Available online at www.sciencedirect.com
ScienceDirect
Organisms, Diversity & Evolution 7 (2007) 55–77
ORGANISMS DIVERSITY & EVOLUTION
www.elsevier.de/ode

Phylogeny of basal eudicots: Insights from non-coding and rapidly evolving DNA

Andreas Worberg[a], Dietmar Quandt[b], Anna-Magdalena Barniske[b], Cornelia Löhne[a], Khidir W. Hilu[c], Thomas Borsch[a,*]

[a]*Nees-Institut für Biodiversität der Pflanzen, Rheinische Friedrich-Wilhelms-Universität Bonn, Meckenheimer Allee 170, 53115 Bonn, Germany*
[b]*Institut für Botanik, Technische Universität Dresden, Zellescher Weg 20b, 01217 Dresden, Germany*
[c]*Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA*

Received 13 June 2006; accepted 10 August 2006

Dedicated to Wilhelm Barthlott on the occasion of his 60th birthday

Abstract

Sequence data of the *trnL* group I intron, the *petD* group II intron, the *trnL-F* and *petB-D* spacers, and the rapidly evolving *matK* gene were analysed from all families of the basal eudicot grade and from representatives of 19 core eudicot orders. The dataset comprised 5654 positions of aligned sequence plus a matrix of 1087 binary indel characters. Mutational hotspots correspond in number and extension to hotspots already known from basal angiosperms and, with respect to secondary structure, are generally located in terminal parts of stem-loop regions. Parsimony, Bayesian, and likelihood analyses depict Ranunculales as sister to all remaining eudicots with maximum support. The branching order in the basal eudicot grade is further resolved as Sabiales, Proteales, Trochodendrales, and Buxales. Nearly all of the backbone nodes gain high confidence, except for the node showing Proteales diverging before Trochodendrales, which is only moderately supported (83% JK). In Ranunculales, the woody Eupteleaceae are first-branching, with Papaveraceae plus Fumariaceae coming next. Within Proteales, *Nelumbo* is clearly resolved as sister to a Platanaceae–Proteaceae clade. Gunnerales are found as the first branch in core eudicots, with maximum support in the combined analysis. This node is also resolved with *matK* alone, but unsupported. It appears that the combined analysis of sequence data from rapidly evolving and non-coding genomic regions leads to significantly improved statistical support values in comparison to earlier studies of basal eudicots using multiple conserved genes.
© 2006 Gesellschaft für Biologische Systematik. Published by Elsevier GmbH. All rights reserved.

Keywords: Angiosperms; *matK*; *trnL-F*; *petD*; Molecular evolution
See also **Electronic Supplement** at doi:10.1016/j.ode.2006.08.001

Introduction

The eudicot clade comprises the vast majority of angiosperm diversity, with an estimated 200,000 species (Drinnan et al. 1994). The clade was first recognized by Donoghue and Doyle (1989) and Doyle and Hotton

*Corresponding author. Tel.: +49 228 732681; fax: +49 228 733120.
E-mail address: borsch@uni-bonn.de (T. Borsch).

1439-6092/$ - see front matter © 2006 Gesellschaft für Biologische Systematik. Published by Elsevier GmbH. All rights reserved.
doi:10.1016/j.ode.2006.08.001

(1991) based on morphological characters. Broad-scale molecular analyses of angiosperms using single genes (Chase et al. 1993; Savolainen et al. 2000a) have consistently recovered the eudicots. The clade has gained maximum support when three or more genes were combined (Qiu et al. 2000; Soltis et al. 2000; Kim et al. 2004). More recently, the analysis of partial *matK* sequences alone yielded 96% JK (Hilu et al. 2003). The eudicots share tricolpate and tricolpate-derived pollen (Donoghue and Doyle 1989; Nandi et al. 1998; Hoot et al. 1999). Thus they have also been called the tricolpate clade (Donoghue and Doyle 1989).

Using DNA data, a number of lineages (Ranunculales, Proteales, Sabiaceae, Buxaceae plus Didymelaceae, and Trochodendraceae plus Tetracentraceae) have been identified as representing the earliest branches in eudicots (Chase et al. 1993; Savolainen et al. 2000b; Soltis et al. 2000, 2003; Hilu et al. 2003), whereas large groups such as asterids, Caryophyllales, rosids, Santalales, and Saxifragales were shown to belong to the core eudicots. The core eudicot node is one of the best-supported nodes within the angiosperm tree (Hilu et al. 2003; Soltis et al. 2003) and obviously marks a major shift in angiosperm evolution. The core eudicot node is also identified by recent analyses of MADS-box genes, where non-core tricolpate clades only have the eu*FUL* gene family and lack the eu*AP1* gene family. Thus, this gene duplication is a synapomorphy for core eudicots (Litt and Irish 2003; Kim et al. 2004; Kramer et al. 2004). Molecular clock dating inferred the eudicots to have an age of 131–125 mya (Magallon et al. 1999; Anderson et al. 2005), whereas the core eudicot node is estimated at 113–116 mya (Magallon et al. 1999; Anderson et al. 2005; Leebens-Mack et al. 2005).

So far, five different coding genes, analysed alone or in combination, have been used to reconstruct relationships of early branching eudicots. The first genes to be analysed were *rbcL* (Chase et al. 1993) and *atpB* (Savolainen et al. 2000a). Their use recovered all lineages belonging to the "basal eudicots", but support for their inter-relationships was not evident. Nevertheless, terminal clades like Ranunculales, Proteales, or Buxaceae–Didymelaceae were identified, and both genes converged on the first-branching position of Ranunculales in eudicots. Hoot et al. (1999) and Soltis et al. (2000) added nuclear 18S sequences. Their analyses showed improved support for most terminal clades. Buxaceae–Didymelaceae and Trochodendraceae were depicted either as successive sisters to core eudicots or in a tritomy with the core eudicots. The clade including Buxaceae–Didymelaceae, Trochodendraceae, and core eudicots gained 87–88% JK support. The respective positions of Sabiaceae and Proteales were not resolved with confidence. Even adding nr26S sequences for a four-gene dataset (Kim et al. 2004) did not improve resolution in the basal eudicot grade. Phylogenetic analysis of a dataset comprising two thirds of the rapidly evolving *matK* gene (Hilu et al. 2003) provided a picture similar to that of the multi-gene analyses. Moreover, *matK* indicated that Buxaceae are sister to core eudicots (91% JK, 1.00 posterior probability (PP)) and provided moderate support (82% JK) for the first-branching position of Ranunculales in eudicots.

