View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Open Access LMU

Reevaluation of the cox1 Group I Intron in Araceae and Angiosperms Indicates a History Dominated by Loss rather than Horizontal Transfer

Natalie Cusimano, Li-Bing Zhang, and Susanne S. Renner

The origin and modes of transmission of introns remain matters of much debate. Previous studies of the group I intron in the angiosperm cox1 gene inferred frequent angiosperm-to-angiosperm horizontal transmission of the intron from apparent incongruence between intron phylogenies and angiosperm phylogenies, patchy distribution of the intron among angiosperms, and differences between cox1 exonic coconversion tracts (the first 22 nt downstream of where the intron inserted). We analyzed the cox1 gene in 179 angiosperms, 110 of them containing the intron (intron^{+}) and 69 lacking it (intron^{-}). Our taxon sampling in Araceae is especially dense to test hypotheses about vertical and horizontal intron transmission put forward by Cho and Palmer (1999. Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial coxl gene during evolution of the Araceae family. Mol Biol Evol. 16:1155–1165). Maximum likelihood trees of Araceae cox1 introns, and also of all angiosperm cox1 introns, are largely congruent with known phylogenetic relationships in these taxa. The exceptions can be explained by low signal in the intron and long-branch attraction among a few taxa with high mitochondrial substitution rates. Analysis of the 179 coconversion tracts reveals 20 types of tracts (11 of them only found in single species, all involving silent substitutions). The distribution of these tracts on the angiosperm phylogeny shows a common ancestral type, characterizing most intron^{+} and some intron^{-} angiosperms, and several derivative tract types arising from gradual back mutation of the coconverted nucleotides. Molecular clock dating of small intron^{+} and intron^{-} sister clades suggests that coconversion tracts have persisted for 70 Myr in Araceae, whose cox1 sequences evolve comparatively slowly. Sequence similarity among the 110 introns ranges from 91% to identical, whereas putative homologs from fungi are highly different, but sampling in fungi is still sparse. Together, these results suggest that the cox1 intron entered angiosperms once, has largely or entirely been transmitted vertically, and has been lost numerous times, with coconversion tract footprints providing unreliable signal of former intron presence.

Introduction

Most mitochondrial introns are self-splicing ribozymes that act as mobile genetic elements (Goddard and Burt 1999; Chevalier and Stoddard 2001; Haugen et al. 2005), with their mobility depending on enzymes that the introns themselves encode. In group I introns, these enzymes belong to the LAGLI-DADG family and function as homing endonuclease, maturase, or both (Delahodde et al. 1989; Wenzlau et al. 1989; Belfort 2003). Phylogenetic analyses suggest that group I introns in fungi, sponges, algae, and land plants have sometimes been transmitted horizontally (Lang 1984; Lambowitz 1989; Nishida and Sugiyama 1995; Vaughn et al. 1995; Adams et al. 1998; Cho, Adams, et al. 1998; Cho, Qiu, et al. 1998; Cho and Palmer 1999; Rot et al. 2006). A group I intron in the cytochrome c oxidase subunit I (cox1) gene is thought to have been transmitted horizontally as many as a 1,000 times among the 13,500 genera and 300,000 species of extant angiosperms (Cho, Qiu, et al. 1998). This extrapolation was based on a survey of the cox1 intron's distribution among 335 genera of land plants in which the authors inferred 32 separate cases of intron acquisition to account for the intron's presence in 48 of 281 species from 278 genera of flowering plants.

Horizontal transmission of introns has been inferred from 3 kinds of evidence (Cho, Qiu, et al. 1998). First, strong incongruence between an intron phylogeny and that of angiosperms suggests independent gains rather than vertical transmission. Second, patchy distribution of an intron on an angiosperm phylogeny and especially the nesting of intron-containing (intron^{+}) species within large clades of intronless (intron^{-}) species point to horizontal acquisition. Third, coconversion tract analysis can provide information on whether a particular site gained or lost an intron (Bell-Pedersen et al. 1989; Adams et al. 1998; Cho and Palmer 1999). Coconversion tracts are short stretches of flanking exon sequence that are converted to the donor DNA sequence. This is because group I introns transfer by way of a recombination/repair process initiated by a staggered double-strand break catalyzed by the intron's homing endonuclease at a target site in the recipient (Szostak et al. 1983; Lambowitz and Belfort 1993; Belfort and Perlman 1995). The cleaved DNA strands of the recipient DNA are partially degraded, creating a gap that is filled in using the donor DNA as the template. If the flanking exon stretches in the donor and recipient differ, then coconversion will create a "footprint" that can stay even after the intron itself is lost again (Cho and Palmer 1999).

Much trust has been placed in coconversion tracts as historical evidence of intron presence. For example, an investigation of the evolutionary history of the cox1 intron in the Araceae (Cho and Palmer 1999) relied on the exonic coconversion tracts in the intron^{+} species, coupled with the absence of any deletion footprints in the intron^{-} species, to infer 3–5 intron gains via horizontal transfers. Reliance on the coconversion tracts here overrode the implication of a parsimony reconstruction, which would have been consistent with a vertical transmission history in Araceae, with 1 gain, followed by 2 losses (Cho and Palmer 1999). Cho and Palmer (1999) also noticed that *Arisaema triphyllum* and *Pistia stratiotes* had identical coconversion tracts and grouped together in the intron phylogeny, suggesting that these 2 introns might be vertically inherited.

To test the hypotheses of Cho and Palmer (1999), namely that the cox1 intron has been transferred horizontally in much of the Araceae family, but vertically in the

Key words: group I intron, mitochondrial genome, cox1 gene, horizontal gene transfer, coconversion, angiosperms.

Arisaema/Pistia clade, we analyzed the *cox1* gene in a dense sample of relevant Araceae, using available multigene phylogenetic frameworks (Renner and Zhang 2004; this study). Surprisingly, with the larger taxon sample employed here, the distribution of the *cox1* intron in Araceae is more parsimoniously explained by ancestral presence, followed by independent losses, than by horizontal gene transfers.

Araceae are an early branching lineage of flowering plants, and we therefore decided to investigate the distribution of the *cox1* intron among early and more recent lineages of angiosperms based on all angiosperm *cox1* sequences available in GenBank (plus new sequences generated in the course of this study). Comparison of the much larger intron phylogeny with the angiosperm phylogeny, the great sequence similarity among angiosperm *cox1* introns, the clustered distribution of exonic coconversion tract types, and hierarchical patterns of decay in the coconversion tracts suggest ancestral presence of the *cox1* intron, followed by numerous losses. This implies that the signal in coconversion tracts (the footprint) is less reliable than previously thought. To infer a temporal framework for *cox1* intron turnover and the loss of coconversion tracts, we estimated maximal times over which the intron could have been gained or lost in Araceae, using angiosperm clade ages as a proxy for intron maximal ages. We also evaluate the hypothesis of Seif et al. (2005) that the *cox1* intron in angiosperms originated in a fungus close to *Rhizopus oryzae*.

Materials and Methods

Taxon Sampling and Sequencing

Taxa selected for this study with their sources and herbarium vouchers (where applicable) are listed in supplementary table S1 (Supplementary Material online), which includes 179 angiosperms, 110 of them intron^{b} and 69 intron^{-}. To deduce the evolutionary history of the *cox1* intron in the *Arisaema/Pistia* clade, we relied on 2 chloroplast loci (the *trnL* intron and adjacent spacer before the *trnF* gene and the *rpl20-5-rps12* intergenic spacer) and one mitochondrial locus (parts of exons b and c of *nad1* and the complete intron between them). We included 30 species of Araceae, many available from Renner and Zhang (2004). Newly generated sequences were produced with the same primers and polymerase chain reaction (PCR) conditions as used in that study. The *cox1* exon and intron (where present) were sequenced for 36 Araceae, and in all, this study includes 56 newly generated sequences (36 *cox1* genes and 20 of other loci).

