibly because of saturation. Duvall et al. (2006) compared four-gene analyses with those excluding the mitochondrial sequences. In the latter analyses, the monocots, eudicots, magnoliids and Chloranthales plus Ceratophyllaceae were placed at different, and even more weakly supported positions.

One difficulty in comparing individual gene trees for deep angiosperm evolution is that the phylogenetic information in each of the individual loci, particularly those of the slowly evolving plant mitochondrion, is insufficient for robust resolution. Removing *atp1* from the four-gene analysis not only removed a source of potential intergenomic conflict, but also reduced phylogenetic information at some nodes. In particular, sequences from a second mitochondrial locus with a similarly low level of saturation would better augment the mitochondrial signal in trigenomic analyses.

We thus sought a mitochondrial locus similar to *atp1* among the other most commonly banked angiosperm loci: *matR*, *cox1*, mtSSU, and mtLSU. Of these, the gene product that is most similar in function to *atp1* is *cox1*. Unlike the mitochondrial ribosomal RNAs or maturase R, *cox1* is both protein-coding and it encodes subunit 1 of cytochrome *c* oxidase, a membrane-embedded component of the respiratory pathway, which thus might be expected to have functional constraints, selection pressures, and molecular evolutionary parameters similar to that of *atp1*. The *cox1* sequences that were already available were primarily from the work of Parkinson et al. (1999) and Barkman et al. (2000).

In this paper, we analyzed these two mitochondrial loci after sampling representatives across the angiosperms. Phylogenetic analyses were performed using parsimony, likelihood, and Bayesian methods. Factors that might be expected to affect phylogenetic congruence, including erroneous sequences, sites subject to RNA editing, long-branch attraction (LBA; Felsenstein, 1978), problematic taxa and type of phylogenetic method were tested. Analyses on different subsets of the complete data matrix were compared to assess the impact on phylogenetic congruence as measured under both parsimony and likelihood.

## MATERIALS AND METHODS

***Sampling and DNA extraction***—Taxa were sampled from the early grade of angiosperms, Ceratophyllaceae, Chloranthaceae, eudicots, magnoliids, and monocots. Early diverging families were emphasized in the larger groups. Specimen vouchers for the 22 newly determined sequences (from 20 taxa) are listed (Appendix 1). DNA was extracted as previously described (Duvall et al., 1993) or using the DNeasy Plant Mini kits according to manufacturer's instructions (Qiagen, Valencia, California, USA).

***Sequence determination***—Sequences of *atp1* were determined largely by previously described methods (Duvall, 2000). Two new primers were designed to obtain more complete terminal sequences. The primer 5_F (5′-AATTCTCACCCAGAGCTGC-3′) corresponds to positions 116–134 in the reference sequence from *Oryza sativa* L. (GenBank accession X51422), and is five bases downstream of the beginning of the start codon. Primer 1474R (5′-GCATCTGGTTCCATCTTTCTTTC-3′) corresponds to positions 1607–1585, which is 31 bases upstream of the stop codon in the reference. These primers were selected from highly conserved regions of previously published complete coding sequences of *atp1* from three grasses and nine eudicots, which are listed with GenBank accession numbers—*Oryza sativa*, X51422 (reference sequence); *Secale cereale* L., X99020; *Zea mays* L., M16222; *Beta vulgaris* L., BA000024; *Panax ginseng* C.A.Mey, AF034118; *Petunia axillaris* (Lam.) Britton, Sterns & Poggenb., U61391; *Nicotiana plumbaginifolia* Willd., X07745; *Helianthus annuus* L., X53537; *Pisum sativum* L., D14698; *Glycine max* Merr., Z14031; *Vigna radiata* (L.) R.Wilczek, AF071550; and *Oenothera biennis* L., X04023. The complete list of *atp1* and *cox1* sequences analyzed, including previously published sequences, is given (Appendix 2).

DNA sequences of *cox1* were largely determined by previously described methods (Bowe et al., 2000). A different forward amplification primer sequence, 130F: 5′-GACCCGGCGATCAAATTCTTG-3′ (Leslie R. Goertzen, Indiana University, personal communication), was useful across a taxonomically broader range of angiosperms. Three new internal primers were designed for use in sequence determination. In each case, these primers were located near the position of a previously published primer to take advantage of a region with minimal interference from secondary structure. The names and sequences of these primers are listed—705R 5′-ACCGAAGAACCGAAAGAGATG-3′; 446R: 5′-GCTGAATCAACTGCTCCTCC-3′ and 1096F: 5′-GACATTGCTCTACATGATACTTATTATG-3′. Sequences were assembled in the program Gene Inspector 1.5 (Textco, West Lebanon, New Hampshire, USA) and aligned with the embedded module Clustal W (Thompson et al., 1994) using default penalty parameters.

***RNA editing***—Three approaches are proposed for the phylogenetic analyses of loci such as *atp1* and *cox1* that are subject to RNA editing. (1) The cDNA sequences may be used, so that data reflect the edited condition, operating under the assumption that the most phylogenetic information is to be found in the edited gene product (Hiesel et al., 1994). (2) Edited sites may be identified and excluded from phylogenetic analyses under the assumption that they impede the accurate inference of phylogeny (Bergthorsson et al., 2003). (3) Unaltered genomic sequences may be analyzed in the pre-edited condition, under the assumption that edited sites have limited phylogenetic impact or are even informational (Bowe and dePamphilis, 1996; Petersen et al., 2006a, b). Because the phylogenetic impact of edited sites is not clear and because relatively few sites are edited (Schuster et al., 1991; Kadowaki et al., 1995; Laser et al., 1995; Giegé and Brennicke, 1999), we favor excluding these sites from analyses. However, for purposes of comparison, edited sites were alternatively included and excluded in phylogenetic analyses to determine if congruence between *atp1* and *cox1* gene trees was in any way affected by these positions.

Sequences from both loci were screened using the program PREP-mt (Mower, 2005). Other methods for prediction of RNA edited sites rely on the presence of signature sequences around edited sites. However, not all edited sites are associated with these sequences (Bowe and dePamphilis, 1996). PREP-mt identifies C to U editing events that increase amino acid conservation. The program predicts the location of edited sites in a plant mitochondrial protein-coding gene by comparing the amino acid sequence inferred from the DNA sequence against a set of aligned experimentally verified proteins from representative green plants. The ability of this program to accurately predict edited sites in mitochondrial loci was recently demonstrated to be above 97.5% (Chaw et al., 2008). PREP-mt analyses were run with the cutoff set to zero, to maximize the predicted number of edited sites. Nucleotide sites that were predicted by PREP-mt to be edited in two or more sequences were initially included and later excluded from phylogenetic analyses for comparative purposes.

***Phylogenetic analyses***—Maximum parsimony (MP), maximum likelihood (ML; Felsenstein, 1973) and Bayesian Markov chain Monte Carlo inference (BI; Yang and Rannala, 1997) methods of analysis were all used. Analyses were run with four programs: PAUP* version 4.0b10 (Swofford, 2002), PAUPRat for

the parsimony ratchet (Nixon, 1999; Sikes and Lewis, 2001), GARLI version 0.951–1 (Zwickl, 2006), and MrBayes version 3.1 (Ronquist and Huelsenbeck, 2003). In all cases, *Amborella trichopoda* Baill. was the specified outgroup. A mutation hotspot, corresponding to nucleotides 571–609 in *atp1* from *Oryza sativa*, in which the frequency of indels and point mutations made alignment uncertain was excluded from all analyses. The *cox1* sequences of *Asimina triloba* Dun. and *Knema latericia* Elmer contain intron sequences that were also excluded as well as the six nucleotides of the coconversion tract that otherwise cause a spurious association between these taxa (Cho et al., 1998).

Heuristic MP analyses for 100 random-addition sequences and tree-bisection-reconnection (TBR) branch swapping were performed for the separate loci using PAUP*. Ten to thirty searches of 200 iterations each were conducted using PAUPRat. Nonparametric parsimony bootstrap analysis (Felsenstein, 1985) was conducted with one random-addition sequence and other default parameters for each of 1 000 pseudoreplicates with TBR swapping. The computer time required for the MP analysis of *cox1* sequences was prohibitive (exceeding 1430 h)—apparently because of time taken in branch swapping on the much larger number of trees that were one step longer than the most parsimonious trees. Because of the high computer burden, the MP bootstrap analysis of the *cox1* sequences was conducted with the "fast heuristic search option" instead of the full heuristic search. A published comparison of this fast method with one in which branch swapping was performed showed effects of only a few percentage points on nodes supported by values at or above 85% (Mort et al., 2000). All reported bootstrap percentages (BP) are parsimony bootstrap values unless specifically indicated to be ML BP.

The best-fit substitution models for the sequences were determined using MODELTEST version 3.7 (Posada and Crandall, 1998). The hypothesis that *atp1* and *cox1* had similar evolutionary histories was supported by the fact that both had the same optimal model, TVM + I + Γ, selected by the Akaike information criterion (Akaike, 1974). Evolutionary parameters for both loci were also largely similar. For *atp1*, these were as follows: nucleotide frequencies—A = 0.2808, C = 0.2182, G = 0.2344, T = 0.2666; substitution rates—A—C = 1.3630, A—G = 1.9560, A—T = 1.2038, C—G = 0.5457, C—T = 1.9560; Pinvar = 0.5261; alpha = 0.6269. The *cox1* model parameters were as follows: nucleotide frequencies—A = 0.2318, C = 0.2113, G = 0.2180, T = 0.3388; substitution rates—A—C = 2.1937, A—G = 1.858, A—T = 1.4045, C—G = 0.8125, C—T =1.8588; Pinvar = 0.5449; alpha = 0.8120. Heuristic ML searches were performed in PAUP* with 100 random addition sequence replicates and TBR swapping using the parameters of the optimal model generated by MODELTEST. To determine the dependence of the ML analysis on these model parameters, ML analyses were also performed with GARLI in which all model parameters were estimated. ML bootstrap analyses were performed in GARLI, estimating all model parameters, for 1000 replicates.