Recently, sequences of introns such as the group I intron in *trnL*, and the group II intron in *petD* were used to infer relationships among basal angiosperms (Borsch et al. 2003, 2005; Löhne and Borsch 2005). The same applies to the *trnT-L* and *trnL-F* spacers (Borsch et al. 2003) which, like the above-mentioned introns, are located in the large single-copy region of the chloroplast genome, and are rapidly evolving. It was shown that mutational dynamics in these spacers and introns follows complex patterns related to structural constraints. Extreme length variability in introns and spacers is confined to certain mutational hotspots which correspond to the least constrained stem-loop elements P6 and P8 in the secondary structure of the group I intron (Quandt et al. 2004), and to the least constrained terminal stem-loop elements of domains I, II, and IV in the group II intron (Löhne and Borsch 2005). Moreover, the *petD* intron dataset yielded one of the largest indel matrices so far generated for angiosperms. Reconstructing the evolution of the underlying microstructural mutations, involving one to many nucleotides, showed a large number of them to be synapomorphic for deep to terminal nodes. Thus, microstructural mutations in rapidly evolving spacers and introns can be expected to be of high phylogenetic utility (Kelchner 2000), as has been shown for indels supporting shallower nodes (Müller and Borsch 2005) as well as for indels in the conserved chloroplast-inverted repeat (Graham et al. 2000). In basal angiosperms it was evident that combining *trnT-F* and *petD* sequences with *matK*, which also is rapidly evolving and has provided good signal in an overall angiosperm analysis (Hilu et al. 2003), can lead to further improved resolution and support of phylogenetic trees (Borsch et al. 2005; Müller et al. 2006). Combining such datasets could therefore have the potential of providing further insight into some of the nodes that are notoriously difficult to resolve in the basal eudicot grade. In comparison to analyses of basal angiosperms, where gymnosperms had to be used as outgroup, a *petD* and *trnL-F* eudicot dataset with basal angiosperms as outgroups could be expected to entail lower *p*-distances, and thus to be easier to align. Because mutational dynamics is strongly influenced by structural constraints inherent to the respective genomic region, at least in introns, hotspots were to be expected in similar positions in eudicots as compared to basal angiosperms.

The aims of the present study were: (1) to produce an alignment of rapidly evolving group I and group II introns, and of spacers, for a taxon sampling

representative of basal eudicots; (2) to examine molecular evolutionary patterns of non-coding genomic regions, and to test their phylogenetic signal in basal eudicots; (3) to reconstruct basal eudicot relationships using a combined set of intron-, spacer-, and complete *matK* sequences, in order to test whether resolution and support can be improved over existing basal eudicot trees.

Material and methods

Taxon sampling and plant material

In total, sequences from five genomic regions were analysed: the *petB-D* spacer, the *petD* group II Intron, the *trnL* group I intron, the *trnL-F* spacer, and the *matK* gene. The dataset comprises 56 angiosperm species, representing 47 families from 31 orders recognized by APG II (2003). For practical reasons the five genomic regions were treated as three partitions, which are usually amplified and sequenced together. Thus, *petB-D* spacer plus *petD* intron are called the "*petD*" partition; *trnL* intron plus *trnL-F* spacer the "*trnL-F* partition". Outgroup taxa were chosen to represent the first-branching angiosperms, the magnoliids, Chloranthaceae, *Ceratophyllum*, and monocots (*Acorus*). All major lineages of basal eudicots are included, comprising 22 species in 14 families. The core eudicots are represented by Gunnerales (3 species) and several families each of the major clades such as Saxifragales (2), Vitales (2), rosids (6), asterids (5), Caryophyllales (2), plus Dilleniales (1), Santalales (1), and Berberidopsidales (1).

Most sequences were generated in this study (Table 1). For *petD*, there are 35 new sequences, whereas 15 were taken from Löhne and Borsch (2005). For *trnL-F*, 34 sequences are new, 15 were originally published by Borsch et al. (2003). For *matK*, 20 completely new sequences were produced, and 12 partial sequences originally generated for the large-scale angiosperm analysis (Hilu et al. 2003) were completed in this study. For the latter purpose, already existing PCR products were sequenced with additional primers, or the upstream halves of the *trnK* intron were amplified from the same DNA already used earlier. Fourteen *matK* sequences were complete already from the study of Müller et al. (2006). Four single sequences of *matK* and one of *trnL-F* as well as complete plastome sequences for *Arabidopsis thaliana*, *Atropa belladonna*, *Nicotiana tabacum*, *Oenothera elata*, *Panax ginseng*, and *Spinacia oleracea* were downloaded from GenBank (Table 1). The *trnL-F* sequence of *Brassica nigra* was used to replace *Arabidopsis thaliana*, as the corresponding whole-genome sequence contained obvious sequencing errors. All taxa included in this study, the respective voucher information and GenBank accession numbers are listed in Table 1.

DNA isolation, amplification, and sequencing

DNA was isolated from fresh or silica gel-dried plant material, using a CTAB method with three extractions (Borsch et al. 2003), designed to yield high amounts of genomic DNA. To study molecular evolution and identify mutational hotspots, complete sequences of spacers and introns are necessary. Amplification was done with primers that were located sufficiently far away from the actual region under study. Sequencing was performed with either the universal primers already used for amplification (such as *petD* primers or various *matK* and *trnL-F* primers) or with additional internal primers, some of which were newly designed using SeqState v1.25 (Müller 2005b). For *petD* both amplification and sequencing were performed with the set of universal primers from Löhne and Borsch (2005). They were supplemented by an internal sequencing primer HE-petD-343R which is located about 340 bp downstream in the *petD* intron, in order to cover polyA stretches in the upstream spacer region. Amplification of *trnL-F* was done with universal primers trnTc and trnTf (Taberlet et al. 1991). Products were then sequenced with trnTd (reversal primer annealing to the trnL 3′ exon; Taberlet et al. 1991), and with trnL460F which is a new universal forward-sequencing primer located about 100 nt upstream of the *trnL*-5′ exon. The *matK* gene was amplified within the *trnK* intron, either entirely or in two overlapping halves. Primers annealing to the *trnK* exons were trnKFbryo (Quandt in press; forward) and trnK2R (Johnson and Soltis 1995). To amplify two overlapping fragments, additional primers were placed about 600 bp downstream (reverse) and about 450 bp downstream (forward) of the *matK* start codon, respectively. Because of deviating sequences several lineage-specific internal primers were used, two of which (ROSmatK530F and ROSmatK655R) were newly designed. For some taxa with deviating sequences and/or microsatellites, further internal sequencing primers had to be designed, such as DIDYmatK570F, DIDYmatK1107F, and DIDYmatK1035R for *Didymeles*. All primers used in this study are listed in the Appendix A (see Electronic Supplement).