Total DNAs of silica-dried material were extracted with the NucleoSpin plant kit according to the manufacturer's protocol (Macherey-Nagel, Düren, Germany), and the complete *cox1* gene was directly amplified with the primer pair 82F (5#-GGAGTGATGGGCACAT GCT-TCT-3#) and *cox1*1.6KR int (5#-AAGGCTGGAGGGC-TTTGT AC-3#). PCRs were performed with 10 mM primers in 25 1l reactions using BioTherm DNA polymerase (Genecraft, Lüdinghausen, Germany). The initial step of 5 min at 95 °C was followed by 35 cycles of 95 °C for 30 s for DNA denaturation, 60 °C for 60 s for primer annealing, and 72 °C for 2 min and 40 s for primer extension. PCR products were controlled by electrophoresis on an ethidiumbromide stained 1% agarose gel with a 1 Kb Plus DNA ladder (Invitrogen, Karlsruhe, Germany). The amplified fragment was approximately 2,340 nt long for intron^{b} and about 1,500 nt long for intron^{-} taxa. Products were purified and quantified electrophoretically using Lambda DNA as standard. If multiple bands were detected, an additional electrophoresis was performed to excise and analyze them separately. Sequencing relied on Big Dye Terminator kits (Applied Biosystems, Warrington, UK) and the following primers (in different combinations depending on the length of the sequences obtained, varying from 400 to 1,000 nt): 42F (5#-GGATCTTCTCCA CTAACCACAAA-3#), 82F (see above), 657R (5#-GCG-GGATCAGAAAAGGTTGTA-3#), IP53 (5#-GGAGCAG-TTGATTTAGC-3#), I589R (5#-GGTAGTCGATGCTT-CATAGC-3#), I361F (5#-GTATTAAAATGCGATCAGG-TGC-3#), I557F (5#-AGGATTCTTTGATGCTGAGGG-3#), I942R (5#-GGATGAATAGAAGAAAGGT-3#), Int1.2KF (5#-AGCATGGCTAGCTTTCCTAGA-3#), 855F (5#-TGG-ATTTCTTGTTTGGGCTCAT-3#), IP56 (5#-GAGCAA-TGTCTAGCC C-3#), 1150F (5#-TCTATGGGAGCCGT-TTTTGC-3#), and *cox1*.6KR (see above). The cycle sequencing products were cleaned by Sephadex G-50 Superfine gel filtration (Amersham, Uppsala, Sweden) on MultiScreen TM-HV membrane plates (Millipore, Bedford, MA) according to the manufacturers' protocols to remove unincorporated nucleotides. Fragments were separated on an ABI 3100 Avant capillary sequencer, assembled and edited using the software Sequencher (Gene Codes, Ann Arbor, MI), and Blast searched in GenBank.

Alignments and Phylogenetic Analyses

Alignments were generated manually in MacClade (Maddison WP and Maddison DR 2003) and adjusted by eye; all have been submitted to TreeBase. Amplification of the *cox1* exon of *Theriophonum dalzelii* failed, and the missing sequence for this species was coded with question marks. We analyzed 4 data matrices. The first comprised the 4 chloroplast and mitochondrial loci sequenced for the *Arisaema/Pistia* clade. The second consisted of 149 angiosperm *cox1* exon sequences including 11 newly sequenced Araceae species and 12 Araceae from Cho, Qiu, et al. (1998). The third consisted of 106 angiosperm *cox1* intron sequences of which 38 were Araceae. The fourth matrix comprised the coconversion tracts of 179 angiosperms, 110 of them intron^{b} and 69 intron^{-}. To assess the phylogenetic signal in the cox1 gene and introns, we used the molecular phylogeny of angiosperms published by Qiu et al. (2005), the angiosperm phylogenetics database of Stevens (2001), and an unpublished phylogeny of Araceae provided by Mayo S (personal communication).

DNA indels or missing data in the *cox1* intron and exon were excluded from phylogenetic analyses. Phylogenetic inference relied on maximum likelihood (ML) searches as implemented in RAxML-VI-HPC version 2.2.3 (Stamatakis 2006). Bayesian analysis relied on MrBayes version 3.1.2 (Ronquist and Huelsenbeck

2003). The best-fitting model for the combined chloroplast and mitochondrial data (4,682 characters, excluding the intron and 22-bp coconversion tract of the *cox1* gene) identified by Modeltest version 3.7 (Posada and Crandall 1998) was the General-Time-Reversible Model (GTR) þ I þ C model whether by hierarchical likelihood ratio testing or the Akaike information criterion. We therefore used this model in Bayesian analyses, whereas ML analyses relied on the GTR þ C model, this being the only model implemented in RAxML. The best-fitting model for the *cox1* intron matrix was the Transversion þ I þ C model (5 substitution types). As the number of substitution types in MrBayes can only be set to 1, 2, or 6, we used the GTR þ I þ C model. Bayesian runs were started from independent random starting trees and repeated at least twice. Markov chain Monte Carlo runs extended for 1 million generations, with trees sampled every 100 generations. We used a flat Dirichlet prior for the relative nucleotide frequencies and rate parameters, a discrete uniform prior for topologies, and an exponential distribution (mean 1.0) for the gamma shape parameter and all branch lengths. Convergence was assessed in several ways: by checking that final likelihoods and majority rule topologies in different runs were similar, that the standard deviations (SD) of split frequencies were ¸ 0.01, that the log probabilities of the data given the parameter values fluctuated within narrow limits, that the convergence diagnostic (the potential scale reduction factor given by MrBayes) approached 1, and by examining the plot provided by MrBayes of the generation number versus the log probability of the data. Trees saved prior to convergence were discarded as burn-in (2,000–5,000 trees), and a consensus tree was constructed from the remaining trees.

Bootstrapping under ML used 1,000 replicates performed in RAxML, with the initial rearrangement settings and the number of categories tested following the manual. Resulting bootstrap values as well as Bayesian posterior probabilities were plotted on the ML tree using the APE package (Paradis et al. 2004) in R (R Development Core Team 2006).

The *cox1* exon data were analyzed under parsimony in PAUP version 4.0b10 (Swofford 2002). Searches were heuristic, using 100 random taxon addition replicates, tree-bisection-reconnection swapping, with the "multiple trees" and the "steepest descent" options in effect. Starting trees were obtained by stepwise addition; the trees in memory were limited to 100.

Coconversion Tract Analysis

Coconversion tracts, that is, the first 22 nt downstream of the intron insertion site, in 179 angiosperms were compared with the exonic tract of an intron$^-$ Araceae, namely *Orontium aquaticum*, following Cho and Palmer (1999). For convenience, the *O. aquaticum* tract type is henceforth referred to as the unaltered, or 0, tract type, without this implying that it is an ancestral condition. The remaining tracts were categorized relative to the *O. aquaticum* type according to the number of alterations present in their 3rd to 18th position (whether 1, 2, 3, ... 6 differences, all in the third position and all silent), presence or absence of a T in the 20th position (silent: C – U RNA editing), and presence or absence of an A in the 21st position (silent). A coconversion tract that comprises 6 nt differences compared with *O. aquaticum*, a T in position 20 and an A in position 21 is thus referred to as "6 þ T þ A." Similarly, "4 þ T" refers to a coconversion tract with 4 substitutions and an T in position 20.

Divergence Time Estimation

For divergence time estimation, we relied on the combined chloroplast and mitochondrial data (4,682 nt for 30 taxa) and the Bayesian relaxed clock approach implemented in "multidivtime" (Thorne et al. 1998; Thorne and Kishino 2002). After calculating substitution model parameters for the DNA data under the F84 þ C model (with 5 rate categories) on the ML topology obtained from the combined data, rooted on *Xanthosoma sagittifolium* and *Caladium bicolor*, we used Thorne's "estbranches" program to estimate branch lengths and their variance, given the specified evolutionary tree and model parameters. The a priori expected number of time units between the root and the tips was set to 0.9, with a SD of 0.5; the prior on the mean root rate was set to 0.0128, by dividing the median distance from the ingroup root to the tips by the time unit. Thorne's manual recommends that the prior for brown mean (and its SD) be set at values that, when multiplied by the approximate time from the root to the present, yield a value between 1 and 2, and we therefore set brown mean to 1.11. The Markov chain length was 1 million cycles, sampled every 100th cycle and with a burn-in of 100,000 cycles; analyses were repeated at least twice.

To obtain absolute times from genetic distances, we used the following constraints: (1) Peltandreae are first known from 60-Myr leaves from Europe, Kazakhstan, North Dakota, and Tennessee (Wilde et al. 2005). This provides a minimal age of 60 Myr for node 1 in figure 1A. (2) *Protarum sechellarum* is endemic to the Seychelles, and the age of this archipelago (Braithwaite 1984) thus provides a maximal age of 85 Myr for node 2. (3) Middle Eocene leaf impressions (*Caladiosoma messelense*; Wilde et al. 2005) that closely match modern Colocasieae provide a minimal age of 45 Myr for node 3 in figure 1A. (4) The oldest fossils of Araceae are 110- to 120-Myr old (Friis et al. 2004), and therefore, 120 Myr was used as a maximal age for the root node. The earliest angiosperms fossils are 141- to 132-Myr old (Hughes 1994).

Results

The *cox1* Intron and Exonic Coconversion Tracts in the *Arisaema/Pistia* Clade

The distribution of the *cox1* intron in the *Arisaema/Pistia* clade is shown in figure 1A (including the relevant outgroups). As predicted by the hypothesis of Cho and Palmer (1999) that the intron might be vertically inherited in this clade, most species are intronþ and have the same coconversion tract. This tract, namely the 6 þ T type, comprises 6 nt differences compared with intron$^-$ species and a T in position 20.