Bayesian inference analyses were also performed (Huelsenbeck and Ronquist, 2001). The general GTR + I + Γ model (nst = 6; rates = invgamma) was used because the TVM model is not implemented in MrBayes. All trees were given equal weight a priori. Sequences were partitioned by gene and site-specific rates were allowed to vary across partitions (ratepr = variable). Default settings were used for other prior probability parameters. All BI analyses were executed for two replicates of 10 000 000 generations each (four chains per replicate run) with trees sampled every 1000 generations. The first 25% of the trees were discarded (burnin = 2501) by default, after which improvement in the range of log-likelihood values was not observed, leaving 7500 trees recovered from each replicate so that the majority rule consensus was of a total of 15 000 trees. Posterior probability (PP) values greater than 89 were reported.

Maximum parsimony, BI, and ML analyses were repeated with the conditions specified above excluding sites predicted to be RNA-edited, and again excluding six taxa that were suspected sources of interlocus phylogenetic conflict. Thus, for each of the two loci three subsets of the data were run with each of the three methods.

Unexpected placements conflicting with multigene phylogenetic results accumulated over the past decade were examined by searching GenBank using the NCBI Blast tool (Altschul et al., 1990) with each sequence of unexpected placement as the query. In some cases we elected to resequence selected loci or chose different published sequences (see section *Sequence authenticity* in Discussion), depending on the outcome of the BLAST searches.

***Parametric bootstrap study***—A parametric bootstrap analysis was performed to determine whether an unexpected association between Acoraceae and *Lilium* in *cox1* gene trees might be attributed to long-branch attraction (LBA) because these two taxa are on the longest branches in these analysis. A subset of 17 species of monocots plus one outgroup were selected for the simulations including *Acorus calamus* L., *A. gramineus* Soland., *Alisma plantago-aquatica* L., *Arisaema triphyllum* Schott, *Asparagus officinalis* L., *Carludovica palmata* Ruiz & Pav., *Dioscorea macrostachya* M.Martens & Galeotti, *Elodea* Juss. sp., *Lilium lancifolium* Thunb., *Joinvillea plicata* (Hook. f.) Newell & Stone, *Liriodendron tulipifera* L., *Oryza sativa*, *Pleea tenuifolia* Michx., *Potamogeton berchtoldii* Fieb., *Sagittaria latifolia* Willd., *Smilax* L. sp., *Spathiphyllum* Juss. cv. Clevelandii and *Tofieldia calyculata* (L.) Wahlenb. Three sets of relationships were separately tested. (1) The two species of *Acorus*, which were sister to each other, were constrained to be sister to the remaining 15 monocots, which is a result that has been recovered in analyses of chloroplast (APG II, 2003; Cai et al., 2006) and selected nuclear sequences (Duvall and Ervin, 2004) and multigene studies including these loci. (2) *Lilium* was constrained to be sister to *Smilax* because the current evidence largely supports a sister group relationship between Liliaceae and Smilaceae (APG II, 2003). (3) The two long-branch taxa, Acoraceae and *Lilium*, were allowed to unite in an unconstrained ML search. For each of these three models, 100 data sets of the same size as the original *cox1* data matrix were generated with the program Seq-Gen version 1.2.5 (Rambaut and Grassly, 1997).

Branch lengths and other parameters for the simulations were obtained from MODELTEST. Each of the 300 simulated data sets was then analyzed using heuristic ML searches with 100 random addition sequence replicates and TBR branch swapping in PAUP*. Model trees were compared to inferred trees to determine if the long branches in *cox1* trees leading to Acoraceae and *Lilium* were attracted to each other or to other long branches in the trees, when other models were specified.

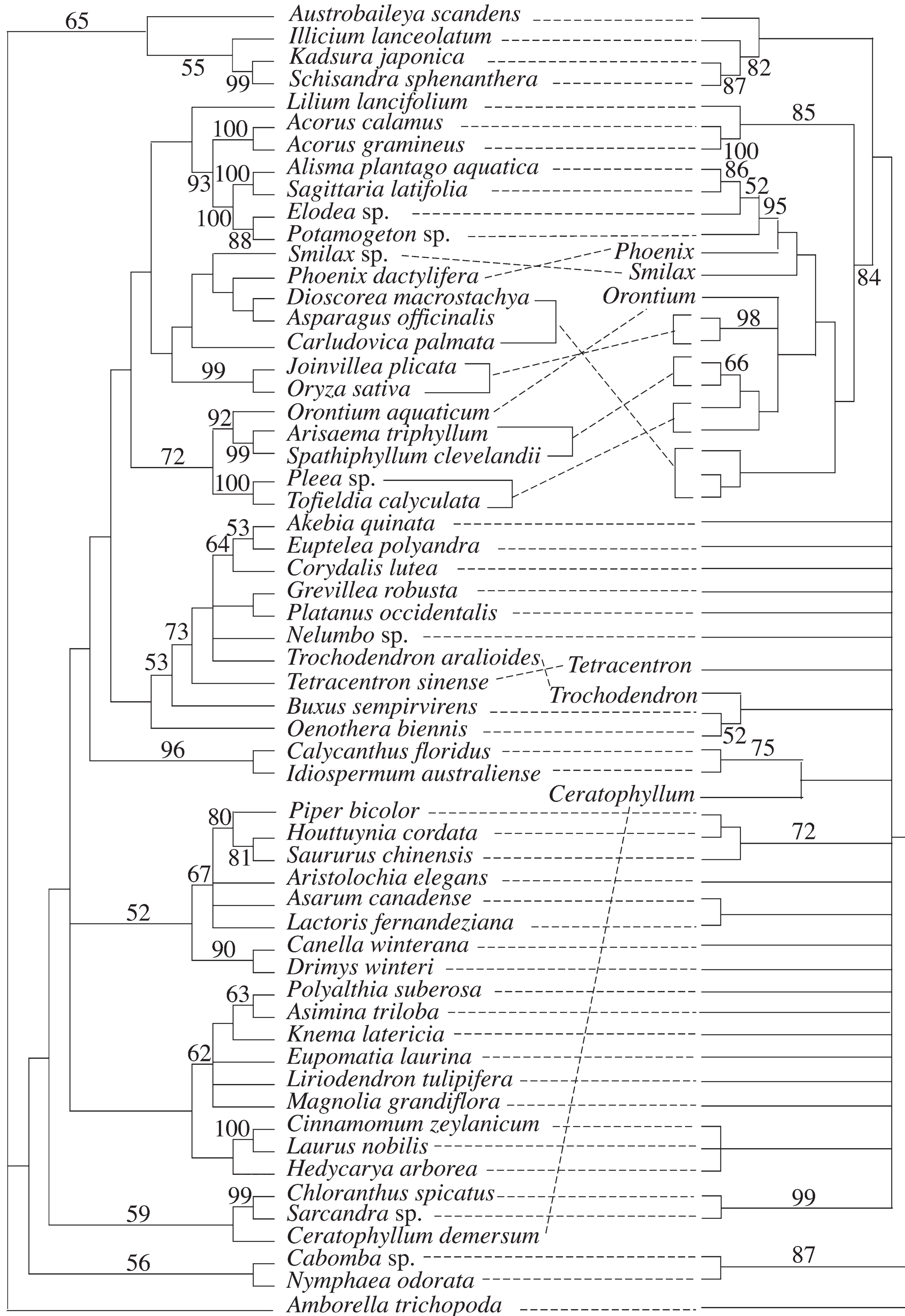
***Congruence tests***—Both parsimony-based and ML-based tests of congruence were conducted on the separate *atp1* and *cox1* gene trees. For a parsimony-based test, the two mitochondrial loci were treated as separate partitions for the incongruence length difference (ILD; Farris et al., 1995) test, implemented in PAUP* as the partition homogeneity test. One random addition sequence was performed for each of 100 replicates. The ILD test was repeated on the data matrix excluding 55 sites predicted to be RNA-edited and was also performed on a subset of 52 taxa, excluding long-branch or otherwise problematic taxa. The ILD test is a useful starting point for exploring congruence (Sanderson and Shaffer, 2002 Hipp et al., 2004; contra Yoder et al., 2001), although it is relatively insensitive to incongruence under certain conditions (Darlu and Lecointre, 2002). A second parsimony-based test, the Templeton test (Templeton, 1983) implemented in PAUP*, was also used to test the congruence of the most parsimonious trees produced from analyses each of the same three subsets of the data matrix.

The Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa, 1999) was selected for likelihood-based tests of congruence. Phylograms and summary model parameters generated by the Bayesian inference analyses were used for SH tests. The SH test is a conservative test that performs well when there are few trees (in this case two per comparison) and is excellent at controlling type 1 statistical errors (Shimodaira and Hasegawa, 2002). The SH tests were implemented in PAUP* on the three sets of single-gene trees described earlier. In each case, reciprocal SH tests were conducted for each mitochondrial locus.

## RESULTS

GenBank accession numbers for the new sequences determined from *atp1* (three) and *cox1* (19) are listed together with those from previously banked sequences (Appendix 2). The aligned data and ML trees were submitted to TreeBase (SN3833). The overall data matrix included perfectly parallel sampling of 52 taxa for the two loci. Additionally, five taxa were "composite terminals" (Nixon and Davis, 1991) in which the two sequences were obtained from different congeners drawn from relatively small monophyletic genera. These six species pairs were *Asparagus officinalis* L., *A. falcatus* L.; *Cabomba* Aubl. sp. *Palmer 688*, *C.* sp. *Qiu 97027* (which may be from the same species); *Nelumbo nucifera* Gaertn., *N. lutea* Pers., *Potamogeton berchtoldii*, *P. crispus* L.; *Sarcandra chloranthoides* Gardner, *S. grandifolia* (Miq.) Subramanyam & A.N.Henry; and *Smilax rotundifolia* L., *S. tamnoides* L.