Amplification and sequencing reactions were performed in a T3 Thermocycler (Biometra; Göttingen, Germany). Amplicons were purified with a QIAquick gel extraction kit (QIAGEN; Hilden, Germany) after running them out on a 1.2% agarose gel and excising the bands. The BeckmannCoulter DTCS QuickStart Reaction kit was used for direct sequencing. Temperature profiles and PCR reaction conditions followed Löhne and Borsch (2005) for *petD*, Borsch et al. (2003) for

Table 1. Taxa analysed (family assignment according to APG II 2003), voucher data, and references

Taxon	Family	Voucher/ Herbarium	Garden/Field Origin	EMBL/GenBank numbers and references		
				matK	*trnL-F*	*petD*
OUTGROUP						
Chimonanthus praecox (L.) Link	Calycanthaceae	T. Borsch 3396 (BONN)	BG Bonn	AF542569 This study update	AM397150 This study	AM396524 This study
Hedycarya arborea Forst.	Monimiaceae	A. Worberg 014 (BONN)	BG Bonn	AM396509 This study	AM397149 This study	AM396523 This study
Umbellularia californica (Hooker and Arn.) Nutt.	Lauraceae	T. Borsch 3471 (BONN)	BG Bonn	AF543752 Müller et al. (2006)	AY145350 Borsch et al. (2003)	AY590850 Löhne and Borsch (2005)
Magnolia virginiana L.	Magnoliaceae	T. Borsch and C. Neinhuis 3280 (VPI, FR)	USA, Maryland	AB020988 Azuma et al. (1999)	AY145354 Borsch et al. (2003)	—
Magnolia officinalis Rehder and Wilson	Magnoliaceae	C. Löhne 53 (BONN)	BG Bonn	—	—	AY590846 Löhne and Borsch (2005)
Chloranthus brachystachys Blume	Chloranthaceae	T. Borsch 3467 (BONN)	BG Bonn	AF543733 Müller et al. (2006)	AY145334 Borsch et al. (2003)	AY590864 Löhne and Borsch (2005)
Acorus gramineus L.	Acoraceae	T. Borsch 3458 (BONN)	BG Bonn	DQ182341 Müller et al. (2006)	AY145336 Borsch et al. (2003)	—
Acorus calamus L.	Acoraceae	C. Löhne 51 (BONN)	BG Bonn	—	—	AY590840 Löhne and Borsch (2005)
Ceratophyllum demersum L.	Ceratophyllaceae	T. Wieboldt 16073 (VPI)	USA, Virginia	AF543732 Müller et al. (2006)	AY145335 Borsch et al. (2003)	AY590841 Löhne and Borsch (2005)
Aristolochia pistolochia L.	Aristolochiaceae	T. Borsch 3257 (FR)	France, Herault	AF543724 Müller et al. (2006)	AY145341 Borsch et al. (2003)	AY590862 Löhne and Borsch (2005)
Austrobaileya scandens C. White	Austrobaileyaceae	T. Borsch 3464 (BONN)	BG Bonn	DQ182344 Müller et al. (2006)	AY145326 Borsch et al. (2003)	AY590867 Löhne and Borsch (2005)
Nymphaea odorata Aiton ssp. *odorata*	Nymphaeaceae	T. Borsch and V.Wilde 3132 (VPI, BONN)	USA, Georgia	—	AY145333 Borsch et al. (2003)	—
Nymphaea odorata Aiton ssp. *tuberosa* (Paine) Wiersema and Hellq.	Nymphaeaceae	T.Borsch, B.Hellquist, J.Wiersema 3389 (BONN)	Canada, Manitoba	DQ185549 Löhne et al. (pers. comm.)	—	AY590873 Löhne and Borsch (2005)
Amborella trichopoda Baill.	Amborellaceae	T. Borsch 3480 (VPI)	UCLA, Sta. Catarina BG	AF543721 Müller et al. (2006)	AY145324 Borsch et al. (2003)	AY590876 Löhne and Borsch (2005)
BASAL EUDICOTS						
Euptelea pleiosperma Siebold and Zucc.	Eupteleaceae	A. Worberg 003 (BONN)	BG Bonn	AM396510 This study	AM397151 This study	AM396525 This study
Akebia quinata Decne.	Lardizabalaceae	T. Borsch 3412 (BONN)	BG Bonn	AF542587 This study update	AM397152 This study	AM396526 This study
Dicentra eximia (Ker Gawl.)Torr.	Papaveraceae	T. Borsch 3468 (BONN)	BG Bonn	DQ182345 Müller et al. (2006)	AY145361 Borsch et al. (2003)	AY590835 Löhne and Borsch (2005)
Papaver triniaefolium Boiss.	Papaveraceae	A. Worberg 018 (BONN)	BG Bonn	AM396511 This study	AM397153 This study	AM396527 This study
Cocculus laurifolius DC.	Menispermaceae	T. Borsch 3406 (BONN)	BG Bonn	AF542588 This study update	AM397159 This study	AM396528 This study
Stephania delavayi Diels.	Menispermaceae	T. Borsch 3550 (BONN)	BG Bonn	AF542584 This study update	AM397154 This study	AM396529 This study

Table 1. (*continued*)