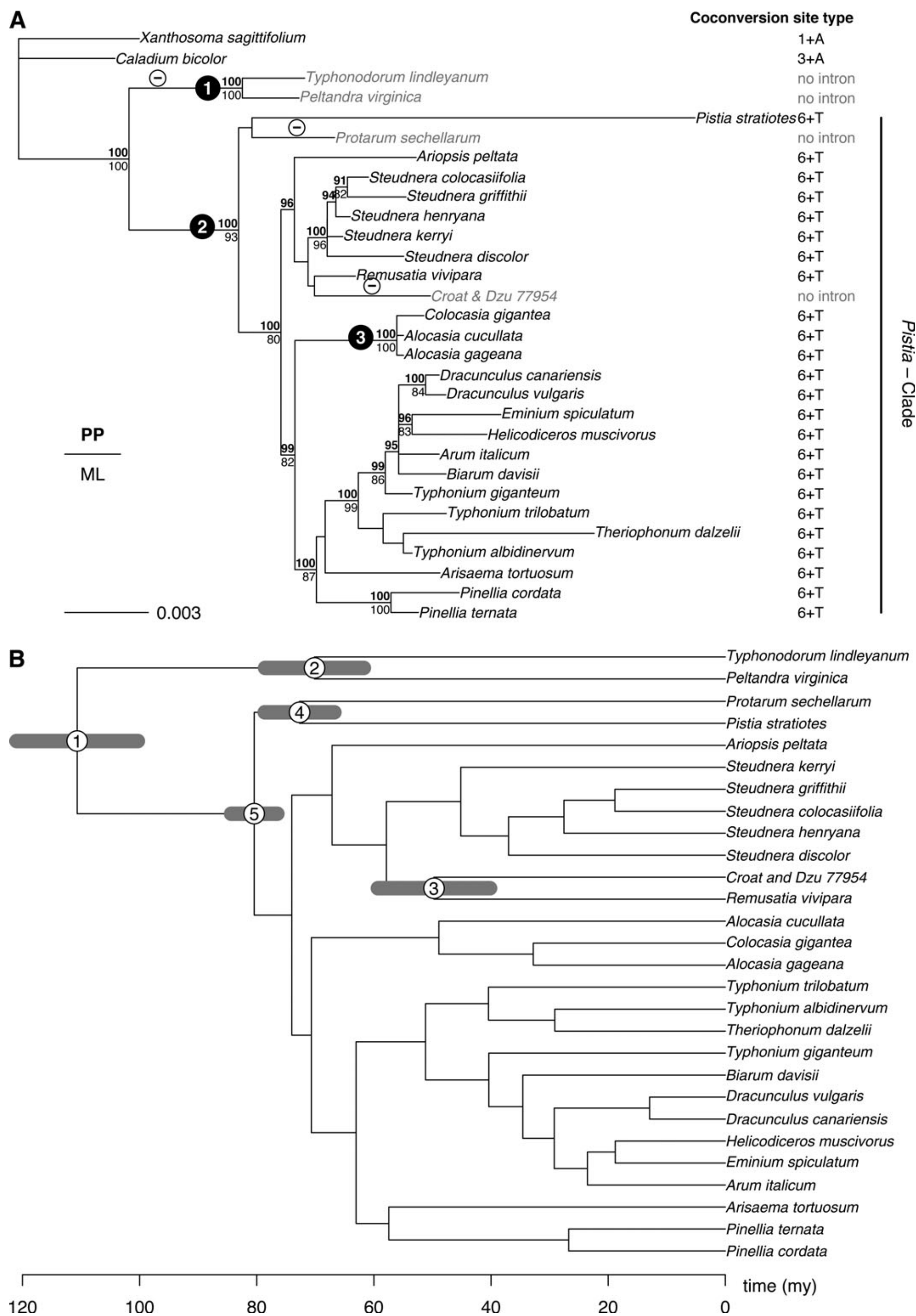

FIG. 1.—(A) ML tree for the Arisaema/Pistia clade based on chloroplast and mitochondrial sequences (4,360 bp) analyzed under a GTR þ I þ C model. Values above branches indicate posterior probabilities ≥ 90, values below branches ML bootstrap values ≥ 75. Numbered nodes (black) were constrained as described in the text. Coconversion tract types (see fig. 2A) of the respective taxa are given on the right. Three inferred intron loss events are marked by a circle-enclosed hyphen. (B) Chronogram for the Arisaema/Pistia clade obtained under a Bayesian relaxed clock applied to the same data and constrained as shown in figure 1A. Nodes 1–5 are discussed in the text, the gray bars indicate SDs around estimates.

The cox1 intron is lacking in Typhonodorum lindleyanum and Peltandra virginica, which form a clade, in P. sechellarum, and in the Vietnamese species Croat and Dzu 77954. For genera with more than one species, we checked at least one additional congeneric for the intron and the coconversion tract, and they all showed the same pattern (figs. 3 and 4). Of the intron^{-} species, 3 have unaltered coconversion tracts, whereas Croat and Dzu 77954 has an A, instead of a C, at position 21 of

A

Int | 5' - Exon - 3' -->

		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29	Coconver-sion site type	Number of species	% species without intron
Orontium aquaticum	-	C A T C C A G A G G T G T A T A T T C C C A T T C T G C C	0	**54**	78,3
Aristolochia elegans	-	C A C C C A G A G G T G T A T A T T C C C A T T C T G C C	1	**1**	1,4
Euptelea polyandra	-	C A T C C A G A A G T G T A T A T T C C C A T T C T G C C	X	**1**	1,4
Euptelea polyandra	-	C A T C C A G A G G T T T A T A T T C C C A T T C T G C C	X	**1**	1,4
Zea mays	-	C A T C C A G A G G T G T A T A T T C T C A T T C T G C C	+T	**7**	10,1
Steudnera colocasiifolia	-	C A T C C A G A G G T G T A T A T T C C A A T T C T G C C	+A	**1**	1,4
Canella winterana	-	C A T C C A G A G G T C T A T A T T C T C A T T C T G C C	X*+T	**1**	1,4
Crossosoma bigelovii	-	C A T C C G A A G G T C T A T A T T C T A A T T C T G C C	X*+T+A	**1**	1,4
Plantago sericea	-	C A C C C T G A A G T C T A T A T T C T C A T T C T G C C	X*+T	**1**	1,4
Asimina triloba	-	C A C C C T G A A G T T T A C A T C C T C A T T C T G C C	6+T	**1**	1,4
			Species without intron:	**69**	
Proteins		HIS | PRO | GLU | VAL | TYR | ILE | PRO			
					% species with intron
Rhamnus cathartica	+	C A T C C A G A G G T G T A T A T T C C C A T T C T G C C	0	**2**	1,8
Barringtonia asiatica	+	C A C C C A G A G G T G T A T A T T C C C A T T C T G C C	1	**1**	0,9
Xanthosoma mafaffa	+	C A C C C A G A G G T G T A T A T T C C A A T T C T G C C	1+A	**2**	1,8
Saranthe sp.	+	C A C C C T G A G G T G T A T A T T C C C A T T C T G C C	2	**1**	0,9
Philodendron hederaceum	+	C A C C C T G A A G T G T A T A T T C C C A T T C T G C C	3	**2**	1,8
Caladium bicolor	+	C A C C C T G A A G T G T A T A T T C C A A T T C T G C C	3+A	**1**	0,9
Plantago cynops	+	C A C C C T G A A G T G T A T A T T C T C A T T C T G C C	3+T	**4**	3,6
Plantago coronopus	+	C A C C C A G A A G T T T A T A T T C T C A T T C T G C C	X+T	**1**	0,9
Heliotropium arborescens	+	C A C C C T G A A G T T T A T A T T C T C A T T C T G C C	4+T	**16**	14,5
Justicia americana	+	C A T C C T G A A G T T T A C A T C C T A A T T C T G C C	X+T+A	**1**	0,9
Hevea brasiliensis	+	C A C C C T G A A G T T T A C A T T C T A A T T C T G C C	5+T+A	**3**	2,7
Cucumis sativus	+	C A C C C T G A A G T T T A C A T C C T A A T T C T G C C	6+T+A	**10**	9,1
Amorphophallus rivieri	+	C A C C C T G A A G T T T A C A T C C T C A T T C T G C C	6+T	**66**	60,0
			Species with intron:	**110**	