Prediction of edited sites by PREP-mt among the *atp1* sequences analyzed here ranged from none in 10 species, to 10 edited sites in three species with a mean value of five predicted


Austrobaileya scandens
Illicium lanceolatum
Kadsura japonica
Schisandra sphenanthera
Lilium lancifolium
Acorus calamus
Acorus gramineus
Alisma plantago aquatica
Sagittaria latifolia
Elodea sp.
Potamogeton sp.
Smilax sp.
Phoenix dactylifera
Dioscorea macrostachya
Asparagus officinalis
Carludovica palmata
Joinvillea plicata
Oryza sativa
Orontium aquaticum
Arisaema triphyllum
Spathiphyllum clevelandii
Pleea sp.
Tofieldia calyculata
Akebia quinata
Euptelea polyandra
Corydalis lutea
Grevillea robusta
Platanus occidentalis
Nelumbo sp.
Trochodendron aralioides
Tetracentron sinense
Buxus sempirvirens
Oenothera biennis
Calycanthus floridus
Idiospermum australiense
Piper bicolor
Houttuynia cordata
Saururus chinensis
Aristolochia elegans
Asarum canadense
Lactoris fernandeziana
Canella winterana
Drimys winteri
Polyalthia suberosa
Asimina triloba
Knema latericia
Eupomatia laurina
Liriodendron tulipifera
Magnolia grandiflora
Cinnamomum zeylanicum
Laurus nobilis
Hedycarya arborea
Chloranthus spicatus
Sarcandra sp.
Ceratophyllum demersum
Cabomba sp.
Nymphaea odorata
Amborella trichopoda
Phoenix
Smilax
Orontium
Tetracentron
Trochodendron
Ceratophyllum
65
55
99
100
93
100
100
88
99
92
99
72
100
53
64
73
53
96
80
81
67
52
90
63
62
100
99
59
56
82
87
85
100
86
52
95
84
98
66
52
75
72
99
87

C to U edits per *atp1* sequence. Eleven of these putatively edited sites were shared by two or more sequences and were excluded from alternate phylogenetic analyses of *atp1* for comparative purposes. The number of analyzed, nonedited sites of *atp1* was 1110, of which 246 were parsimony informative.

Prediction of edited sites by PREP-mt among the *cox1* sequences analyzed here was more variable than for *atp1*, ranging from none, in *Alisma plantago-aquatica* and *Canella winterana* Gaertn., to 40 edited sites in *Amborella trichopoda* with a mean value of 24 predicted C to U edits per sequence. Forty-four of these putatively edited sites were shared by two or more sequences and were excluded from alternate phylogenetic analyses of *cox1*. The number of analyzed, nonedited positions of *cox1* was 1299 sites, of which 162 were parsimony informative.

Single-locus MP analysis of *atp1* produced 24 equally parsimonious trees of 897 steps (ensemble consistency index [Kluge and Farris, 1969] excluding uninformative characters, CI = 0.474; ensemble retention index [Farris, 1989], RI = 0.698; see strict consensus tree, Fig. 1). Forty-seven nodes were resolved in the strict consensus of these trees. Parsimony ratchet analysis found the same most parsimonious trees. Posterior probabilities from 15 000 trees were mapped onto the Bayesian inference phylogram (Fig. 2). Maximum likelihood analysis of *atp1* produced one tree (−lnL = 6495.08), which is posted at the TreeBASE database (http://www.treebase.org) together with the results of the ML bootstrap analysis.

Single-locus MP analysis of *cox1* produced 47 743 trees of 587 steps (CI = 0.510; RI = 0.671) after 100 random addition sequence replicates. The strict consensus of these trees had only 32 of 57 possible resolved nodes (Fig. 1). Parsimony ratchet analysis found the same most parsimonious trees for the 52 taxon subset, but found trees one step longer than the trees of the MP analysis in PAUP* for the two 58 taxon subsets. Posterior probabilities from 15 000 trees were mapped onto the Bayesian inference phylogram (Fig. 3). Maximum likelihood analysis of *cox1* produced one tree (−lnL = 5531.23), which is posted at TreeBASE together with the results of the ML bootstrap analysis. Because the BI and ML analyses were topologically congruent for both genes, the former will be the focus of further discussion in this paper.

The ILD tests indicated significant incongruence between the two mitochondrial partitions in the complete data matrix (Table 1). All of the summed tree lengths from the random partitions in the ILD tests were greater than the sum of the trees from the original partitions. Parsimony-based Templeton tests also indicated significant incongruence between *atp1* and *cox1* gene trees in all cases ($P < 0.001$, Table 1). SH tests likewise found the *atp1* and *cox1* gene trees to be incongruent (Table 1).

In the ML trees of the parametric bootstrap study, long-branch taxa failed to associate whenever they were not initially united in the model tree. In the unconstrained model, Acoraceae were sister to *Lilium* in 99% of the inferred trees (Table 2) and paraphyletic with *Lilium* in the tree from the remaining replicate. When Acoraceae were constrained in a sister group position to the other monocots, that placement was retrieved in 71% of the inferred trees. In the other 29%, *Acorus* was variously united with *Carludovica*, *Dioscorea*, *Potamogeton*, and/or *Smilax*, but never with *Lilium*. When *Lilium* was constrained to a sister group position with *Smilax*, 100% of the trees retained that aspect of the topology and Acoraceae were variously sister to one or more of Alismatales, Poales, and/or *Asparagus*, but never to *Lilium*. Thus under each of the three models, the specified phylogenetic positions of *Acoraceae* and *Lilium* were retrieved in 71–100% of the simulations.

## DISCUSSION

***Comparison of single-gene analyses***—The deeper nodes in the *atp1* tree were generally better resolved and more strongly supported than those in the *cox1* tree. Nine clades corresponding to previously recognized orders or taxa of higher rank were resolved by analyses of the *atp1* sequences, confirming published multigene analyses (Figs. 1, 2; Table 3). These were Austrobaileyales (*Austrobaileya*, *Illicium*, *Kadsura* and *Schisandra*), Canellales (*Canella* and *Drimys*), Chloranthales (*Chloranthus* and *Sarcandra*), eudicots (10 spp., Appendix 2), magnoliids (19 spp.), monocots (19 spp.), Nymphaeales (*Nymphaea* and *Cabomba*), Poales (*Oryza* and *Joinvillea*) and Piperales (six spp.). Note that in most cases, the MP BP offers a more conservative indication of support than the PP for the same clade. Only five of these clades, Austrobaileyales, Chloranthales, monocots Nymphaeales and Poales, were also supported in the *cox1* trees (Figs. 1, 2; Table 2). *Ceratophyllum* was resolved as the sister taxon to Chloranthaceae in the *atp1* trees (BP = 59; PP = 100) but was sister to Calycanthaceae in the *cox1* MP trees (BP = 75) and unresolved in the BI analysis.

The *atp1* trees strongly supported (BP > 95; PP = 100) five families as clades (Table 3). Among monocots Acoraceae (*A. calamus* and *A. gramineus*.) and Alismataceae (*Alisma* and *Sagittaria*) were strongly supported as clades. In Laurales, two families, Calycanthaceae (*Calycanthus* and *Idiospermum*) and Lauraceae (*Cinnamomum* and *Laurus*) were strongly supported. The monophyly of Schisandraceae (Austrobaileyales) was also strongly supported. There was at least moderate congruent support (BP ≥ 75; PP ≥ 99) in the *cox1* trees for four of these, but the two species of Lauraceae were in a weakly supported, unresolved trichotomy with *Hedycarya* (Monimiaceae) in both MP and BI analyses (Figs. 1, 3).

Also receiving strong support were two clades corresponding to portions of families. In each of these cases, the topology of the *cox1* tree shows an unexpected, though not necessarily strongly supported, paraphyletic arrangement for the subtaxon with other taxa. These included two of the three Araceae, *Arisaema* and *Spathiphyllum* (*atp1*: BP = 99; PP = 100; *cox1*: BP = 66; PP = 100; Araceae were weakly paraphyletic with Tofieldiaceae in the *cox1* MP and BI trees). Three of the six Piperales, *Houttuynia*, *Piper* and *Saururus,* were supported as monophyletic (*atp1*: BP = 80; PP = 100; *cox1*: BP = 72; PP = 100; Piperales were broadly paraphyletic with other dicots in the *cox1* ML tree). However, the internal arrangements of the latter clade differed between the two trees (Figs. 1–3).

←

Fig. 1. Strict consensus maximum parsimony trees for *atp1* (left side: consensus of 24 trees of length 897 steps) and *cox1* (right side: consensus of 47 743 trees of length 587 steps) for 58 taxa. Nucleotide sites putatively subject to RNA editing were excluded from both analyses. Numbers along the branches are bootstrap values >50%. Dashed lines connect identical terminals or identical groups of terminals. In some cases the name of the taxon (genus only) is repeated on the *cox1* tree for clarification.