Taxon	Family	Voucher/ Herbarium	Garden/Field Origin	EMBL/GenBank numbers and references		
				matK	*trnL-F*	*petD*
Xanthoriza simplicissima Woodhouse	Ranunculaceae	T. Borsch 3394 (BONN)	BG Bonn	AF542567 This study update	AM397155 This study	AM396530 This study
Mahonia japonica DC.	Berberidaceae	T. Borsch 3405 (BONN)	BG Bonn	—	AM397156 This study	AM396531 This study
Mahonia japonica DC.	Berberidaceae	GenBank	BG Bonn	AB038184 Kita and Kato (2001)	—	—
Podophyllum peltatum L.	Berberidaceae	T. Borsch 3393 (BONN)	BG Bonn	—	AM397157 This study	AM396532 This study
Podophyllum peltatum L.	Berberidaceae	GenBank	BG Bonn	AB069843 K. Kosuge (pers.comm.)	—	—
Sabia japonica Maxim.	Sabiaceae	Y-L. Qiu 91025 (NCU)	NCU	AM396512 This study	AM397158 This study	AM396533 This study
Meliosma cuneifolia Franch.	Sabiaceae	A. Worberg 001 (BONN)	BG Bochum	AM396513 This study	AM397160 This study	AM396534 This study
Nelumbo nucifera Gaertn. ssp. *nucifera 'Alba'*	Nelumbonaceae	A. Worberg s.n. (BONN)	BG Bonn	AM396514 This study	AM397161 This study	AM396535 This study
Nelumbo nucifera Gaertn. ssp. *lutea* (Willd.) Borsch and Barthlott	Nelumbonaceae	T. Borsch and Summers 3220 (FR)	USA, Missouri	AF543740 Müller et al. (2006)	AY145359 Borsch et al. (2003)	AY590836 Löhne and Borsch (2005)
Embothrium coccineum Forst.	Proteaceae	A. Worberg 004 (BONN)	BG Bonn	AM396515 This study	AM397162 This study	AM396536 This study
Grevillea banksii R. Br.	Proteaceae	T. Borsch 3413 (BONN)	BG Bonn	AF542583 This study update	AM397163 This study	AM396537 This study
Platanus orientalis L.	Platanaceae	A. Worberg 005 (BONN)	BG Bonn	AM396503 This study	AM397164 This study	AM396538 This study
Platanus occidentalis L.	Platanaceae	Slotta s.n. (VPI)	USA, Virginia	AF543747 Müller et al. (2006)	AY145358 Borsch et al. (2003)	AY590834 Löhne and Borsch (2005)
Tetracentron sinense Oliver	Trochodendraceae	T. Borsch 3494 (BONN)	BG Freiburg	AM396504 This study	AM397165 This study	AM396539 This study
Trochodendron aralioides Siebold and Zucc.	Trochodendraceae	T. Borsch 3478 (BONN)	BG Bonn	AF543751 Müller et al. (2006)	AY145360 Borsch et al. (2003)	AY590833 Löhne and Borsch (2005)
Didymeles integrifolia J. St.-Hil.	Didymelaceae	J. Rabenantoandro et al. 916 (MO)	Madagascar	AM396505 This study	AM397166 This study	AM396540 This study
Buxus sempervirens L.	Buxaceae	T. Borsch 3465 (BONN)	BG Bonn	AF543728 Müller et al. (2006)	AY145357 Borsch et al. (2003)	AY590832 Löhne and Borsch (2005)
Pachysandra terminalis Siebold and Zucc.	Buxaceae	T. Borsch 3407 (BONN)	BG Bonn	AF542581 This study update	AM397167 This study	AM396541 This study
CORE EUDICOTS						
Gunnera tinctoria (Molina) Mirb.	Gunneraceae	N. Korotkov 50 (BONN)	BG Bonn	AM396506 This study	AM397168 This study	AM396542 This study
Myrothamnus flabellifolia Welw.	Myrothamnaceae	A. Worberg 011 (BONN)	BG Bonn	AM396507 This study	AM397169 This study	AM396543 This study
Myrothamnus moschata Baill.	Myrothamnaceae	E. Fischer and W. Höller (BONN)	BG Bonn	AF542591 This study update	AM397170 This study	AM396544 This study
Cercidiphyllum japonicum Siebold and Zucc.	Cercidiphyllaceae	T. Borsch s.n. (BONN)	BG Bonn	AM396508 This study	AM397171 This study	AM396545 This study
Chrysosplenium alternifolium L.	Saxifragaceae	T. Borsch s.n. (BONN)	Germany	AM396496 This study	AM397172 This study	AM396546 This study
Vitis riparia A. Gray	Vitaceae	T. Borsch 3458 (BONN)	BG Bonn	AF542593 This study update	AM397173 This study	AM396547 This study

Table 1. (*continued*)

Taxon	Family	Voucher/ Herbarium	Garden/Field Origin	EMBL/GenBank numbers and references		
				matK	*trnL-F*	*petD*
Leea coccinea Planch.	Leeaceae	T. Borsch 3418 (BONN)	BG Bonn	AM396497 This study	AM397174 this study	AM396548 This study
Dillenia philippinensis Rolfe	Dilleniaceae	A. Worberg 010 (BONN)	BG Bonn	AM396498 This study	AM397175 This study	AM396549 This study
Aextoxicon punctatum Ruiz and Pav.	Aextoxicaceae	T. Borsch 3459 (BONN)	BG Bonn	DQ182342 Müller et al. (2006)	AY145362 Borsch et al. (2003)	AY590831 Löhne and Borsch (2005)
Osyris alba L.	Santalaceae	A. Worberg 015 (BONN)	BG Bonn	AM396499 This study	AM397176 This study	AM396550 This study
CARYOPHYLLIDS						
Rhipsalis paradoxa Salm-Dyck.	Cactaceae	A. Worberg s.n. (BONN)	BG Bonn	—	AM397177 This study	AM3965551 This study
Rhipsalis floccosa Salm-Dyck.	Cactaceae	GenBank	—	AY01534 Nyffeler (2002)	—	—
Spinacia oleracea L.	Chenopodiaceae	GenBank	—	AJ400848; Schmitz-Linneweber et al. (2001)	AJ400848; Schmitz-Linneweber et al. (2001)	AJ400848; Schmitz-Linneweber et al. (2001)
ROSIDS						
Erodium cicutarium (L.) L'Hér	Geraniaceae	T. Borsch 3483 (BONN)	Germany, Eifel	AM396500 This study	AM397178 This study	AM396552 This study
Brassica nigra (L.) W.D.J. Koch	Brassicaceae	GenBank	—	—	AF451579 Yang et al. (2002)	—
Arabidopsis thaliana (L.) Heynh.	Brassicaceae	GenBank	—	NC000932 Sato et al. (1999)	—	NC000932 Sato et al. (1999)
Stachyurus chinensis Franch.	Stachyuraceae	A. Worberg s.n. (BONN)	BG Bonn	AM396501 This study	—	AM396555 This study
Stachyurus chinensis Franch.	Stachyuraceae	GenBank	—	—	AB066335 Ohi et al. (2003)	—
Coriaria myrtifolia L.	Coriariaceae	T. Borsch 3415 (BONN)	BG Bonn	AF542600 This study update	AM397179 This study	AM396553 This study
Larrea tridentata Coult.	Zygophyllaceae	A. Worberg 012 (BONN)	BG Bonn	AM396502 This study	AM397180 This study	AM396554 This study
ASTERIDS						
Impatiens noli-tangere L.	Balsaminaceae	T. Borsch 3485 (BONN)	BG Bonn	AF542608 This study update	AM397181 This study	AM396556 This study
Ilex aquifolium L.	Aquifoliaceae	T. Borsch 3419 (BONN)	BG Bonn	AF542607 This study update	AM397182 This study	AM396557 This study
Oenothera elata Kunth	Onagraceae	GenBank	—	NC002693 Hupfer et al. (2000)	NC002693 Hupfer et al. (2000)	NC002693 Hupfer et al. (2000)
Panax ginseng C.A. Mey.	Araliaceae	GenBank	—	AY582139 Kim and Lee (2004)	AY582139 Kim and Lee (2004)	AY582139 Kim and Lee (2004)
Atropa belladonna L.	Solanaceae	GenBank	—	NC004561; Schmitz-Linneweber et al. (2002)	NC004561; Schmitz-Linneweber et al. (2002)	NC004561; Schmitz-Linneweber et al. (2002)
Nicotiana tabacum L.	Solanaceae	GenBank	—	NC001879 Shinozaki et al. (1986)	NC001879 Shinozaki et al. (1986)	NC001879 Shinozaki et al. (1986)