B

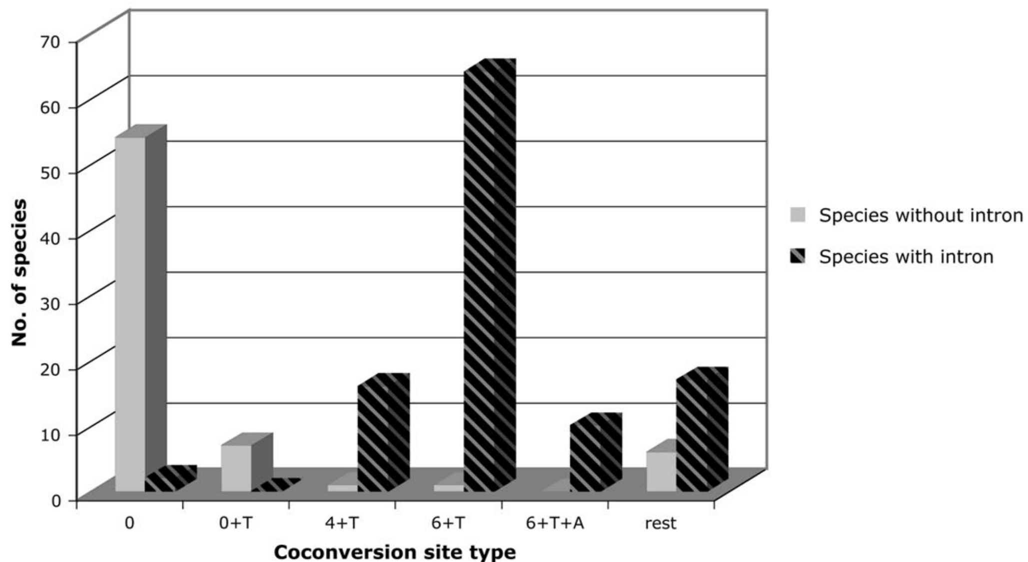

FIG. 2.—Coconversion tract types of intron$^{−}$ and intron^{þ} angiosperm cox1 exons. (A) Top panel: patterns of substitutions in the 22 nt downstream from the 3# end of the intron insertion site in 10 species selected to represent the tract types found in 69 intron$^{−}$ species. Bottom panel: tracts in 13 species selected to represent the tract types found in 110 intron^{þ} species. Most common patterns are given a specific name; the others are labeled as X. An asterisk following the x denotes substitutions in the third position that are different from the common nucleotides. Protein translation is given below the top panel. (B) Histogram of the number of intron^{þ} and intron$^{−}$ species possessing a certain type of coconversion tract. Only the 5 commonest types are explicitly shown (0, 0 þ T, 4 þ T, 6 þ T, and 6 þ T þ A), the remained are subsumed under “rest”.

its tract (fig. S1), which may be circumstantial evidence that it once had an intron in its cox1 gene (below). Of the outgroups, *C. bicolor* has the 3 þ A tract type and *X. sagittifolium* and *X. mafaffa* the 1 þ A type (figs. 1A and 2A and supplementary fig. S1 [Supplementary Material online]).

Cox1 Exonic Coconversion Tracts throughout Angiosperms

Analysis of all available angiosperm cox1 sequences (GenBank, 1 March 2007) revealed 20 coconversion tract

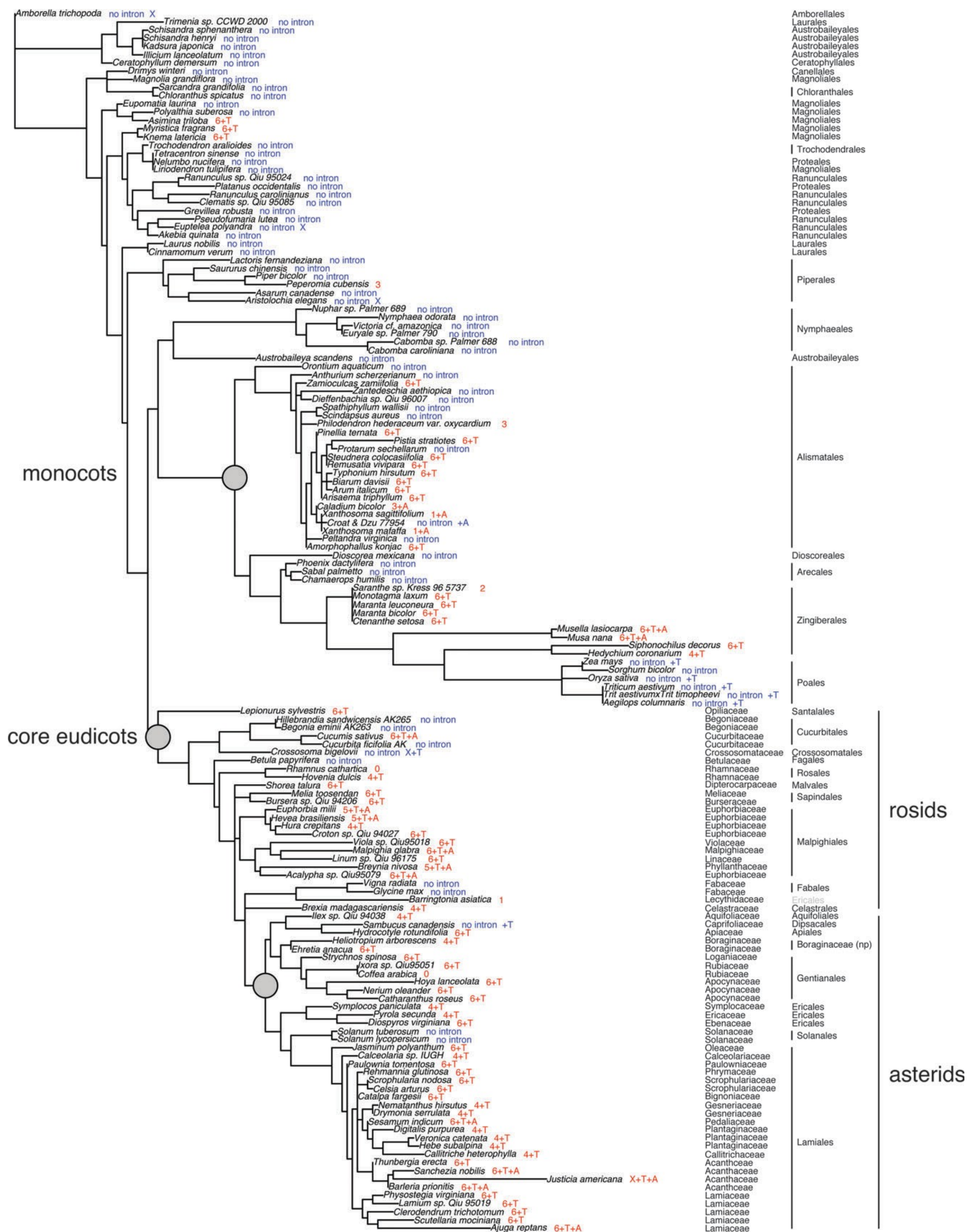

FIG. 3.—Parsimony tree for 148 angiosperms obtained from cox1 exon sequences (1,288 characters, 244 parsimony informative), rooted on *Amborella trichopoda*. Intron absence (blue), intron presence (red), coconversion tract type, family, and order are shown to the right of each species. Orders with a vertical line before their name were recovered as monophyletic.