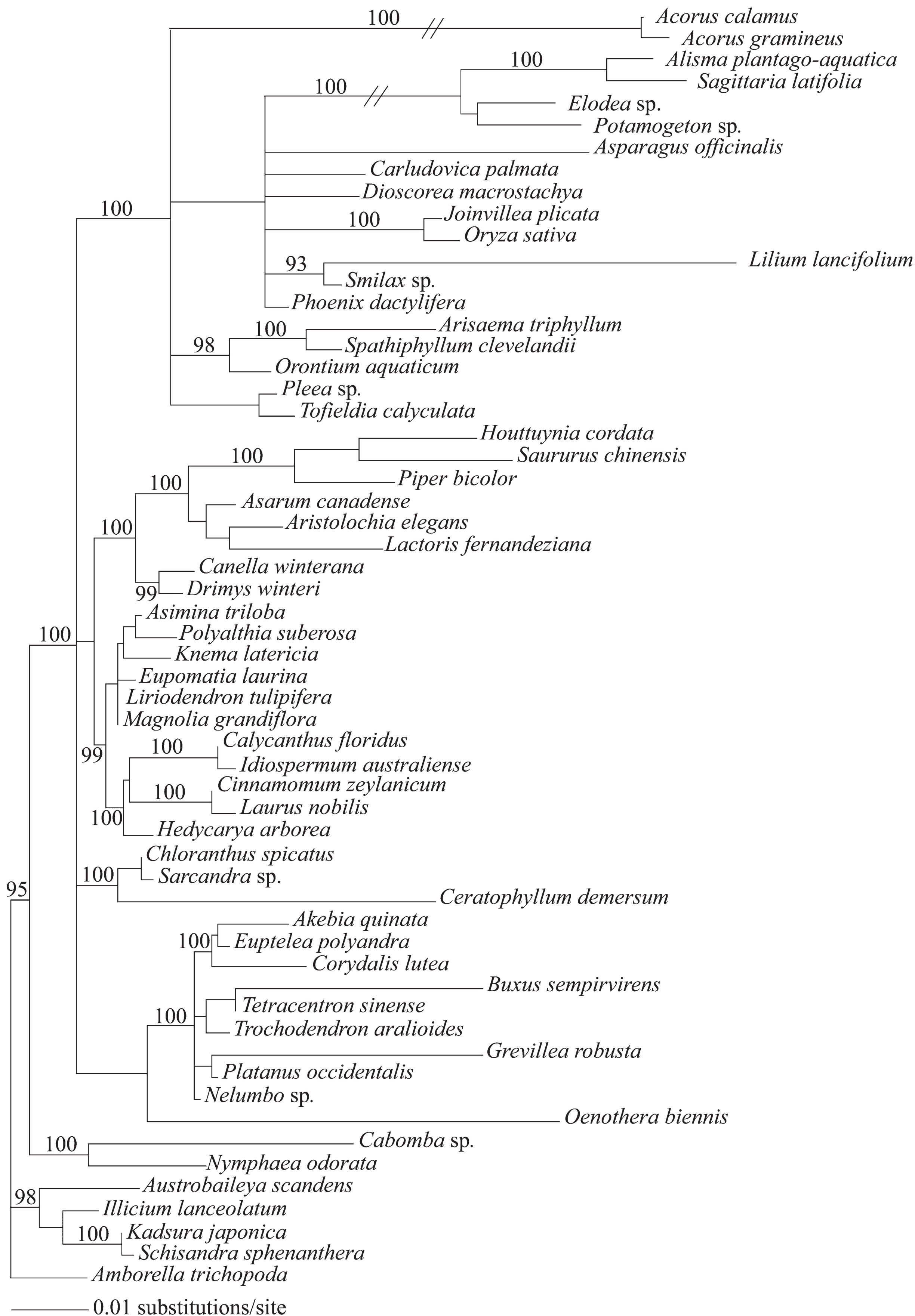

100
Acorus calamus
Acorus gramineus
100
100
Alisma plantago-aquatica
Sagittaria latifolia
Elodea sp.
Potamogeton sp.
Asparagus officinalis
Carludovica palmata
Dioscorea macrostachya
100
100
Joinvillea plicata
Oryza sativa
93
Lilium lancifolium
Smilax sp.
Phoenix dactylifera
98
100
Arisaema triphyllum
Spathiphyllum clevelandii
Orontium aquaticum
Pleea sp.
Tofieldia calyculata
100
100
Houttuynia cordata
Saururus chinensis
Piper bicolor
Asarum canadense
Aristolochia elegans
Lactoris fernandeziana
100
Canella winterana
Drimys winteri
99
100
Asimina triloba
Polyalthia suberosa
Knema latericia
Eupomatia laurina
Liriodendron tulipifera
Magnolia grandiflora
99
100
Calycanthus floridus
Idiospermum australiense
100
Cinnamomum zeylanicum
Laurus nobilis
100
Hedycarya arborea
Chloranthus spicatus
Sarcandra sp.
100
95
Ceratophyllum demersum
100
Akebia quinata
Euptelea polyandra
Corydalis lutea
Buxus sempirvirens
100
Tetracentron sinense
Trochodendron aralioides
Grevillea robusta
Platanus occidentalis
Nelumbo sp.
Oenothera biennis
100
Cabomba sp.
Nymphaea odorata
98
Austrobaileya scandens
Illicium lanceolatum
100
Kadsura japonica
Schisandra sphenanthera
Amborella trichopoda
0.01 substitutions/site

TABLE 1. The *p*-values resulting from parsimony-based (ILD and Templeton) and likelihood-based (SH) tests of congruence between *atp1* and *cox1* rounded to two decimal places. Note that all tests were significant at $P < 0.05$.

| Test | 58 Sequences, RNA-edited sites included. | 58 Sequences, RNA-edited sites excluded. | 52 Sequences, RNA-edited sites excluded. |
|---|---|---|---|
| ILD | 0.01 | 0.04 | 0.02 |
| Templeton | 0.00 | 0.00 | 0.00 |
| SH | 0.00 | 0.00 | 0.01 |

TABLE 2. Percentage recovery of model trees in simulations in which three sets of relationships were separately tested to assess whether the unexpected association between Acoraceae and *Lilium* might be attributed to long-branch attraction.

| Model tree | Inferred tree | | |
|---|---|---|---|
| | Acoraceae sister to monocots | *Lilium* sister to *Smilax* | Acoraceae sister to *Lilium* |
| Acoraceae sister to monocots | 71 | 0 | 0 |
| *Lilium* sister to *Smilax* | 0 | 100 | 0 |
| Acoraceae sister to *Lilium* | 0 | 0 | 99 |

Aside from these points of similarity, the *atp1* and *cox1* trees had numerous topological differences, but these were rarely strongly supported in both trees. Differences between the two mitochondrial gene trees caused the two partitions to consistently fail tests of congruence. This incongruence was unexpected, given the general similarities between these two loci and their gene products. Several hypotheses were tested in an effort to eliminate artifactual sources of phylogenetic incongruence.

***Sequence authenticity***—The first of these was a test of sequence authenticity. For example, three banked sequences of *cox1*, those of *Liriodendron* (AF193959, Parkinson et al., 1999), *Nelumbo nucifera* (AF193950, Parkinson et al., 1999) and *Tetracentron* (AY009455, Barkman et al., 2000) were unexpectedly identical across their entire aligned lengths of 1413 bases. Including these three sequences in phylogenetic analyses resulted in paraphyletic arrangements of Magnoliales and eudicots in MP, ML, and BI *cox1* trees (results not shown). This situation was resolved by resequencing one of these (*Liriodendron* [DQ980415]) and substituting alternative banked sequences for the other two (*Nelumbo lutea* [AY009447] and *Tetracentron* [AF193964]) in the data matrix.

Whether these banked sequences were actual and atypical *cox1* paralogues or simply misidentified sequences is unknown, but the following example shows a specific instance of misidentification. At the time of the initial analyses for this paper, there were two banked sequences of *cox1* for *Austrobaileya scandens* C. T. White (AY009434, Barkman et al., 2000; AF193954, Parkinson et al., 1999). These two sequences differed from each other by 28 substitutions across 1390 bp. In BLAST searches, the former sequence targeted other members of Austrobaileyales whereas the latter targeted *Ceratophyllum* and so was not chosen for this paper. By this writing, the source taxon for the AF193954 entry has been changed in GenBank, to *Ceratophyllum demersum* L. However, the accession number remains incorrect in printed records (Table S1, p. S2, supplementary material in Parkinson et al., 1999; M. Duvall, unpublished BLAST report, July 2006).

The *Lilium cox1* sequence determined for this study also showed strongly supported phylogenetic incongruence with *atp1*. An unanticipated and moderately supported association was observed uniting Acoraceae and *Lilium* in *cox1* MP, ML, and BI trees (BP = 85; PP = 91; Fig. 3). Three *cox1* sequences from *Acorus* were produced in two different laboratories from source materials harvested from *A. americanus* Raf., *A. calamus*, and *A. gramineus* and were 98% identical across 1351 bp. The *cox1* sequences of *A. americanus* and *A. calamu*s differed at only three sites so that only the latter was included in our phylogenetic analyses. The similarities among these three independently produced sequences argue for their authenticity. The *cox1* sequence from *Lilium* proved to be reproducible from a new DNA extract obtained from newly harvested, confidently identified leaf tissue of the same species, and so the reproducibility of this sequence was confirmed.

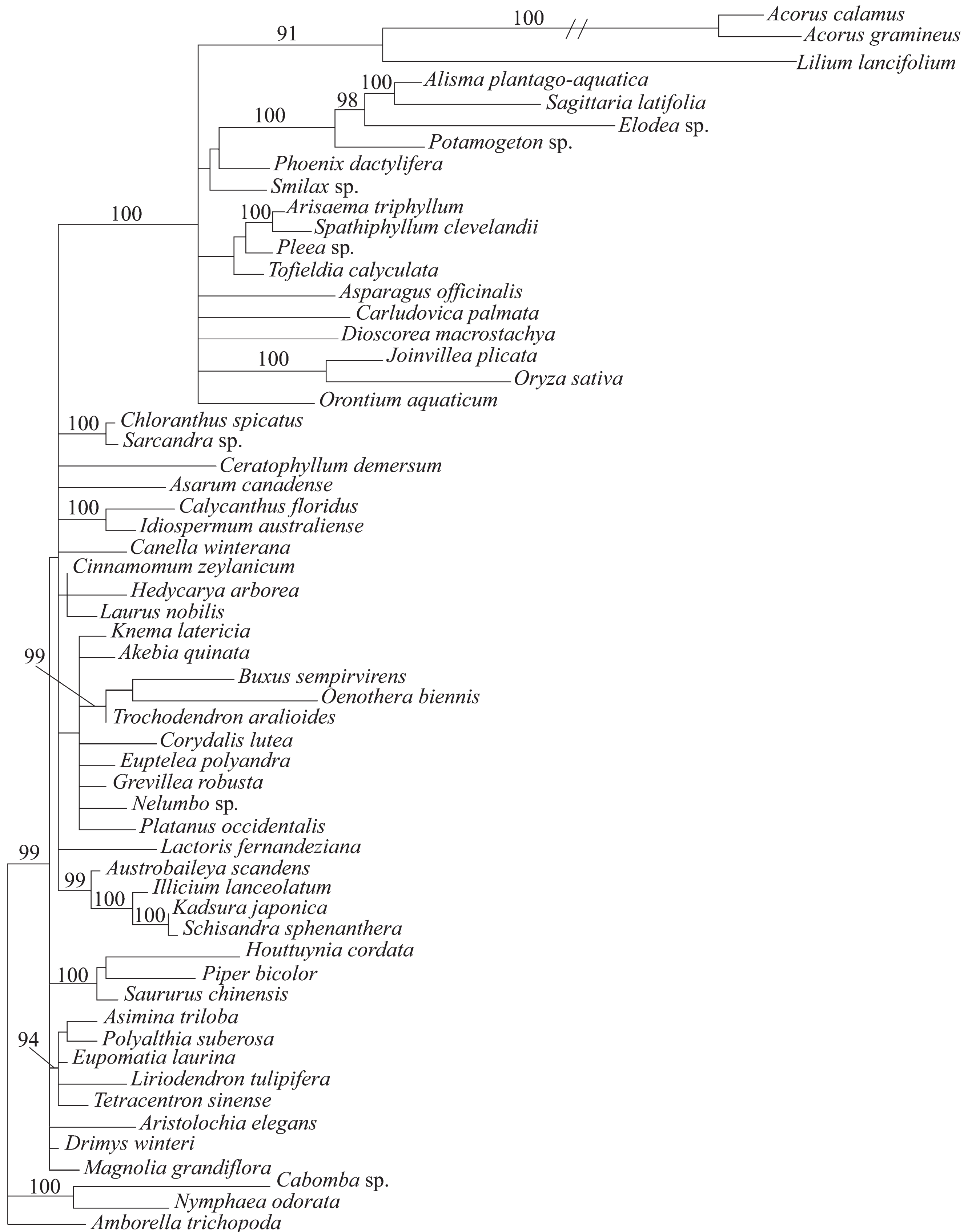
***Long-branch attraction***—Acoraceae and a clade of four Alismatales (*Alisma*, *Elodea*, *Potamogeton*, and *Sagittaria*) are on the longest branches in the *atp1* trees probably due to higher rates of substitution in these two lineages (see Fig. 2). These long branches are sister lineages in the MP *atp1* tree (BP = 93; Fig. 1), but not in the BI tree. Their position as a highly derived clade of monocots in the MP tree is unexpected, but this position is only weakly supported (BP < 50). The separation of these two long branches in the ML and BI trees (Fig. 2) suggests that this is a LBA artifact to which parsimony is more susceptible than likelihood (Huelsenbeck, 1997). However, the previously established relationship between these clades makes their strongly supported association in the MP *atp1* tree unsurprising and would complicate an investigation of this as a putative systematic error.