trnL-F, Hilu et al. (2003), and Quandt (in press) for *matK*. Extension products were run on BeckmannCoulter CEQ 8000 automated sequencers in Bonn or Dresden. Sequences were edited manually with PhyDE v0.972 (Müller et al. 2005).

Alignment, indel coding, and phylogenetic analysis

The presence of microstructural changes, such as deletions, single sequence repeats, other insertions, and inversions, necessitates special attention to the alignment of sequences. Alignment was carried out by eye using PhyDE v0.972, applying the rules outlined in Borsch et al. (2003) and Löhne and Borsch (2005). These alignment rules are based on recognizing sequence motifs that result from microstructural changes (Golenberg et al. 1993; Kelchner and Clark 1997; Kelchner 2000) rather than globally applying fixed gap costs. The history of microstructural changes has been reconstructed within several orthologous genomic regions of the slowly chloroplast-inverted repeat (Graham et al. 2000) and the more rapidly evolving spacers and introns of the single-copy regions (Löhne and Borsch 2005; Stech and Quandt 2006; Borsch et al. in press), indicating mutational patterns common to the chloroplast genome. However, presently available alignment algorithms and software applications (e.g. Wheeler et al. 1996–2003; Benson 1997; Morgenstern 1999) are not yet able to recognize these patterns, and provide only unsatisfactory approximations of primary homology. Therefore they were not used here. Approaches to include ambiguously alignable regions in character matrices (Lutzoni et al. 2000; Aagesen 2004) were not pursued here either. Sequence stretches with unclear primary homology were excluded from tree inference in order to achieve maximum accuracy. To utilize indel characters, we applied the simple indel-coding method (Simmons and Ochoterena 2000) via SeqState v1.25 (Müller 2005b). The resulting indel matrix was then used in combination with the nucleotide-sequence matrix for parsimony analyses and Bayesian Inference (BI).

To infer most parsimonious trees, we used the Parsimony Ratchet (Nixon 1999) as implemented in the program PRAP (Müller 2004). Ratchet settings were 20 random-addition cycles of 200 ratchet replicates, and upweighting 25% of the characters. If more than one shortest tree was found, strict consensus trees were created. Nodes were evaluated by jackknifing in PAUP* (Swofford 2001) with 36.79% deletion of characters and 10,000 replicates, saving only 1 tree per replicate. This approach follows recent studies on the reliability of jackknife percentages (Freudenstein and Simmons 2004; Müller 2005a). Since the effect of adding indel characters that further support already well-supported nodes cannot be tested by comparing jackknife percentages, decay values were calculated with the help of PRAP, using the ratchet settings mentioned above. Before combining individual partitions, incongruence length difference tests (partition homogeneity test) were performed in 1000 random-addition replicates using PAUP*. Maximum likelihood analyses were executed assuming a general time-reversible (GTR) model and a rate variation among sites following a gamma distribution (four categories represented by mean). GTR + Γ + I was chosen as the model that best fits the data, as evaluated before by Modeltest v3.6 (Posada and Crandall 1998). Employing the MTgui interface (Nuin 2005), settings proposed by Modeltest were exported to the PAUP* command file. ML bootstrap analysis (1000 replicates) was carried out as fast-heuristic search, employing the same settings as above. BI was performed using the program MrBayes v3.1 (Ronquist and Huelsenbeck 2003), applying the GTR + Γ + I model for sequence data and the restriction site model (“F81”) for the indel matrix. Four runs (10^6 generations each) with four chains each were run simultaneously, starting from random trees. Chains were sampled every 10th generation. Calculations of the consensus tree and the posterior probability (PP) of clades were done based upon the trees sampled after the burn-in set by default at 250,000 generations. Only PPs of 0.90 and higher were considered significant (alpha = 0.1). Trees were drawn using TreeGraph (Müller and Müller 2004).

Results

Sequence variability

The length of the five genomic regions studied here varies greatly (Table 2), and so do the amounts of length variability within individual spacers and introns. The *petB-D* spacer ranges from 174 to 226 nt, the *petD* intron from 639 to 799 nt, and the *trnL* intron from 303 to 643 nt, whereas the *trnL-F* spacer extends from 186 to 746 nt in length, thus displaying by far the greatest differences across taxa. The *matK* gene also exhibits significant length variation, ranging from 1494 to 1638 nt. Indels in *matK* correspond to codons, maintaining the open reading frame. Table 3 provides baseline sequence statistics for the spacers, introns, and the *matK* gene. Percentage of variable characters (substitutions) was highest in the *trnL-F* spacer, intermediate in the introns and the *petB-D* spacer. Variability of *matK* sequences equals the non-coding partitions concerning substitutions. Transition/transversion ratios mostly range from 1.1 to 1.3 although they are considerably higher in sequences of the *petD* intron. GC content, too, is highest in the *petD* intron (39%), it is lowest in the *petB-D* spacer (29.3%). Partition

Table 2. Actual lengths of genomic regions used, and positions of mutational hotspots in the respective sequences