Taxon	Coconversion site type	Family	Order
Knema latericia	6+T	Myristicaceae	Magnoliales
Myristica fragrans	6+T	Myristicaceae	Magnoliales
Ehretia anacua	6+T	Boraginaceae	not placed
Bursera sp. Qiu 94206	6+T	Burseraceae	Sapindales
Lepionurus sylvestris	6+T	Opiliaceae	Santalales
Melia toosendan	6+T	Meliaceae	Sapindales
Croton sp. Qiu 94027	6+T	Euphorbiaceae	Malpighiales
Jasminum polyanthum	6+T	Oleaceae	Lamiales
Musella lasiocarpa	6+T+A	Musaceae	Zingiberales
Musa nana	6+T+A		Zingiberales
Malpighia glabra	6+T+A	Malpighiaceae	Malpighiales
Diospyros virginiana	6+T	Ebenaceae	Ericales
Hevea brasiliensis	5+T+A	Euphorbiaceae	Malpighiales
Heliotropium arborescens	4+T	Boraginaceae	not placed
Hura crepitans	4+T	Euphorbiaceae	Malpighiales
Calceolaria sp. IUGH	4+T	Calceolariaceae	Lamiales
Hovenia dulcis	4+T	Rhamnaceae	Rosales
Rhamnus cathartica	0		Rosales
Barringtonia asiatica	1	Lecythidaceae	Ericales
Ilex sp. Qiu 94038	4+T	Aquifoliaceae	Aquifoliales
Symplocos paniculata	4+T	Symplocaceae	Ericales
Pyrola secunda	4+T	Ericaceae	Ericales
Pilea fontana	4+T	Urticaceae	Rosales
Hebe subalpina	4+T	Plantaginaceae	Lamiales
Veronica catenata	4+T		Lamiales
Digitalis purpurea	4+T		Lamiales
Callitriche heterophylla	4+T	Callitrichaceae	Lamiales
Nematanthus hirsutus	4+T	Gesneriaceae	Lamiales
Drymonia serrulata	4+T		Lamiales
Brexia madagascariensis	4+T	Celastraceae	Celastrales
Hedychium coronarium	4+T	Zingiberaceae	Zingiberales
Euphorbia milii	5+T+A	Euphorbiaceae	Malpighiales
Breynia nivosa	5+T+A	Phyllanthaceae	Malpighiales
Sanchezia nobilis	6+T+A	Acanthaceae	Lamiales
Justicia americana	X+T+A		Lamiales
Barleria prionitis	6+T+A		Lamiales
Thunbergia erecta	6+T		Lamiales
Linum sp. Qiu 96175	6+T	Linaceae	Malpighiales
Siphonochilus decorus	6+T	Zingiberaceae	Zingiberales
Cucumis sativus	6+T+A	Cucurbitaceae	Cucurbitales
Acalypha sp. Qiu95079	6+T+A	Euphorbiaceae	Malpighiales
Sesamum indicum	6+T+A	Pedaliaceae	Lamiales
Strychnos spinosa	6+T	Loganiaceae	Gentianales
Catharanthus roseus	6+T	Apocynaceae	Gentianales
Nerium oleander	6+T	Apocynaceae	Gentianales
Hoya lanceolata	6+T	Asclepiadaceae	Gentianales
Coffea arabica	0	Rubiaceae	Gentianales
Ixora sp. Qiu95051	6+T	Rubiaceae	Gentianales
Paulownia tomentosa	6+T	Paulowniaceae	Lamiales
Lamium sp. Qiu 95019	6+T	Lamiaceae	Lamiales
Scutellaria mociniana	6+T		Lamiales
Clerodendrum trichotomum	6+T		Lamiales
Physostegia virginiana	6+T		Lamiales
Ajuga reptans	6+T+A		Lamiales
Scrophularia nodosa	6+T	Scrophulariaceae	Lamiales
Celsia arturus	6+T		Lamiales
Rehmannia glutinosa	6+T	Phrymaceae	Lamiales
Catalpa fargesii	6+T	Bignoniaceae	Lamiales
Monotagma laxum	6+T	Marantaceae	Zingiberales
Saranthe sp. Kress 96 5737	2		Zingiberales
Maranta bicolor	6+T		Zingiberales
Maranta leuconeura	6+T		Zingiberales
Ctenanthe setosa	6+T		Zingiberales
Hydrocotyle rotundifolia	6+T	Apiaceae	Apiales
Viola sp. Qiu95018	6+T	Violaceae	Malpighiales
Shorea talura	6+T	Dipterocarpaceae	Malvales
Typhonium trilobatum	6+T	Araceae	Alismatales
Typhonium albidinervum	6+T		Alismatales
Alocasia gageana	6+T		Alismatales
Colocasia gigantea	6+T		Alismatales
Alocasia cucullata	6+T		Alismatales
Typhonium giganteum	6+T		Alismatales
Arisaema tortuosum	6+T		Alismatales
Arisaema triphyllum	6+T		Alismatales
Arisaema speciosum	6+T		Alismatales
Arum dioscoridis	6+T		Alismatales
Pistia stratiotes	6+T		Alismatales
Arum concinnatum	6+T		Alismatales
Arum italicum	6+T		Alismatales
Biarum davisii	6+T		Alismatales
Dracunculus canariensis	6+T		Alismatales
Helicodiceros muscivorus	6+T		Alismatales
Dracunculus vulgaris	6+T		Alismatales
Biarum tenuifolium	6+T		Alismatales
Theriophonum infaustum	6+T		Alismatales
Eminium spiculatum	6+T		Alismatales
Typhonium hirsutum	6+T		Alismatales
Typhonium venosum	6+T		Alismatales
Pinellia ternata	6+T		Alismatales
Pinellia cordata	6+T		Alismatales
Ariopsis peltata	6+T		Alismatales
Steudnera kerryi	6+T		Alismatales
Steudnera henryana	6+T		Alismatales
Remusatia vivipara	6+T		Alismatales
Steudnera griffithii	6+T		Alismatales
Colocasia esculenta	6+T		Alismatales
Steudnera colocasiifolia	6+T		Alismatales
Steudnera discolor	6+T		Alismatales
Amorphophallus konjac	6+T		Alismatales
Zamioculcas zamiifolia	6+T		Alismatales
Philodendron hederaceum var. oxycardium	3		Alismatales
Xanthosoma mafaffa	1+A		Alismatales
Xanthosoma saggitifolium	1+A		Alismatales
Caladium bicolor	3+A		Alismatales
Peperomia cubensis	3	Piperaceae	Piperales
Peperomia polybotrya	3+T		Piperales

FIG. 4.—ML tree for 106 angiosperms based on mitochondrial cox1 intron sequences (967 bp) analyzed under a TvM + I + C model. Values above branches indicate posterior probabilities $\geq$ 90, values below branches ML bootstrap values $\geq$ 75. Coconversion tract types (see fig. 2A) of the respective taxa are given on the right. Families with a vertical line before their name were recovered as monophyletic. (A–D) are clades of interest discussed in the text.

types of which 11 are only found in single species. Figure 2A summarizes the tract types of the 110 intron^{þ} and the 69 intron^{-} *cox1* sequences. Overall, 112 tracts are of the þT type (101 in intron^{þ} *cox1* genes and 11 in intron^{-} genes) and 19 are of the þA type (17 intron^{þ} and 2 intron^{-}). In 15, the A occurs together with the T, whereas in 4 cases, all in the Araceae, the A occurs without the T (one of them in an intron^{-} species and 3 in intron^{þ} species, see figs. 2A and 3). Depending on the intron^{-} angiosperm used for comparison, the C - T transversion in position 20 will be counted as part of the coconversion tract or not. Thus, comparison with *Zea mays* results in an apparent coconversion tract with 6 differences (Cho, Qiu, et al. 1998), whereas comparison with *O. aquaticum* suggests a coconversion tract with 7 differences (Cho and Palmer 1999; this study).

Of the intron^{þ} angiosperms, the majority (60%) has the 6 þ T tract type (34 of these are Araceae), whereas 9% (10 species) have the 6 þ T þ A tract type (fig. 2B). In other words, more than two-thirds of the intron^{þ} angiosperms have changes in all 6 third positions available in the conconversion tract changed (compared with intron^{-} species). Five species (*Breynia nivosa*, *Hevea brasiliensis*, *Justicia americana*, *Pilea fontana*, and *Plantago coronopus*) are of the *x* types, which appear to have undergone a back mutation in the middle of a 6 þ T coconversion tract (fig. 2A). Sixteen are of the 4 þ T type, 4 of the 3 þ T type, and the remaining ones belong to rare types, such as 1 þ A. Two of the 110 intron^{þ} angiosperms (*Coffea arabica* and *Rhamnus* sp.) exhibit no differences in their coconversion tract.

Of the intron^{-} angiosperms, 78% have unaltered tracts like *O. aquaticum* (fig. 2), whereas 22% (15) show differences in their coconversion tract, that is, have a deletion footprint. Of the footprints, one is of the 6 þ T type, one of the 4 þ T type, 7 have only the T in tract position 20, and 4 have single substitutions in different positions. Two species (*Canella winterana* and *Crossosoma bigelovii*) have coconversion tracts that apparently underwent back mutations from 6 þ T and 6 þ T þ A type tracts, and 3 are of the X* types (fig. 2A), that is, have a G instead of an A or a T in tract position 6 and/or a C instead of a G or a T in tract position 12. So far, *Plantago* is the only genus with both intron^{-} and intron^{þ} species, and it also shows particularly variable coconversion tracts (fig. 2A, supplementary fig. S1 [Supplementary Material online]).

Phylogenetic Analyses of Angiosperm *cox1* Exon and Intron Sequences

The hypothesis that the *cox1* intron in angiosperms was gained by multiple lateral transfers predicts incongruence between the angiosperm phylogeny and the intron phylogeny. By contrast, congruence among angiosperm and intron phylogenies points to vertical inheritance. We therefore performed phylogenetic analyses of *cox1* exons and introns and compared them with angiosperm relationships inferred from larger data sets (Stevens 2001; Qiu et al. 2005).

A phylogeny based on *cox1* exon sequences (149 angiosperm species, 1,288 characters, excluding the 22 nt of the coconversion tract), rooted on *Amborella*, is shown in figure 3. It recovers the monocots, core eudicos, and ordinal relationships in agreement with angiosperm phylogenies based on other data (Stevens 2001; Qiu et al. 2005). Generic groupings within more densely sampled families such as Araceae (23 *cox1* sequences) agree with relationships obtained in larger data sets. For example, *Orontium* places as the first diverging Araceae, followed by *Anthurium* and *Zamioculcas*. *Dieffenbachia* and *Zantedeschia*, as well as *Spathiphyllum* and *Scindapus* form sister pairs. The *Arisaema*/*Pistia* clade and its relatives also group together. All this fits with a molecular phylogeny of Araceae (Mayo S, personal communication). The earliest diverging angiosperm lineages containing the *cox1* intron are the Magnoliales (*Asimina*, *Knema*, and *Myristica*) and Piperales (*Peperomia*).