A less obscure example is to be found in the *cox1* trees where Acoraceae and *Lilium* are on the two longest branches (see Fig. 3) as sister lineages. The unexpected and moderately supported association between these taxa (BP = 85; PP = 91) suggested a LBA artifact. We chose four approaches to test for evidence of LBA. First, MP, BI, and ML methods were all employed in analyses of the *cox1* sequences. All produced trees with this unexpected association. Even though ML is less biased to LBA, it is not completely unaffected by this type of systematic error (Bergsten, 2005), and so this finding is inconclusive.

Second, an attempt was made to break up the long branches by including *cox1* sequences from related taxa. The *cox1* sequence of *A. americanus* was too similar to that of *A. calamus* to divide the branch leading to Acoraceae (see earlier). Some other lilioid taxa, such as *Erythronium*, also have high substitution rates so that adding them to the analysis has the effect of replacing one long branch by another (trees not shown). The *cox1* sequence of *Fritillaria crassifolia* Boiss. & Huet differed from that of *Lilium lancifolium* by 10 substitutions. A ML bootstrap analysis was performed on a subset of the *cox1* sequences to reduce the computation time. Adding the *Fritillaria*

←

Fig. 2. Bayesian inference phylogram for 58 *atp1* sequences. Branches are depicted as proportional to their calculated mean lengths except in two instances where unusually long branches were truncated and marked (//) to make the figure more readable. Numbers along the branches are posterior probability (PP) values >89% from the Bayesian inference analysis.



100
91
Acorus calamus
Acorus gramineus
Lilium lancifolium
100
Alisma plantago-aquatica
98
Sagittaria latifolia
100
Elodea sp.
Potamogeton sp.
Phoenix dactylifera
Smilax sp.
100
100
Arisaema triphyllum
Spathiphyllum clevelandii
Pleea sp.
Tofieldia calyculata
Asparagus officinalis
Carludovica palmata
Dioscorea macrostachya
100
Joinvillea plicata
Oryza sativa
Orontium aquaticum
100
Chloranthus spicatus
Sarcandra sp.
Ceratophyllum demersum
Asarum canadense
100
Calycanthus floridus
Idiospermum australiense
Canella winterana
Cinnamomum zeylanicum
Hedycarya arborea
Laurus nobilis
Knema latericia
99
Akebia quinata
Buxus sempirvirens
Oenothera biennis
Trochodendron aralioides
Corydalis lutea
Euptelea polyandra
Grevillea robusta
Nelumbo sp.
Platanus occidentalis
99
Lactoris fernandeziana
99
Austrobaileya scandens
100
Illicium lanceolatum
100
Kadsura japonica
Schisandra sphenanthera
Houttuynia cordata
100
Piper bicolor
Saururus chinensis
Asimina triloba
94
Polyalthia suberosa
Eupomatia laurina
Liriodendron tulipifera
Tetracentron sinense
Aristolochia elegans
Drimys winteri
Magnolia grandiflora
100
Cabomba sp.
Nymphaea odorata
Amborella trichopoda
0.005 substitutions/site

TABLE 3. Parsimony bootstrap (MP BP) and posterior probability (PP) values for taxonomically defined clades that were retrieved in one or both of the mitochondrial gene trees.

| Clade (no. species) | *atp1* MP BP | *atp1* PP | *cox1* MP BP | *cox1* PP |
|---|---|---|---|---|
| Austrobaileyales (4 species) | 65 | 98 | <50 | 99 |
| Schisandraceae (2 species) | 99 | 100 | 87 | 100 |
| Canellales (2 species) | 90 | 99 | NR | NR |
| Chloranthales (2 species) | 99 | <90 | 99 | 100 |
| Eudicots (10 species) | <50 | <90 | NR | NR |
| Magnoliids (19 species) | NR | <90 | NR | NR |
| Calycanthaceae (2 species) | 96 | 100 | 75 | 100 |
| Lauraceae (2 species) | 100 | 100 | NR | NR |
| Nymphaeales (2 species) | 56 | 100 | 87 | 100 |
| Piperales (6 species) | 67 | 100 | NR | NR |
| Monocots (19 species) | <50 | 100 | 84 | 100 |
| Acoraceae (2 species) | 100 | 100 | 100 | 100 |
| Poales (2 species) | 99 | 100 | 98 | 100 |
| Alismataceae (2 species) | 100 | 100 | 86 | 100 |

*Note:* NR = not resolved

sequence failed to disrupt the acoroid/lilioid association (ML BP = 83; Fig. 4), although because Liliaceae were still resolved on a relatively long branch, it was potentially still susceptible to LBA.

Third, a parametric bootstrap study tested whether the two long-branch taxa associated in spite of models in which (1) Acoraceae were constrained as the sister to the remaining monocots and (2) *Lilium* and *Smilax* were constrained to be sister to each other. These two models test more expected phylogenetic placements for both taxa (APG II, 2003). However, in neither constrained model did the two long-branch taxa ever associate with each other after ML analysis of 100 simulated *cox1* matrices each (Table 2). (See section *Problematic species* later for a description of the results of the fourth approach, exclusion of long-branch taxa.) So while we were unable to eliminate LBA as a source of possible systematic error for these taxa, we were also not able to find evidence for it.

***RNA editing***—The potential impact of RNA editing on phylogenetic topology was also investigated. A total of 55 sites, 11 in *atp1* and 44 in *cox1*, which were predicted to be RNA-edited by unexpected differences in coding sequences, were excluded from the 58-taxon data set and the phylogenetic analyses were repeated. Exclusion of putatively RNA-edited nucleotides somewhat increased resolution and support. For example, in BI analyses of *cox1*, the clade of angiosperms other than *Amborella* and Nymphaeales was not resolved, whereas in the analysis without edited sites, this clade was strongly supported (PP = 99). RNA-edited sites might be expected to increase the incidence of apparently parallel mutations that would conflict with real synapomorphic substitutions presumably lowering both resolution and support. However, the exclusion of these sites still failed to produce mitochondrial gene trees that were congruent under either MP (ILD and Templeton tests) or ML (SH tests) methods (Table 1).

***Problematic species***—In some cases phylogenetic incongruence is restricted to specific taxa in an analysis because of unique historical events such as hybridization, horizontal gene transfer, or lineage sorting. Removal of these taxa eliminates the source of conflict so that congruent gene trees can be produced. While there are many differences and surprising relationships in the *atp1* and *cox1* trees, most of them have little support. Two examples of at least moderately supported conflicts between the *atp1* and *cox1* gene trees were identified. First, the position of *Oenothera* in a clade with *Buxus* and *Trochodendron* in the *cox1* trees was somewhat unexpected (BP < 50; PP = 99). Because *Oenothera* was the only relatively derived eudicot in the analysis, its position in this clade may be an artifact of uneven sampling. However, another trochondendroid, *Tetracentron*, was excluded from this clade. Moreover, *Oenothera* occupied a weakly supported position as the sister to the remaining eudicots in the *atp1* trees (BP < 50; PP < 90) as might be expected for an isolated derived eudicot. *Oenothera* was judged to be problematic on the basis of sampling, as were *Buxus* and *Trochodendron* by association in the *cox1* tree.

Second, Acoraceae were strongly supported in an unexpected sister position with *Lilium* in the *cox1* tree, possibly because of LBA, as noted before (BP = 85; PP = 91). Acoraceae associated with a subset of Alismatales in the MP *atp1* trees (BP = 93) and are unresolved in the BI tree (PP < 50). *Lilium* was in an expected, though more weakly supported sister position with *Smilax* in the *atp1* BI tree (PP = 93) and not resolved in the MP tree. The positions of both *Lilium* and *Smilax* were only weakly supported in the *atp1* MP tree. Thus, *Acorus calamus*, *A. gramineus*, and *Lilium* were considered to be problematic.

Because Acoraceae, *Lilium*, *Buxus*, *Oenothera*, and *Trochodendron* were involved in conflicting clades that were at least moderately supported in some of the single locus trees, the phylogenetic analyses for the nonedited sites of both loci were repeated after exclusion of these taxa. However, even after removal of these putatively problematic taxa and the RNA-edited sites, the resulting trees were still incongruent under both parsimony and likelihood-based tests (Table 1).

***Conclusion***—We present evidence of interlocus incongruence between two mitochondrial protein-coding metabolic genes with the same optimal evolutionary model. Ultimately, we were unable to attribute incongruence to erroneous sequences, attraction between long-branch taxa or other specific problematic taxa, inclusion of RNA-edited sites, or choice of phylogenetic method. Such incongruence may be attributed to fundamentally different molecular evolutionary histories between these apparently similar mitochondrial loci, such as might be caused by ancient lineage sorting, horizontal gene transfers (HGT), or unrecognized gene duplications.