Taxon	*petB-D* spacer	*petD* intron	*trnL* intron	*trnL-F* spacer	*matK* gene	Position H1 *petB-D*	Position H1 *petD*	Position H2 *petD*	Position H3 *petD*	Position H1 *trnL*	Position H2 *trnL*	Position H3 *trnL*	Position H1 *trnL-F*	Position H2 *trnL-F*	Position H3 *trnL-F*
Amborella trichopoda	224	733	474	375	1506	110–124	233–263	389–419	623–630	130–132	232–240	284–295	1–52	270–276	299–302
Nymphaea odorata ssp. *tuberosa*	204	639	520	380	1530	101–109	223–262	386–414	528–535	139–145	240–244	281–336	1–63	269–276	311–318
Austrobaileya scandens	176	710	475	390	1524	62–70	223–246	375–400	598–605	132–140	233–241	278–291	1–73	279–286	326–333
Ceratophyllum demersum	190	694	528	442	1545	90–98	238–268	393–421	578–590	133–141	242–256	306–351	1–101	303–313	367–374
Acorus calamus	190	726	520	377	1536	86–94	233–257	385–410	611–618	137–143	242–272	317–331	1–48	260–269	308–315
Chloranthus brachystachys	195	715	493	351	1524	87–95	231–254	386–406	604–611	135–139	249–257	294–311	1–40	241–248	287–294
Aristolochia pistolochia	200	699	510	372	1530	86–101	226–244	368–398	589–596	141–148	260–278	320–345	1–56	257–264	303–310
Magnolia officinalis	198	701	490	356	1524	86–94	227–250	373–398	590–597	132–138	240–258	295–313	1–51	245–252	292–299
Umbellularia californica	197	716	482	363	1524	86–94	242–260	384–409	601–608	132–143	241–254	291–310	1–47	252–259	299–306
Hedycarya arborea	198	706	481	388	1524	86–94	236–254	378–403	595–602	132–140	238–251	288–306	1–55	277–284	324–331
Chimonanthus praecox	198	698	477	328	1518	86–94	226–244	368–393	588–595	133–139	236–254	291–304	1–42	211–218	260–267
Euptelea pleiosperma	197	702	500	380	1524	86–94	226–250	375–400	592–599	136–142	244–262	299–317	1–52	247–254	299–306
Akebia quinata	213	709	503	371	1521	92–100	224–248	373–394	598–605	133–143	256–274	311–326	1–52	253–260	300–308
Dicentra eximia	213	709	474	359	1524	91–113	222–251	376–401	598–605	136–142	237–249	288–307	1–44	231–238	283–290
Papaver triniaefolium	191	718	519	363	1527	86–92	239–263	388–413	608–615	140–146	261–280	315–336	1–62	245–253	293–300
Cocculus laurifolius	220	702	490	386	1530	99–107	226–250	376–401	592–599	135–141	243–265	301–313	1–61	248–253	298–305
Stephania delavaji	223	704	501	378	1545	102–110	226–250	376–401	593–600	140–146	248–270	306–318	1–63	256–263	308–315
Xanthoriza simplicissima	193	728	501	345	1587	86–94	226–258	383–408	612–625	136–145	242–260	293–318	1–20	217–224	263–272
Mahonia japonica	197	690	479	746	1527	90–98	231–256	385–414	579–586	115–126	222–240	275–301	1–52	297–324	474–501
Podophyllum peltatum	219	737	466	387	1640	93–105	231–262	388–413	627–634	111–111	200–223	263–290	1–52	272–279	312–319
Sabia japonica	189	706	503	367	1536	83–91	223–247	372–397	595–602	132–139	248–266	308–326	1–52	250–257	297–304
Meliosma cuneifolia	193	714	513	380	1524	79–87	226–250	375–400	598–610	145–154	263–281	323–336	1–53	251–258	303–309
Nelumbo nucif ssp. *nucif*	194	719	524	402	1524	87–95	231–255	375–400	608–615	136–142	256–279	321–346	1–62	285–292	332–339
Nelumbo nucif ssp. *lutea*	192	718	525	401	1524	86–94	231–259	379–404	607–614	136–142	256–276	318–347	1–62	285–292	332–339
Embothrium coccineum	191	734	492	366	1530	84–92	244–268	393–418	623–630	136–138	241–259	306–323	1–39	235–242	296–303
Grevillea banksii	193	733	494	421	1530	86–94	242–267	392–417	622–629	141–143	243–261	308–325	1–48	304–312	352–359

Table 2. (*continued*)

Taxon	*petB-D* spacer	*petD* intron	*trnL* intron	*trnL-F* spacer	*matK* gene	Position H1 *petB-D*	Position H1 *petD*	Position H2 *petD*	Position H3 *petD*	Position H1 *trnL*	Position H2 *trnL*	Position H3 *trnL*	Position H1 *trnL-F*	Position H2 *trnL-F*	Position H3 *trnL-F*
Platanus orientalis	200	708	500	366	1539	93–101	226–250	375–400	598–605	121–126	235–254	296–323	1–39	235–242	296–303
Platanus occidentalis	200	709	523	366	1539	93–101	226–250	375–400	598–605	136–143	257–276	318–346	1–39	235–242	296–303
Tetracentron sinense	200	704	442	397	1516	86–103	221–240	365–394	593–600	136–145	246–246	249–265	1–52	265–272	311–318
Trochodendron aralioides	204	709	439	369	1516	93–105	226–250	375–400	598–605	135–143	244–244	247–262	1–57	269–276	315–322
Didymeles integrifolia	217	743	529	353	1524	103–118	226–251	376–400	602–640	136–142	255–277	319–395	1–41	237–244	283–290
Buxus sempervirens	193	726	505	378	1524	86–94	226–250	379–411	615–622	136–142	251–269	311–328	1–55	256–263	308–315
Pachysandra terminalis	193	704	507	370	1524	86–94	226–250	375–400	594–601	136–142	251–269	312–330	1–51	248–255	300–307
Gunnera tinctoria	196	721	511	359	1536	84–92	231–257	381–406	610–617	136–142	251–269	311–330	1–49	249–256	290–296
Myrothamnus flabellifolia	202	725	498	349	1590	93–103	226–250	374–399	598–621	136–142	244–262	304–327	1–50	231–245	280–286
Myrothamnus moschata	193	731	492	353	1530	86–94	226–250	374–406	605–628	136–142	244–256	298–321	1–50	231–246	284–290
Cercidiphyllum japonicum	198	716	507	356	1515	86–94	208–232	356–381	590–613	138–144	253–271	313–330	1–52	239–246	286–292
Chrysosplenium alternifolium	193	696	464	186	1530	86–94	215–239	359–378	585–592	139–147	242–242	270–287	1–44	84–87	127–130
Vitis riparia	188	734	517	323	1509	85–93	224–253	381–409	615–630	136–142	263–275	317–335	1–50	191–199	244–255
Leea coccinea	189	733	505	377	1506	86–94	226–255	383–408	615–630	136–142	251–263	305–323	1–51	260–264	304–310
Dillenia philippinensis	191	799	495	412	1527	84–92	234–267	391–416	617–685	132–140	243–261	306–317	1–51	278–285	331–337
Aextoxicon punctatum	193	716	509	355	1509	86–94	230–254	378–403	605–612	143–151	260–278	320–337	1–51	239–246	286–292
Osyris alba	193	726	528	376	1521	88–96	225–249	384–409	615–622	151–157	259–276	321–339	1–60	262–269	309–311
Rhipsalis paradoxa	212	789	643	365	1530	94–102	250–285	410–466	663–676	150–158	273–273	323–473	1–41	250–259	299–304
Spinacia oleracea	212	743	303	336	1518	99–107	240–258	382–438	629–634	119–127	167–167	167–167	1–42	205–209	251–256
Erodium cicutarium	220	734	496	369	1494	107–115	239–261	386–410	621–631	136–146	255–273	334–357	1–47	247–261	300–305
Coriaria myrtifolia	191	743	570	377	1521	86–92	225–256	380–420	631–640	132–137	246–264	306–378	1–50	259–266	306–314
Arabidopsis thaliana	188	709	311	343	1515	82–89	225–278	401–427	596–605	130–130	195–195	195–195	1–82	225–232	257–279
Oenothera elata	198	755	519	376	1539	85–93	241–265	389–412	621–651	139–147	258–271	332–348	1–58	259–278	318–324
Larrea tridentata	226	724	526	287	1515	105–112	234–276	400–419	614–621	146–151	261–279	315–354	1–51	151–157	205–225
Stachyurus chinensis	204	754	512	365	1515	91–99	224–263	387–418	640–651	136–142	251–269	311–335	1–58	232–247	287–297
Impatiens nolitangere	184	785	493	361	1509	76–84	234–277	405–433	653–671	145–151	255–267	300–316	1–49	254–260	302–308
Ilex aquifolium	194	720	491	361	1515	92–100	222–253	379–406	605–617	125–129	238–256	292–314	1–46	243–250	292–298
Atropa belladonna	190	742	496	362	1530	93–101	223–252	376–396	623–630	129–134	249–267	302–319	1–46	246–253	295–301
Nicotiana tabacum	190	742	502	356	1529	93–101	223–252	376–396	623–630	129–134	249–261	308–325	1–46	241–248	289–295
Panax ginseng	174	751	506	361	1512	81–89	241–272	395–433	640–647	139–140	249–267	309–326	1–47	240–249	291–297