A phylogeny based on *cox1* intron sequences (106 species, 967 characters, excluding 173 gapped positions) and rooted on Myristiaceae (Magnoliales) is shown in figure 4. Except for 10 species that form a basal grade (*Erethia*, *Bursera*, *Lepionurus*, *Melia*, *Croton*, *Jasminum*, *Musella*, and *Musa*), 4 large clades are apparent (labeled A–D in fig. 4): Clade A includes all Araceae plus the 2 *Peperomia* species (probably reflecting long-branch attraction). Within clade A, species with a 6 þ T coconversion tract cluster together. Clade B includes the remaining monocots as well as all eudicots. Clade C includes only species with a 4 þ T coconversion tract, except for *Rhamnus* (no substitution in its coconversion tract) and *Barringtonia* (coconversion tract with 1 substitution). The closest relatives of clade C (not statistically supported) also have 4 þ T tracts or 5 þ T þ A tracts. Clade D, finally, unites 22 species with mainly 6 þ T (one with 6 þ T þ A) coconversion tracts, again with 2 exceptions: *Coffea* (no substitution in its coconversion tract) and *Saranthe* (2 substitutions). Natural groups recovered within clade D are Lamiales (all Lamiaceae and Scrophulariaceae, among others), Gentianales, and Zingiberales. Members of 9 families form well-supported clades, namely Acanthaceae, Araceae, Gesneriaceae, Marantaceae, Musaceae, Piperaceae, Plantaginaceae, Rhamnaceae, and Scrophulariaceae.

Sequence similarity among the 110 *cox1* introns of the angiosperms ranges from 91% to identical (GenBank maximal identities with Blast values of zero in each case), whereas the genetically closest nonangiosperm *cox1* introns (all in fungi) differ greatly from each other and from angiosperms (supplementary fig. S2 [Supplementary Material online] and below, The Possible Origin of the *Cox1* Intron from Fungi and Intron Functionality). Araceae appear to have especially low *cox1* intron mutation rates, judging from mean branch lengths of 0.012 (±0.004) in clade A (excluding the fast mutating and phylogenetically misplaced *Peperomia*, above), compared with 0.038 (±0.013) in clade B (fig. 4).

Hierarchical Distribution of *cox1* Exonic Coconversion Tracts and Time Frame of *cox1* Intron Loss in the Araceae

When plotting Araceae exonic tract types on an Araceae phylogeny (Mayo S, personal communication) short

tract types are found in derived positions, long tract types in basal positions. For example, *Xanthosoma* with a short coconversion tract of 1 þ A in both species sequenced is derived relative to *Amorphophallus* (6 þ T). Similarly, *Philodendron hederaceum* var. *oxycardium*, with a short coconversion tract (3 substitutions; fig. 2A), is derived relative to *Zamioculcas* with 6 þ T.

A relaxed molecular clock applied to the 4,360-nt matrix of combined chloroplast and mitochondrial data (*cox1* exon, *trnL* intron and spacer, *rpl20-rps12* intergenic spacer, and *nad1* b/c exon and intron) yielded an age of 111 (SD 91–131) Myr for the stem of the intron^{-} *T. lindleyanum*/*P. virginica* clade and of 70 (SD 60–91) Myr for the divergence between these 2 species (fig. 1B, nodes 1 and 2). The divergence of the intron^{-} Vietnamese species Croat and Dzu 77954 from *Remusatia vivipara* is estimated as having occurred 49 (SD 30–67) Myr ago (fig. 1B, node 3). The fourth intron^{-} species, *P. sechellarum*, is not securely placed by our data (fig. 1A) but may have diverged from the remaining *Arisaema*/*Pistia* clade at about 73 (SD 60–82) Myr ago (fig. 1B, node 4).

The Possible Origin of the *cox1* Intron from Fungi and Intron Functionality

Vaughn et al. (1995) who first reported on the *cox1* intron in angiosperms assumed that its endonuclease was functional because of the presence of 2 LAGLI-DADG motifs (Belfort and Perlman 1995). Experimental confirmation of homing ability is still lacking. Blasting of the hypothetical protein from *Arum concinnatum* (306 residues) yielded a Blast value of 5×10^{-120} (71% identical and 83% positives) with open reading frame (ORF) 305 of the rice mold *R. oryzae* (a basal fungal lineage, formerly placed in Zycomycetes, family Mucoraceae), 3×10^{-103} (63% identical and 78% positives) with an "unknown" region (fide GenBank) in the Oyster mushroom *Pleurotus ostreatus* (Agaricomycetes and Basidiomycota), 8×10^{-75} (48% identical and 64% positives) with ORF 318 of *Monoblepharella* sp. (Chytridiomycota and Monoblepharidaceae), and of 3×10^{-67} (46% identical and 63% positives) with the *cox1* aI4 intronic protein of *Saccharomyces cerevisiae* (Ascomycota and Saccharomycetaceae), which encodes site-specific DNA endonuclease and RNA maturase activities (Wenzlau et al. 1989). The putative *cox1* intron endonucleases of angiosperms have sequence similarities of 98% (*P. oxycardium*) to 86% (*Peperomia grisoargentea*) with that of *A. concinnatum*, and the entire *cox1* intron sequence of *A. concinnatum* has sequence similarities of 77% with *Rhizopus*, 71% with *Pleurotus* and *Cryptococcus* (both Basidiomycota), and 66% with *Monoblepharella* (supplementary fig. S2, Supplementary Material online).

Discussion

The *cox1* Intron in the Araceae—A Long History of Vertical Inheritance

Considering first the distribution of the *cox1* intron on a phylogeny of the *Arisaema*/*Pistia* clade, sampled for all its genera (Renner and Zhang 2004), it is parsimoniously explained by vertical inheritance as suggested by Cho and Palmer (1999). All intron^{þ} species in this clade have the same 6 þ T coconversion tract type. An intron loss occurred in Croat and Dzu 77954, which is embedded among intron^{þ} relatives and has a coconversion tract with a single substitution (an A in position 21). Two further losses apparently occurred in *Protarum seychellarum* and in the common ancestor of the outgroup species *Typhonodorum* and *Peltandra* (fig. 1A). Judging from the fossil-constrained relaxed molecular clock, the *cox1* intron has persisted in the genomes of the *Arisaema*/*Pistia* clade for at least 80 Myr (fig. 1B, node 5). If it is ancestral in the Araceae, not just the *Arisaema*/*Pistia* clade, as suggested by the intron's phylogenetic signal, which matches the Araceae family tree (Results), it may have persisted for 110 Myr (oldest Araceae fossils, 110–120 Myr; Friis et al. 2004). The timing of at least one intron loss can also be inferred. The stem lineage of the intron^{-} *Typhonodorum*/*Peltandra* clade, which comprises just 3 species, is between 110- and 70-Myr old (Results). Its sister clade consists of a similarly species-poor group (*Ambrosina* with 1 species, *Arophyton* with 3 species, and *Arisarum* also with 3 species) that appears to have the intron (Natalie Cusimano, unpublished data for *Arisarum vulgare*). Intron loss in the *Typhonodorum*/*Peltandra* clade could have occurred some 70 Myr ago, with the coconversion tracts found in *Typhonodorum* and *Peltandra* persisting since then.

The Araceae have 4 exonic coconversion tract types, 6 þ T, 3 þ A, 3, or 1 þ A (fig. 2A). Based on a small taxon sample, a tract with 7 substitutions (6 þ T) appeared synapomorphic for the *Arisaema*/*Pistia* clade (Cho and Palmer 1999), but when all angiosperm coconversion tracts are compared with the same reference Araceae, *O. aquaticum*, as used in Cho and Palmer (1999), it is clear that the 6 þ T type is the predominant *cox1* exonic coconversion tract of intron^{þ} angiosperms (figs. 2–4). It also appears that the *cox1* exonic coconversion tracts in Araceae may be hierarchically nested, with species having 3 or 1 differences in their tracts phylogenetically more derived than species with 6 differences. Such a pattern might be expected if the 6 þ T type coconversion tract arose when the *cox1* intron first inserted itself into some ancestral Araceae (or angiosperm; see below) and was then passed on vertically, occasionally undergoing back mutation (which would lead to "shorter" coconversion tracts, namely 6 þ T - 5 þ T - 4 þ T - 3 þ T, etc., fig. 2A).