The persistent strongly supported placement of Acoraceae within a clade of lilioid monocots in *cox1* gene trees (Figs. 3, 4) is in opposition to the evidence of morphology and other genes (APG II, 2003). Such a discrepancy is suggestive of either a duplication event or a HGT (Graur and Li, 2000). The former is less likely because it presupposes that there is an atypical,

←

Fig. 3. Bayesian inference phylogram for 58 *cox1* sequences. Branches are depicted as proportional to their calculated mean lengths except in two instances where unusually long branches were truncated and marked (//) to make the figure more readable. Numbers along the branches are posterior probability (PP) values >89% from the Bayesian inference analysis. Note the unexpected and well-supported position of Acoraceae as the sister group to *Lilium lancifolium*.

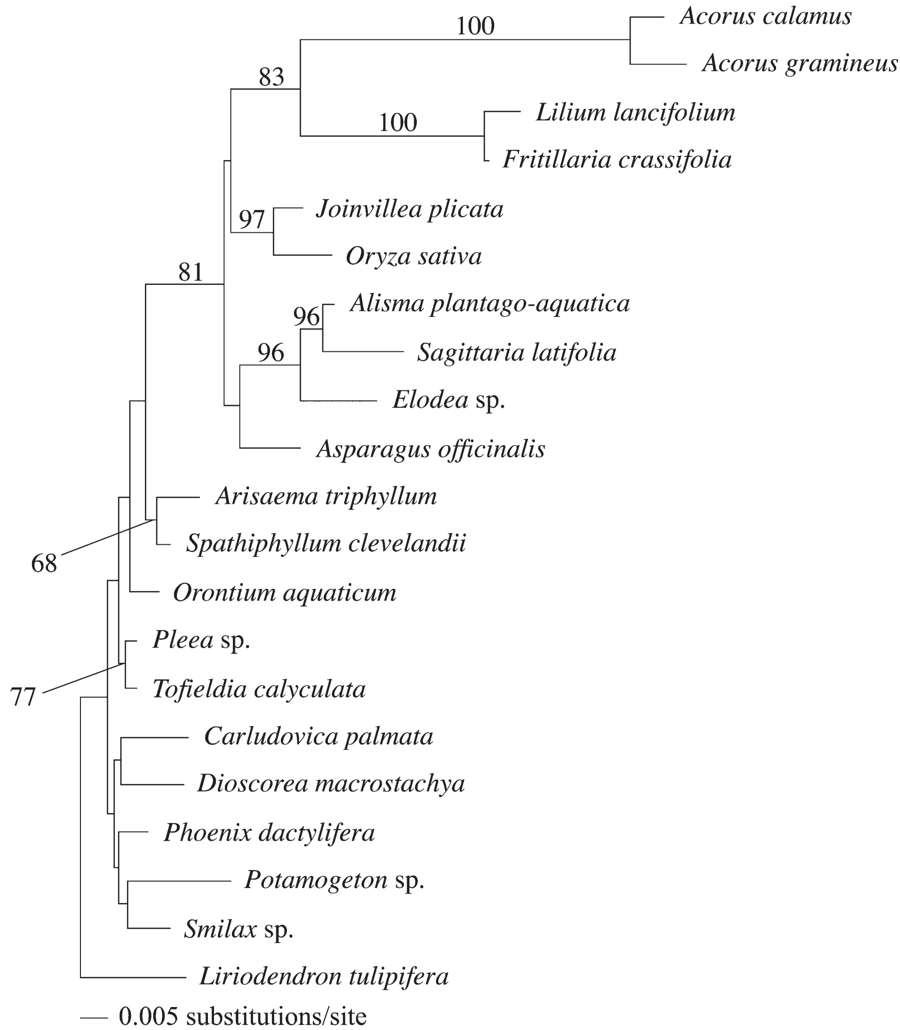


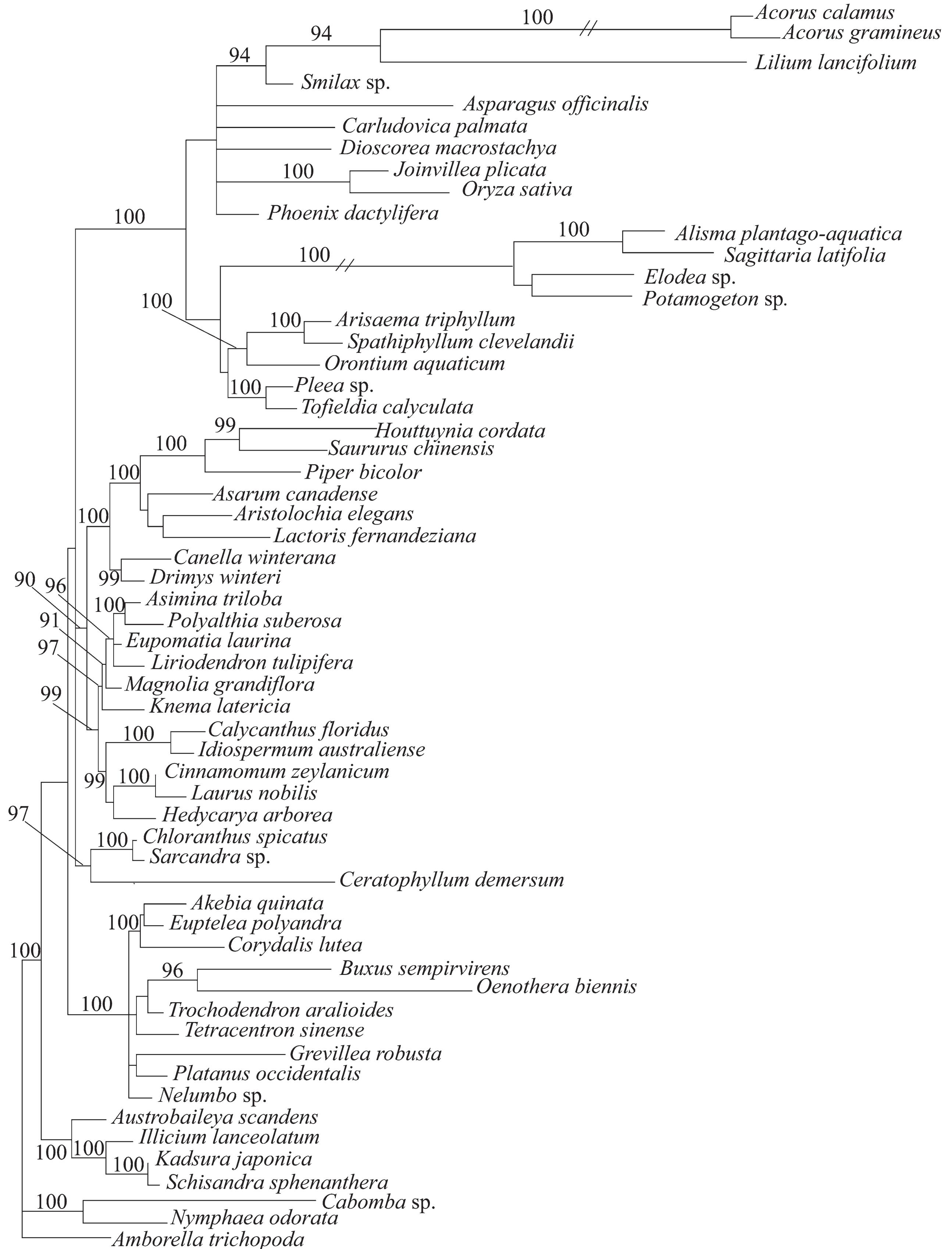

Fig. 4. Maximum likelihood tree for 20 *cox1* sequences from monocots plus one outgroup. Branches are depicted as proportional to their calculated lengths. Numbers along the branches are maximum likelihood bootsrap values (ML BP) >50%. Note the unexpected and moderately well-supported position of Acoraceae as the sister group to Liliaceae (ML BP = 83).

unidentified paralogue (probably of *cox1*) that was coincidentally and preferentially amplified and sequenced in multiple species. Horizontal gene transfer events have been identified among other plant mitochondrial genes (e.g., Bergthorsson et al., 2003, 2004). The tree of Fig. 4 suggests that the transfer occurred between an ancestor of the two lilioid taxa and an ancestor of the Acoraceae analyzed here. Because *Lilium* diverged from *Fritillaria* an estimated five million years ago (Vinnersten and Bremer, 2001), this places an upper bound on the age of this event. Continuing this line of reasoning, because exclusion of these lineages fails to produce congruent mitochondrial gene trees, other events, possibly including other HGTs, must also be invoked. Moreover, because other instances of strongly supported topological differences are not otherwise apparent, these other events may be of greater age and are now partly obliterated by subsequent mutational histories. Confirmation of putatively ancient events that are now identifiable only as a background of unexpected, but weakly supported relationships may be impractical.

In general the combination of congruent loci is expected to increase phylogenetic signal and decrease homoplasy. A combined *atp1*, *cox1* analysis was performed with MP ("fast heuristic" bootstrap), BI, and ML over the 2409 unedited sites to see the effect of combining these incongruent sequences into a single analysis (BI results are given in Fig. 5; MP and ML results are posted at TreeBase). In the combined tree, the placements of the problematic taxa identified above resembled those of the *cox1* tree, rather than the *atp1* tree (Acoraceae are united with *Lilium*, PP = 94; *Oenothera* were united with *Buxus*, PP = 96). Moreover, combining *atp1* and *cox1* increased the tree length

→

Fig. 5. Bayesian inference phylogram from a combined analysis of *atp1* and *cox1* loci for 58 taxa. Branches are depicted as proportional to their calculated mean lengths except in two instances where unusually long branches were truncated and marked (//) to make the figure more readable. Numbers along the branches are PP values > 89% from the Bayesian inference analysis.