Table 3. Variation and relative contribution (excluding mutational hotspots) of the five genomic regions studied

Parameter	Region				
	petB-D spacer	*petD* intron	*trnL* intron	*trnL-F* spacer	*matK* gene
Average sequence length (bp)	207	787	540	439	1528
Standard deviation	14	48	77	80	22
Average sequence length excluding hotspots (bp)	190	657	451	304	1528
Standard deviation	11	19	33	55	22
Number of characters	504	1162	908	1211	1857
% variable characters (corrected)	29.4 (77.9)	43.9 (77.6)	38.7 (77.8)	26.6 (105.9)	65.1 (79.1)
% informative characters (corrected)	18.3 (48.4)	30.1 (53.2)	26.7 (53.6)	20.0 (79.6)	50.8 (61.7)
Number of indels coded	121	257	244	356	109
% GC content	29.3	39.0	36.7	35.0	34.2
T_i/T_v ratio	1.079	1.632	1.335	1.230	1.338

Number and quality of characters, indels coded and GC content, as well as transition/transversion ratio all calculated for the sequence length with hotspots excluded. Due to high numbers of insertions characteristic to non-coding regions, underestimation of variable characters occurs. As a better approximation, amount of variability is also calculated on the average length of sequences (without hotspots), and shown as corrected.

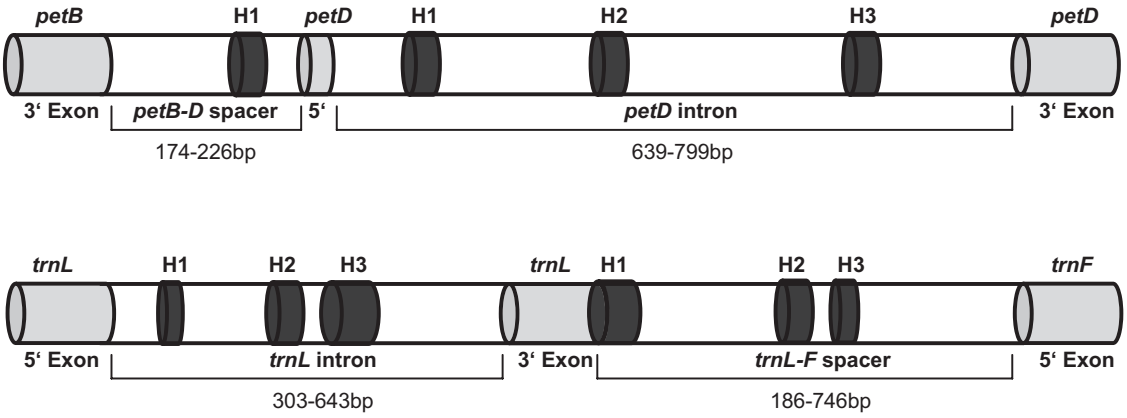

Fig. 1. Scheme of regions with position of mutational hotspots.

homogeneity tests indicate no significant level of heterogeneity between the three analysed regions (*petD* versus *trnL-F*: $P = 0.49$; *petD* versus *matK*: $P = 0.21$; *trnL-F* versus *matK*: $P = 0.81$).

Several short mutational hotspots have been identified in the spacers and introns (Fig. 1), in which an accurate primary homology assessment was not possible, either because of length-variable poly-A/T stretches (microsatellites) or difficulties in motif recognition as a result of frequent and overlapping microstructural mutations comprising several nucleotides. One mutational hotspot was determined in the *petB-D* spacer, three other hotspots in the *petD* intron. The *trnL* intron contains three mutational hotspots, as does the *trnL-F* spacer. Extension and absolute position (referring to nucleotide positions in the absolute sequence lengths starting at the 5′ end of a genomic region) of the corresponding hypervariable sequence parts constituting each hotspot are provided in Table 2. Sequence stretches within most hotspots are largely around 10–20 nt in length, in some taxa up to 50 nt. Hotspot H3 in the *trnL* intron is clearly the most variable, containing up to 100 nt in some taxa. Also, the first hotspot in *trnL-F*, comprising the 5′ end of the spacer, is somewhat more variable than the other hotspots (Table 2).