Based on the data available now, Araceae exonic coconversion tracts are less static than thought previously (Cho, Qiu, et al. 1998; Cho and Palmer 1999; Palmer et al. 2000), when it was argued that, "Regardless of how closely related they are, any 2 taxa whose coconversion tracts differ probably acquired their introns separately. For example, *Amorphophallus* and *Xanthosoma* are sister taxa with 85% bootstrap support and thus are inferred to have received their introns by vertical transmission according to all parsimony models of intron distribution (fig. 2B–E). However, because their coconversion tracts differ, and substantially so (fig. 5 [compare our figs. 2A and 4]), we conclude that they most likely acquired their introns by 2 separate, and recent, horizontal transfers. By the same logic, we conclude that *Philodendron* and *Zamioculcas*, which cluster weakly in the shortest angiosperm tree

(fig. 2A), also acquired their introns separately (fig. 6).” (Cho and Palmer 1999: 1161).

The *cox1* Intron in the Angiosperms—Predominant Loss, Not Horizontal Transfer

The hypothesis of multiple gains of the angiosperm *cox1* intron via horizontal gene transfer (Cho, Qiu, et al. 1998; Cho and Palmer 1999; Palmer et al. 2000; Richardson and Palmer 2007) was based on relatively small taxon samples, making it appear that, “Given that we have still sampled only a tiny fraction of the ~300,000 species of angiosperms, we are confident that the intron has been horizontally acquired at least hundreds of times during angiosperm evolution and probably over 1,000 times. Equally remarkably, all these transfers seem to have occurred very recently, in the last 10 Myr or so of angiosperm evolution.” (Palmer et al. 2000: 6965). Evidence for independent gains came mostly from phylogenetic incongruence between intron and angiosperm phylogenies, patchy distribution of the intron, and analyses of exonic coconversion tracts, similar to the arguments used in the case of Araceae (Cho and Palmer 1999).

Considering the argument from phylogenetic incongruence between *cox1* intron and angiosperm phylogenies, the current data suggest a different interpretation. The 3 pairs of angiosperm genera (*Ilex*/*Hydrocotyle*, *Symplocus*/*Diospyros*, and *Maranta*/*Hedychium*) for which data of Cho, Qiu, et al. (1998) showed strong disagreement between the intron and the angiosperm phylogeny are not recovered with the current larger taxon sample (fig. 4), and although the intron phylogeny contains many phylogenetically incorrect groups, it recovers an even larger number of correct clades at the species, genus, family, and even ordinal level (fig. 4). The odd groupings found by Cho, Qiu, et al. (1998) and in the current intron phylogeny (fig. 4) are probably due to low sequence variability of the *cox1* intron leading to random groupings and to a few taxa with higher mutation rates, causing long-branch attraction.

Regarding the coconversion tracts previously seen as evidence for or against a vertical or horizontal intron history, renewed analysis leads to a different conclusion. Several of the family-level clades recovered in the *cox1* intron tree (fig. 4) include species that differ in their coconversion tracts (as shown in the figure). This is the case in Araceae, Marantaceae, Acanthaceae, and Rhamnaceae. The simplest explanation of this is that in each case the intron is inherited vertically, with the exonic tracts decaying stochastically over time. Conversely, in the densely sampled order Lamiales (fig. 4), 6 þ T and 4 þ T coconversion tracts sort by family, and such slow exonic tract decay also predominates in the generally slowly evolving Araceae (at least in terms of their *cox1* sequences), which continue to pass on an ancient exonic tract. Taxa with high mitochondrial mutation rates, on the other hand, also undergo rapid changes in their coconversion tracts, as seen in *Plantago* (Cho, Adams, et al. 1998, Cho et al. 2004; our supplementary fig. S1 [Supplementary Material online]). There is also a slight positive correlation between a clade’s species sampling density and its tract type diversity (supplementary fig. S3, Supplementary Material online).

The only finding suggestive of horizontal *cox1* intron transfer is a clade of phylogenetically unrelated taxa in the intron phylogeny that comprises many species with a 4 þ T tract type (clade C in fig. 4; the clade also includes a few other tract types). We compared the *cox1* sequences of these taxa and found that they share 4 synapomorphic changes in loops L3 and L5 of the intron’s predicted secondary structure (Vaughn et al. 1995). These 4 substitutions, which do not seem to correlate with other changes in the intron or its coconversion tract, explain the high bootstrap support of the 4 þ T clade (3 nonhomoplastic changes will lead to a bootstrap support of 95%; Felsenstein 1985). The level of support for the 4 þ T clade is thus, in fact, not very high. A second observation arguing against horizontal transfer is that the subgroups inside the 4 þ T clade are monophyletic at the family level (Rhamnaceae, Plantaginaceae, and Gesneriaceae) or even the ordinal level (Lamiales). Vertical inheritance of the 4 þ T tract type, coupled with insufficient phylogenetic signal in the *cox1* intron to recover relationships at hierarchical levels above the order, thus remain the simplest explanation for all groupings in figure 4.

Together, these results suggest that differences in *cox1* coconversion tracts do not necessarily imply independent horizontal gene transfer and that phylogenetic evidence fits with a vertical history of the intron in angiosperms or at least fails to contradict it with statistical support. A largely vertical history also fits with the similar length of the intron across all angiosperms, its position at the same site in the *cox1* gene, and its generally high nucleotide similarity. Had there been thousands of horizontal transfers of the intron (perhaps over the past 10 Myr; Palmer et al. 2000), the intron phylogeny would hardly recover as many natural groups as it does nor would one expect all angiosperm introns to be essentially equally distant from the closest fungal *cox1* intron (supplementary fig. S2, Supplementary Material online). The “high frequency angiosperm-to-angiosperm horizontal transfer” hypothesis for the *cox1* gene (Richardson and Palmer 2007) also faces the difficulty of the still unknown transferring agent, although this is not a strong argument against lateral transfer.

Possible Mechanisms of *cox1* Intron Loss

One of the reasons why Cho, Qiu, et al. (1998) preferred a hypothesis of multiple intron gains over multiple losses was that each plant cell contains thousands of mitochondrial genomes. Mitochondrial genes that have lost an intron should therefore suffer an onslaught of homing introns coming from other genomes in the same cell as long as the introns’ homing endonucleases are intact. However, so far no experimental data show that the ORF-encoded protein in angiosperms *cox1* introns functions as an endonuclease. Conceivably, the angiosperm *cox1* intron ORF long ago lost its endonuclease function and now acts only as maturase for the splicing process (Delahodde et al. 1989; Wenzlau et al. 1989; Haugen et al. 2005). If this were the case, intron reinsertion by homing would not longer be possible.

Molecular mechanisms for intron loss are either recombination between an intron^{þ} and intron^{-} gene or recombination between the genomic copy of an intron^{þ} gene and a reverse transcribed copy of spliced mRNAs (Dujon

1989; Roy and Gilbert 2005; Roy and Penny 2007); another mechanism is genomic deletion as in the intron presence–absence polymorphism in *Drosophila* (Llopart et al. 2002). The *cox1* intron is always gained or lost in one step because it is self-splicing and can only function if the entire intron is inserted. For the angiosperms, we assume that the intron is lost by gene conversion (i.e., by 1 of the above 2 recombination mechanisms). That most intron^{-} angiosperms having the 0 tract type (fig. 1A, top panel), while most intron^{þ} angiosperms the 6 þ T tract type (fig. 1A, bottom panel), suggests that one reflects an event during intron insertion, the other an event correlated with intron loss. (It is also possible that the original intron donor and the first angiosperm recipient had identical *cox1* tracts, and there was no coconversion. The 6 þ T tract would then simply be an ancestral angiosperm *cox1* sequence and the 0 type would be the footprint of intron loss.) In the long run, selection on the host should favor intron loss. Based on the just discussed biased distribution of tract types, we suggest that the *cox1* coconversion tract is usually lost during the intron excision process. A stage in the angiosperm life cycle at which such loss might logically occur is during megaspore or zygote formation when the number of mitochondria is reduced and changes in the mitochondrial genome may therefore spread more easily. (Only maternal mitochondria closest to the egg cell become part of the zygote.)

Fungi as Donors of the *cox1* Intron in Angiosperms

Regarding the possible donor of the angiosperm and/or Araceae *cox1* intron, the current hypothesis is that it came from a fungus (Vaughn et al. 1995; Adams et al. 1998; Cho, Qiu, et al. 1998; Seif et al. 2005). This idea is based on 2 observations. First, the *cox1* intron is the only group I intron in vascular plant mitochondrial DNAs, whereas in fungi, group I introns in the *cox1* gene are common. Second, most angiosperms have symbiotic interactions with fungi, providing a conceivable way of intron transfer from a fungus to an angiosperm. A recent study that analyzed fungi group I introns with ORFs, including 8 in the *cox1* gene, found a *cox1* intron in *R. oryzae* (*cox1*-i1-ORF 305) that was similar to the *cox1* intron of angiosperms (Seif et al. 2005), leading to the suggestion that the angiosperm *cox1* intron "originated in a zygomycete close to *Rhizopus*." Renewed Blast searching of angiosperm *cox1* introns (28 August 2007) still yields the widespread mold *R. oryzae* and the Oyster mushroom *P. ostreatus* as the closest relatives outside of angiosperms (supplementary fig. S2, Supplementary Material online). However, sampling in the fungi is extremely sparse and sequence homology low.