100
Acorus calamus
Acorus gramineus
94
94
Lilium lancifolium
Smilax sp.
Asparagus officinalis
Carludovica palmata
Dioscorea macrostachya
100
Joinvillea plicata
Oryza sativa
100
Phoenix dactylifera
100
Alisma plantago-aquatica
Sagittaria latifolia
100
Elodea sp.
100
Potamogeton sp.
100
Arisaema triphyllum
Spathiphyllum clevelandii
Orontium aquaticum
100
Pleea sp.
Tofieldia calyculata
99
Houttuynia cordata
100
Saururus chinensis
100
Piper bicolor
Asarum canadense
Aristolochia elegans
100
Lactoris fernandeziana
Canella winterana
99
Drimys winteri
90
96
100
Asimina triloba
Polyalthia suberosa
91
Eupomatia laurina
97
Liriodendron tulipifera
Magnolia grandiflora
99
Knema latericia
100
Calycanthus floridus
Idiospermum australiense
100
Cinnamomum zeylanicum
99
Laurus nobilis
Hedycarya arborea
97
100
Chloranthus spicatus
Sarcandra sp.
Ceratophyllum demersum
Akebia quinata
100
Euptelea polyandra
Corydalis lutea
100
96
Buxus sempirvirens
Oenothera biennis
Trochodendron aralioides
100
Tetracentron sinense
Grevillea robusta
Platanus occidentalis
Nelumbo sp.
Austrobaileya scandens
Illicium lanceolatum
100
100
100
Kadsura japonica
Schisandra sphenanthera
100
Cabomba sp.
Nymphaea odorata
Amborella trichopoda
0.005 substitutions/site

by 36 steps over the sum of the lengths of the two individual trees because of the increased homoplasy. The homoplasy indices for the MP trees, excluding uninformative characters, were 0.5257 (*atp1*) and 0.4896 (*cox1*). The homoplasy index for the combined analysis was 0.5268, so that combining these loci actually slightly increased homoplasy over the more homoplasious single gene tree.

Trigenomic analysis has been a generally productive approach in phylogenetics (e.g., Qiu et al., 2005). Combining data from different loci is potentially productive in phylogenetics particularly when there are no strongly supported conflicts or other indications of incongruence (Duvall, 2000; Duvall et al., 2006). The rationale for including sequences with low substitution rates, such as mitochondrial genes, is that these will have a lower degree of saturation and will better resolve deep divergences. Mitochondrial coding sequences as similar as *atp1* and *cox1* are not expected to produce gene trees that are fundamentally and pervasively incongruent. However, there is strongly supported conflict between the separate *atp1* and *cox1* gene trees and persistent indications of incongruence so that combining these mitochondrial sequences is clearly counterproductive to the goal of increasing phylogenetic signal at specific deep nodes in the angiosperm phylogeny. We believe that mitochondrial sequences in particular should not be indiscriminately combined with each other or the sequences of other loci in spite of the theoretical appeal of this total evidence approach (Kluge, 1989). Single-gene analyses must first be performed to permit tests of inter-locus phylogenetic incongruence. If incongruence is found, hypotheses should be devised and tested seeking the likely cause, so that appropriate analytical strategies, such as the exclusion of certain subsets of the data, may be devised. However, when the source of the incongruence cannot be identified, such as with *atp1* and *cox1*, perhaps entirely different approaches, such as whole genome analyses (e.g., Hansen et al., 2007; Moore et al., 2007), will prove to be more productive.

## LITERATURE CITED


Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723.

Altschul, S. F., W. Gish, W. Miller, E. Myers, and D. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

APG II [Angiosperm Phylogeny Group]. 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.

Barkman, T. J., G. Chenery, J. R. McNeal, J. Lyons-Weiler, W. J. Ellisens, G. Moore, A. D. Wolfe, and C. W. dePamphilis. 2000. Independent and combined analyses of sequences from all three genomic compartments converge on the root of the flowering plant phylogeny. *Proceedings of the National Academy of Sciences, USA* 97: 13166–13171.

Bergsten, J. 2005. A review of long-branch attraction. *Cladistics* 21: 163–193.

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.

Bergthorsson, U., A. Richardson, G. Young, L. Goertzen, and J. D. Palmer. 2004. Massive horizontal gene transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella. Proceedings of the National Academy of Sciences, USA* 101: 17747–17752.

Bowe, L. M., M. Coat, and C. W. dePamphilis. 2000. Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Sciences, USA* 97: 4092–4097.

Bowe, L. M., and C. W. dePamphilis. 1996. Effects of RNA editing and gene processing on phylogenetic reconstruction. *Molecular Biology and Evolution* 13: 1159–1166.

Cai, Z., C. Penaflor, J. V. Kuehl, J. Leebens-Mack, J. E. Carlson, C. W. DePamphilis, J. L. Boore, and R. K. Jansen. 2006. Complete chloroplast genome sequences of *Drimys, Liriodendron*, and *Piper*: Implications for the phylogeny of magnoliids. *BMC Evolutionary Biology* 6: 77.

Chaw, S.-M., A. Chun-Chieh Shih, D. Wang, Y.-W. Wu, S.-M. Liu, and T.-Y. Chou. 2008. The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. *Molecular Biology and Evolution* 25: 603–615.

Cho, Y., Y.-L. Qiu, P. Kuhlman, and J. D. Palmer. 1998. Explosive invasion of plant mitochondria by a group I intron. *Proceedings of the National Academy of Sciences, USA* 95: 14244–14249.

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

Davis, J. I., D. W. Stevenson, G. Petersen, O. Seberg, L. M. Campbell, J. V. Freudenstein, D. H. Goldman, C. R. Hardy, F. A. Michelangeli, M. P. Simmons, C. D. Specht, F. Vergara-Silva, and M. A. Gandolfo. 2004. A phylogeny of the monocots, as inferred from *rbcL* and *atpA* sequence variation, and a comparison of methods for calculating jackknife and bootstrap values. *Systematic Botany* 29: 467–510.

Duvall, M. R. 2000. Seeking the dicot sister group of the monocots. *In* K. L. Wilson and D. A. Morrison [eds.], Monocots: Systematics and evolution, 25–32. CSIRO Publishing, Collingwood, Victoria, Australia.

Duvall, M. R., M. T. Clegg, M. W. Chase, W. D. Clark, W. J. Kress, H. G. Hills, L. E. Eguiarte, J. F. Smith, B. S. Gaut, E. A. Zimmer, and G. H. Learn Jr. 1993. Phylogenetic hypotheses for the monocotyledons constructed from *rbcL* sequence data. *Annals of the Missouri Botanical Garden* 80: 607–619.

Duvall, M. R., and A. Ervin. 2004. 18S gene trees are positively misleading for monocot/dicot phylogenetics. *Molecular Phylogenetics and Evolution* 30: 97–106.

Duvall, M. R., S. Mathews, N. Mohammad, and T. Russell. 2006. Placing the monocots; conflicting signal from trigenomic analyses. *Aliso* 22: 77–88.

Farris, J. 1989. The retention index and the rescaled consistency index. *Cladistics* 5: 417–419.

Farris, J., M. Källersjö, A. Kluge, and C. Bult. 1994. Testing significance of incongruence. *Cladistics* 10: 315–319.

Felsenstein, J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology* 22: 240–249.

Felsenstein, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution; International Journal of Organic Evolution* 39: 783–791.

Giegé, P., and A. Brennicke. 1999. RNA editing in *Arabidopsis mitochondria* effects 441 C to U changes in ORFs. *Proceedings of the National Academy of Sciences, USA* 96: 15324–15329.

Graur, D., and W.-H. Li. 2000. Fundamentals of molecular evolution, 2nd ed. Sinauer, Sunderland, Massachusetts, USA.

Hajibabaei, M., J. Xia, and G. Drouin. 2006. Seed plant phylogeny: Gnetophytes are derived conifers and a sister group to Pinaceae. *Molecular Phylogenetics and Evolution* 40: 208–217.

Hansen, D., S. Dastidar, Z. Cai, C. Penaflor, J. Kuehl, J. Boore, and R. Jansen. 2007. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Molecular Phylogenetics and Evolution* 45: 547–563.

HIESEL, R., A. VON HAESELER, AND A. BRENNICKE. 1994. Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. *Proceedings of the National Academy of Sciences, USA* 91: 634–638.

HIPP, A., J. HALL, AND K. SYTSMA. 2004. Congruence versus phylogenetic accuracy: Revisiting the incongruence length difference test. *Systematic Biology* 53: 81–89.

HUELSENBECK, J. P. 1997. Is the Felsenstein zone a fly trap? *Systematic Biology* 46: 69–74.

HUELSENBECK, J., AND F. RONQUIST. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* 17: 754–755.

KADOWAKI, K., K. OZAWA, S. KAZAMA, N. KUBO, AND T. AKIHAMA. 1995. Creation of an initiation codon by RNA editing in the *coxl* transcript from tomato mitochondria. *Current Genetics* 28: 415–422.

KLUGE, A. 1989. A concern for evidence and a phylogenetic hypothesis for relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology* 38: 7–25.

KLUGE, A., AND J. FARRIS. 1969. Quantitative phyletics and the evolution of Anurans. *Systematic Zoology* 18: 1–32.

LASER, B., G. OETTLER, AND U. KÜCK. 1995. RNA editing of the mitochondrial *atp*A/*atp*9 co-transcript of Triticale, carrying the *timopheevi* cytoplasmic male sterility cytoplasm from wheat. *Plant Physiology* 107: 663–664.

MOORE, M., C. BELL, P. SOLTIS, AND D. SOLTIS. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences, USA* 104: 19363–19368.

MORT, M. E., P. S. SOLTIS, D. E. SOLTIS, AND M. L. MABRY. 2000. Comparison of three methods for estimating internal support on phylogenetic trees. *Systematic Biology* 49: 160–171.

MOWER, J. 2005. PREP-Mt: Predictive RNA editor for plant mitochondrial genes. *BMC Bioinformatics* 6: 96 Website http://www.biomedcentral.com/1471-2105/6/96.

NIXON, K. C. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* 15: 407–414.

NIXON, K. C., AND J. I. DAVIS. 1991. Polymorphic taxa, missing values and cladistic analysis. *Cladistics* 7: 233–241.

PARKINSON, C., K. ADAMS, AND J. D. PALMER. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants. *Current Biology* 9: 1485–1488.

PETERSEN, G., O. SEBERG, J. I. DAVIS, D. H. GOLDMAN, D. W. STEVENSON, L. M. CAMPBELL, F. A. MICHELANGELI, C. D. SPECHT, M. W. CHASE, M. F. FAY, J. C. PIRES, J. V. FREUDENSTEIN, C. R. HARDY, AND M. P. SIMMONS. 2006a. Mitochondrial data in monocot phylogenetics. *Aliso* 22: 52–62.