Regardless of how many times and from which fungus the *cox1* intron entered the angiosperms, such entry was hardly a straightforward process because of differences in the genetic code used by fungi and angiosperms (Fox 1987). For intron homing to function, the encoded endonuclease must be translated and differences in codes may cause difficulties in translation. Further difficulties are the existence of C - U RNA editing in plant but not fungal mitochondria (Gray 1996) and differences in promoter sequences recognized by the fungal and plant mitochondrial transcriptional apparatus (Tracy and Stern 1995). Nevertheless, there is indirect evidence that angiosperm-to-fungus intron transfer can occur (Nishida and Sugiyama 1995). The *cox1* introns of other spermatophytes, for example, *Marchantia polymorpha* (Ohta et al. 1993), are more distant from angiosperm *cox1* introns than are fungal *cox1* introns.

Conclusion

For Araceae, the fit between the *cox1* intron and the Araceae phylogeny and the highly conserved coconversion tracts together suggest vertical intron inheritance over 110 Myr, with several independent losses within Araceae. Current data for all angiosperms likewise point to a history dominated by vertical intron inheritance followed by repeated intron loss. The alternative hypothesis of numerous horizontal acquisitions has difficulties explaining the observed congruence between the intron and the angiosperm phylogeny as well as the evidence from the 20 coconversion tract types found across angiosperms. Coconversion tracts can no longer be regarded as static footprints. Instead their analysis in a phylogenetic framework provides evidence of their gradual decay and loss, most likely at the excision stage and by reverse transcription–mRNA-mediated coconversion. The hypothesis that fungi are the source of the angiosperm *cox1* intron fits with current data, but sampling in fungi is still extremely sparse, and specific donor lineages can therefore not yet be named.

Supplementary Material

Supplementary table S1, figures S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Correction in proof: Due to an erroneous sequence in GenBank (AY009433 of Asimina triloba; corrected by T. Barkman on 31 Dec. 2007), Asimina triloba was counted as lacking the *cox1* intron, when instead it does have an intron in its *cox1* gene.

Literature Cited

Adams KL, Clements MJ, Vaughn JC. 1998. The *Peperomia* mitochondrial coxI group I intron: timing of horizontal transfer and subsequent evolution of the intron. J Mol Evol. 46:689–696.

Belfort M. 2003. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. Genes Dev. 17:2860–2863.

Belfort M, Perlman PS. 1995. Mechanisms of intron mobility. J Biol Chem. 270:30237–30240.

Bell-Pedersen D, Quirk SM, Aubrey M, Belfort M. 1989. A site-specific endonuclease and coconversion of flanking exons associated with the mobile *td* intron of phage T4. Gene. 82:119–126.

Braithwaite CJR. 1984. Geology of the Seychelles. In: Stoddart DR, editor. Biogeography and ecology of the Seychelles Islands. The Hague (The Netherlands): Junk. p. 17–38.

Chevalier BS, Stoddard BL. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res. 29:3757–3774.

Cho Y, Adams KL, Qiu Y-L, Kuhlman P, Vaughn JC, Palmer JD. 1998. A highly invasive group I intron in the

mitochondrial cox1 gene. In: Moller I-M, Gardestrom P, Glimelius K, Glaser E, editors. Plant mitochondria: from gene to function. Leiden (The Netherlands): Backhuys Publishers. p. 19–23.
Cho Y, Mower JP, Qiu Y-L, Palmer JD. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc Natl Acad Sci USA. 101:17741–17746.
Cho Y, Palmer JD. 1999. Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial coxl gene during evolution of the Araceae family. Mol Biol Evol. 16:1155–1165.
Cho Y, Qiu Y-L, Kuhlman P, Palmer JD. 1998. Explosive invasion of plant mitochondria by a group I intron. Proc Natl Acad Sci USA. 95:14244–14249.
Delahodde A, Goguel V, Becam AM, Creusot F, Perea J, Banroques J, Jacq C. 1989. Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. Cell. 56:431–441.
Dujon B. 1989. Group I introns as mobile genetic elements: facts and mechanistic speculations—a review. Gene. 82:91–114.
Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 39:783–791.
Fox TD. 1987. Natural variation in the genetic code. Annu Rev Genet. 21:67–91.
Friis EM, Pedersen KR, Crane PR. 2004. Araceae from the Early Cretaceous of Portugal: evidence on the emergence of monocotyledons. Proc Natl Acad Sci USA. 101:16565–16570.
Goddard MR, Burt A. 1999. Recurrent invasion and extinction of a selfish gene. Proc Natl Acad Sci USA. 96:13880–13885.
Gray MW. 1996. RNA editing in plant organelles: a fertile field. Proc Natl Acad Sci USA. 93:8157–8159.
Haugen P, Simon DM, Bhattacharya D. 2005. The natural history of group I introns. Trends Genet. 21:110–119.
Hughes NF. 1994. The enigma of angiosperm origins. Cambridge: Cambridge University Press.
Lambowitz AM. 1989. Infectious introns. Cell. 56:323–326.
Lambowitz AM, Belfort M. 1993. Introns as mobile genetic elements. Annu Rev Biochem. 62:587–622.
Lang BF. 1984. The mitochondrial genome of the fission yeast Schizosaccharomyces pombe: highly homologous introns are inserted at the same positions of the otherwise less conserved cox1 genes in Schizosaccharomyces pombe and Aspergillus nidulans. EMBO J. 3:2129–2136.
Llopart A, Comeron JM, Brunet FG, Lachaise D, Long M. 2002. Intron presence–absence polymorphism in Drosophila driven by positive Darwinian selection. Proc Natl Acad Sci USA. 99:8121–8126.
Maddison WP, Maddison DR. 2003. MacClade: analysis of phylogeny and character evolution. Version 3.0. Sunderland (MA): Sinauer Associates.
Nishida H, Sugiyama J. 1995. A common group I intron between a plant parasitic fungus and its host. Mol Biol Evol. 12: 883–886.
Ohta E, Oda K, Yamato K, Nakamura Y, Takemura M, Nozato N, Akashi K, Ohyama K, Michel F. 1993. Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts. Nucleic Acids Res. 21:1297–1305.
Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K. 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc Natl Acad Sci USA. 97:6960–6966.
Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.
Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics. 14:817–818.
Qiu Y-L, Dombrovska O, Lee JH, et al. (20 co-authors). 2005. Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. Int J Plant Sci. 166: 815–842.
R Development Core Team. 2006. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
Renner SS, Zhang L-B. 2004. Biogeography of the Pistia clade (Araceae) based on chloroplast and mitochondrial DNA sequences and Bayesian divergence time inference. Syst Biol. 53:422–432.
Richardson AO, Palmer JD. 2007. Horizontal gene transfer in plants. J Exp Bot. 58:1–9.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.
Rot C, Goldfarb I, Ilan M, Huchon D. 2006. Putative cross-kingdom horizontal gene transfer in sponge (Porifera) mitochondria. BMC Evol Biol. 6:71.
Roy SW, Gilbert W. 2005. The pattern of intron loss. Proc Natl Acad Sci USA. 102:713–718.
Roy SW, Penny D. 2007. Patterns of intron loss and gain in plants: intron loss–dominated evolution and genome-wide comparison of O. sativa and A. thaliana. Mol Biol Evol. 24:171–181.
Seif E, Leigh J, Liu Y, Roewer I, Forget L, Lang BF. 2005. Comparative mitochondrial genomics in zygomycetes: bacteria-like RNase P RNAs, mobile elements and a close source of the group I intron invasion in angiosperms. Nucleic Acids Res. 33:734–744.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22:2688–2690.
Stevens PF. 2001. onwards. 2001. Angiosperm Phylogeny Website Version 8. June 2007, [Internet]. Available from: http://www.mobot.org/MOBOT/research/APweb/
Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.
Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. 1983. The double-strand-break repair model for recombination. Cell. 33:25–35.
Thorne JL, Kishino H. 2002. Divergence time estimation and rate evolution with multilocus data sets. Syst Biol. 51:689–702.
Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol. 15:1647–1657.
Tracy RL, Stern DB. 1995. Mitochondrial transcription initiation: promoter structures and RNA polymerases. Curr Genet. 28:205–216.
Vaughn JC, Mason MT, Sper-Whitis GL, Kuhlman P, Palmer JD. 1995. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric coxI gene of Peperomia. J Mol Evol. 41:563–572.
Wenzlau JM, Saldanha RJ, Butow RA, Perlman PS. 1989. A latent intron-encoded maturase is also an endonuclease needed for intron mobility. Cell. 56:421–430.
Wilde V, Kvacek Z, Bogner J. 2005. Fossil leaves of Araceae from the European Eocene and notes on other aroid fossils. Int J Plant Sci. 166:157–183.