PETERSEN, G., O. SEBERG, J. I. DAVIS, AND D. W. STEVENSON. 2006b. RNA editing and phylogenetic reconstruction in two monocot mitochondrial genes. *Taxon* 55: 871–886.

POSADA, D., AND K. A. CRANDALL. 1998. MODELTEST: Testing the model of DNA substitution. *Bioinformatics (Oxford, England)* 14: 817–818.

QIU, Y.-L., O. DOMBROVSKA, J. LEE, L. LI, B. A. WHITLOCK, F. BERNASCONI-QUADRONI, J. S. REST, C. C. DAVIS, T. BORSCH, K. W. HILU, S. S. RENNER, D. E. SOLTIS, P. S. SOLTIS, M. J. ZANIS, J. J. CANNONE, R. R. GUTELL, M. POWELL, V. SAVOLAINEN, L. W. CHATROU, AND M. W. CHASE. 2005. Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. *International Journal of Plant Sciences* 166: 815–842.

QIU, Y.-L., J. LEE, F. BERNASCONI-QUADRONI, D. E. SOLTIS, P. S. SOLTIS, M. J. ZANIS, E. A. ZIMMER, Z. CHEN, V. SAVOLAINEN, AND M. W. CHASE. 2000. Phylogeny of basal angiosperms: Analyses of five genes from three genomes. *International Journal of Plant Sciences* 161: S3–S27.

RAMBAUT, A., AND N. GRASSLY. 1997. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer Applications in the Biosciences* 13: 235–238.

RONQUIST, F., AND J. HUELSENBECK. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)* 19: 1572–1574.

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

SAVOLAINEN, V., AND M. W. CHASE. 2003. A decade of progress in plant molecular phylogenetics. *Trends in Genetics* 19: 717–724.

SCHUSTER, W., R. TERNES, V. KNOOP, R. HIESEL, B. WISSINGER, AND A. BRENNICKE. 1991. Distribution of RNA editing sites in *Oenothera* mitochondrial mRNAs and rRNAs. *Current Genetics* 20: 397–404.

SHIMODAIRA, H., AND M. HASEGAWA. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114–1116.

SHIMODAIRA, H., AND M. HASEGAWA. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 51: 492–508.

SIKES, D. S. AND P. O. LEWIS. 2001. Software manual for PAUPRat: A tool to implement parsimony ratchet searches using PAUP*. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, USA. Website http://users.iab.uaf.edu/~derek_sikes/software.htm

SOLTIS, P. S., D. E. SOLTIS, AND M. W. CHASE. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402: 402–404.

SOLTIS, P. S., D. E. SOLTIS, M. J. ZANIS, AND S. KIM. 2000. Basal lineages of angiosperms: Relationships and implications for floral evolution. *International Journal of Plant Sciences* 161 (6 Supplement): S97–S107.

SWOFFORD, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer, Sunderland, Massachusetts, USA.

TEMPLETON, A. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* 37: 221–244.

THOMPSON, J., D. HIGGINS, AND T. GIBSON. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673–4680.

VINNERSTEN, A., AND K. BREMER. 2001. Age and biogeography of major clades in Liliales. *American Journal of Botany* 88: 1695–1703.

WOLFE, K. H., W.-H. LI, AND P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054–9058.

YANG, Z., AND B. RANNALA. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Molecular Biology and Evolution* 14: 717–724.

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

ZANIS, M. J., P. S. SOLTIS, Y.-L. QIU, E. A. ZIMMER, AND D. E. SOLTIS. 2003. Phylogenetic analyses and perianth evolution in basal angiosperms. *Annals of the Missouri Botanical Garden* 90: 129–150.

ZWICKL, D. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, University of Texas at Austin, Austin, Texas, USA. GARLI version 0.951-1Website http://www.molecularevolution.org/mbl/software/garli/.

APPENDIX 1. Voucher information for taxa sequenced in this study. Vouchers are deposited at DEK unless otherwise indicated.

***Taxon***—*Voucher specimen*, Herbarium other than DEK.

***Acorus americanus*** (Raf.) Raf., *M. Duvall s. n.* ***Acorus gramineus*** Soland., *M. Duvall s. n. 26 May 2006*. ***Alisma plantago-aquatica*** L., *No voucher*. ***Asparagus falcatus*** L., *W. J. Hahn 6881*, WIS. ***Carludovica palmata*** Ruiz & Pavon, *Y.-L. Qiu 97021*, IND. ***Elodea*** sp., *M. Duvall s. n. 26 May 2006*. ***Fritillaria crassifolia*** Boiss. & Huet, *M. Duvall s. n.* ***Hedycarya arborea*** J.R. & G. Forst., *Y.-L. Qiu 90028,* NCU. ***Houttuynia cordata*** Thunb., *Y.-L. Qiu 92016*, NCU. ***Idiospermum australiense*** (Diels) S.T. Blake, *Y.-L. Qiu 91042,* NCU. ***Joinvillea plicata***, *L. Thien 84,* NO. ***Lilium tigrinum*** Ker Gawl. (= ***L. lancifolium*** Thunb.), *M. Duvall s. n. 1 Aug. 2005*. ***Liriodendron tulipifera***, *M. Duvall s. n. DEK000372*. ***Nymphaea odorata*** Aiton, *D. Les s. n.,* CONN. ***Phoenix dactylifera*** L., *M. Duvall s. n. 26 May 2006*. ***Pleea tenuifolia*** Michx., *Y.-L. Qiu 96128*, IND. ***Potamogeton crispus*** L., *Y.-L. Qiu 96086*, IND. ***Sagittaria latifolia*** Willd., *M. Duvall s. n. 13 Aug. 2002*. ***Smilax tamnoides*** L., *M. Duvall s. n. 31 May 2006*. ***Tofieldia calyculata*** Wahlenb., *Y.-L. Qiu 97041*, IND.

APPENDIX 2. GenBank accession numbers of sequences analyzed in this paper. The accession numbers for newly determined sequences begin with "DQ" or "EU." The single accession numbers listed for *Acorus americanus* and *Fritillaria crassifolia* are for *atp1*; there are no banked sequences of *cox1* for these two species.

***Taxon***—GenBank accessions: *atp1, cox1*.

MONOCOTS. ***Acorus americanus,*** EU081863, —. ***Acorus calamus,*** AF197621, AF193944. ***Acorus gramineus,*** AF197622, DQ630729. ***Alisma plantago-aquatica,*** AF197717, DQ630730. ***Arisaema triphyllum*** (L.) Shott, AY009246, AY009454. ***Asparagus* sp.,** AF197713, DQ508957. ***Carludovica palmata,*** AF197707, DQ508954. ***Dioscorea macrostachya,*** AY009417, AY009442. ***Elodea* sp.,** DQ508946, DQ630731. ***Fritillaria crassifolia,*** EU081862, —. ***Lilium lancifolium,*** AY394729, DQ508958. ***Joinvillea plicata,*** AY394728, DQ508950. ***Orontium aquaticum*** L., AF197705, AJ007554. ***Oryza sativa,*** AB076666, BA000029. ***Phoenix dactylifera,*** DQ508947, AY166800. ***Pleea tenuifolia,*** AF197703, DQ508952. ***Potamogeton* sp.,** AF197715, DQ508953. ***Sagittaria latifolia,*** DQ508948, DQ630732. ***Smilax* sp.,** AF039251, DQ508959. ***Spathiphyllum clevelandii,*** AF197706, AJ007553. ***Tofieldia calyculata,*** AF197704, DQ508960.

EARLY ANGIOSPERM GRADE. ***Amborella_trichopoda,*** AY009407, AY009430. ***Austrobaileya scandens,*** AF197664, AY009434. ***Cabomba* sp.,** AF197641, AF193949. ***Illicium lanceolatum*** A. C. Sm., AF209101, AY009445. ***Kadsura japonica*** Juss., AF197661, AF193952. ***Nymphaea odorata,*** AF209102, DQ508951. ***Schisandra sphenanthera*** Rehder & E.H.Wilson, AF197662, AF193951.

CERATOPHYLLACEAE. ***Ceratophyllum demersum,*** AF197627, AY009438.

CHLORANTHALES. ***Chloranthus spicatus*** Makino, AY299746, AY009439. ***Sarcandra grandifolia,*** AF197666, AF193958.

EUDICOTS. ***Akebia quinata*** Decne., AF197642, AY009429. ***Buxus sempervirens*** L., AF197636, AF193962. *Corydalis lutea* DC., AY009416, AY009441. ***Euptelea polyandra*** Sieb. & Zucc., AF197650, AF193963. ***Grevillea robusta*** A. Cunn., AF197712, AY009449. ***Nelumbo* sp.,** AF197654, AY009447. ***Oenothera biennis,*** X04023, AF020571. ***Platanus occidentalis*** L., AF197655, AY009450. ***Tetracentron sinense*** Oliver, AF197647, AF193964. ***Trochodendron aralioides*** Sieb. & Zucc., AF197648, AF020581.

MAGNOLIIDS. ***Asarum canadense*** L., AF197671, AY009432. ***Asimina triloba*** Dun., AF197696, AY009433. ***Aristolochia elegans*** Mast., AY009408, AY009431. ***Calycanthus floridus*** L., AF197678, AY009436. ***Canella winterana*** (L.) Gaertn., AF197676, AY009437. ***Cinnamomum zeylanicum*** Blume, AY009415, AY009440. ***Drimys winteri*** J.R.Forst. & G.Forst., AF197673, AY009443. ***Eupomatia laurina*** Hook., AY299767, AY009444. ***Hedycarya arborea*** J.R.Forst. & G.Forst., AF197689, DQ508949. ***Houttuynia cordata*** Thunb., AF197632, DQ508955. ***Idiospermum australiense*** S. T. Blake, AF197680, DQ508956. ***Knema latericia,*** AF197697, AJ223430. ***Lactoris fernandeziana*** Phil., AF197710, AY009446. ***Laurus nobilis*** L., AF197682, AF193956. ***Liriodendron tulipifera,*** AY394730, DQ980415. ***Magnolia grandiflora*** L., AF209100, AF020568. ***Piper bicolor*** Yunck., AY009421, AY009448. ***Polyalthia suberosa*** (Roxb.) Thwaites, AF197694, AF193957. ***Saururus chinensis*** Hort. ex Loud., AY009424, AY009452.