Length mutations occur frequently in all taxa and genomic regions studied, ranging from 1 to 150 nt but mostly consisting of single sequence repeats 4–6 nt in length. The combined indel matrix of all five regions comprises 1087 characters. Fig. 2 shows one of the most length-variable parts in the alignment, found in the *trnL-F* spacer (alignment positions 1092–1172). Several indels are synapomorphic for specific clades, such as indel number 306 which is a deletion unique to both species of *Myrothamnus*. Other indels are autapomorphic, though partly overlapping at one end. Examples are indels 248 and 308, which are independent deletions in *Chrysosplenium* and *Ceratophyllum*, respectively. Further prominent examples are a 150 nt insertion in the P8 loop of the *trnL* intron in *Rhipsalis* (not illustrated), and a 154 bp deletion in *Brassica*.

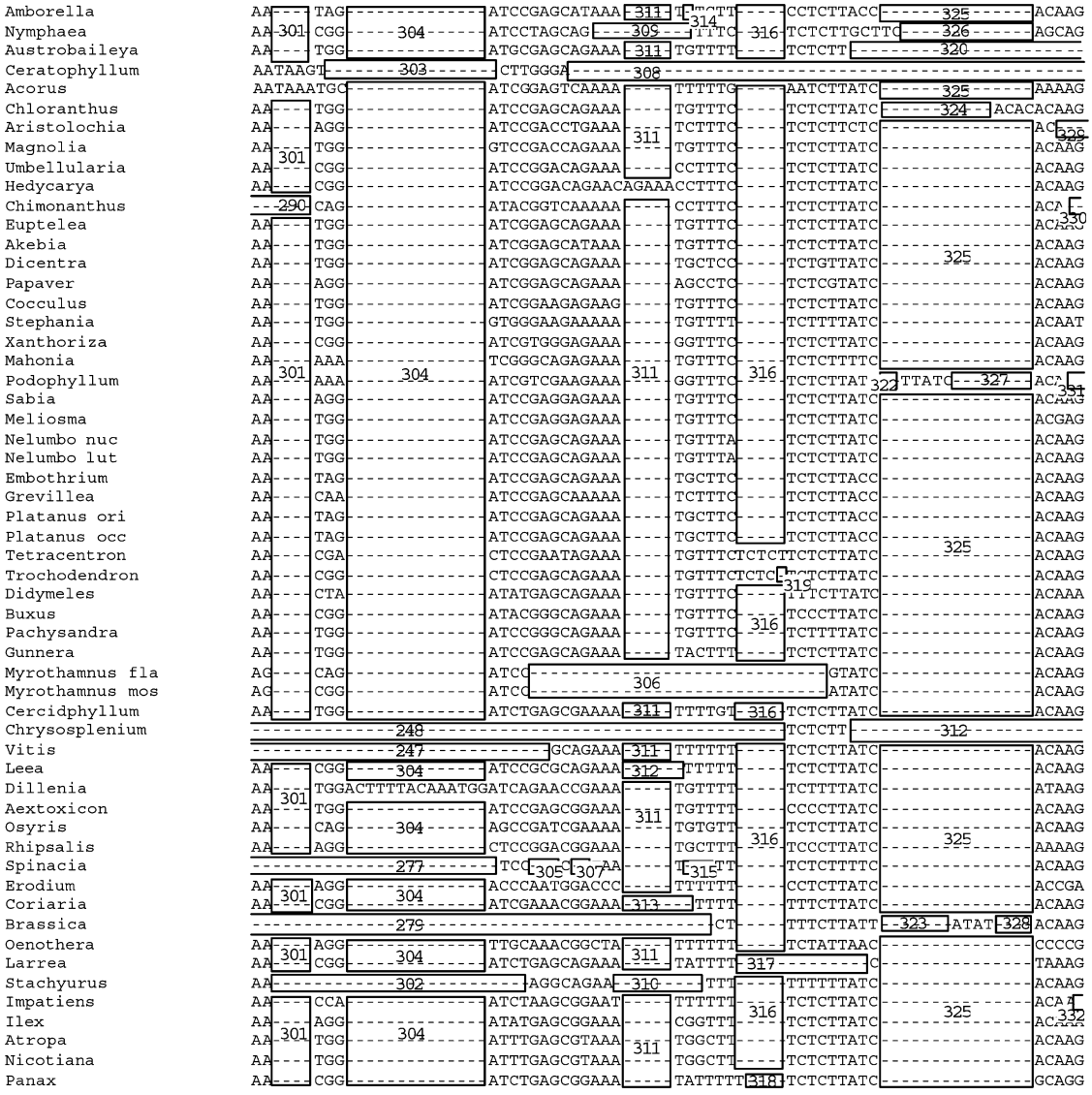

Fig. 2. Illustration of observed indels; example from the *trnL-F* region at positions 1092–1172. Simple Indel Coding after Simmons and Ochoterena (2000) using SeqState v1.25 (Müller 2005b). Indel number 306 is synapomorphic for the genus *Myrothamnus*, lacking in all other genera; Indel number 316 shared by *Trochodendron* and *Tetracentron*, which display a simple sequence repeat there (TCTCT) and one substituted point deletion (indel 319) for *Trochodendron* is missing in all other taxa studied.

Phylogeny of basal eudicots

The combined data matrix (*petD* + *trnL-F* + *matK*) provided 5654 characters (excluding mutational hotspots). Of these characters 2542 were variable and 1869 were parsimony informative. In addition, 1087 binary indel characters were added to the dataset. Relative contributions of individual genomic regions are shown in Table 3. Maximum parsimony analysis of the combined dataset resulted in one shortest tree of 12,363 steps (CI = 0.458, RI = 0,466; Fig. 3). Ranunculales gains high support (100 JK, 10 DI) as sister to the remainder of eudicots, followed by Sabiales (83 JK, 2 DI), and Proteales including Nelumbonaceae (100 JK, 14 DI). Branching next are Trochodendrales (91 JK, 5 DI), followed by Buxales including Didymelaceae. Core eudicots are strongly supported (100 JK, 52 DI), with Gunnerales as the first-branching core eudicot clade (100 JK, 7 DI). The backbone of core eudicots is resolved, but support stays low. Nevertheless, major lineages such as Vitales (100 JK, 69 DI), Saxifragales (95 JK, 11 DI), Caryophyllales (100 JK, 118 DI), rosids (96 JK, 20 DI), and asterids (96 JK, 11 DI) are identified with high confidence. Inside Ranunculales, Eupteleaceae are branching off first (81 JK, 3 DI), followed by Papaveraceae s.l. (100 JK, 13 DI). Sabiales are clearly

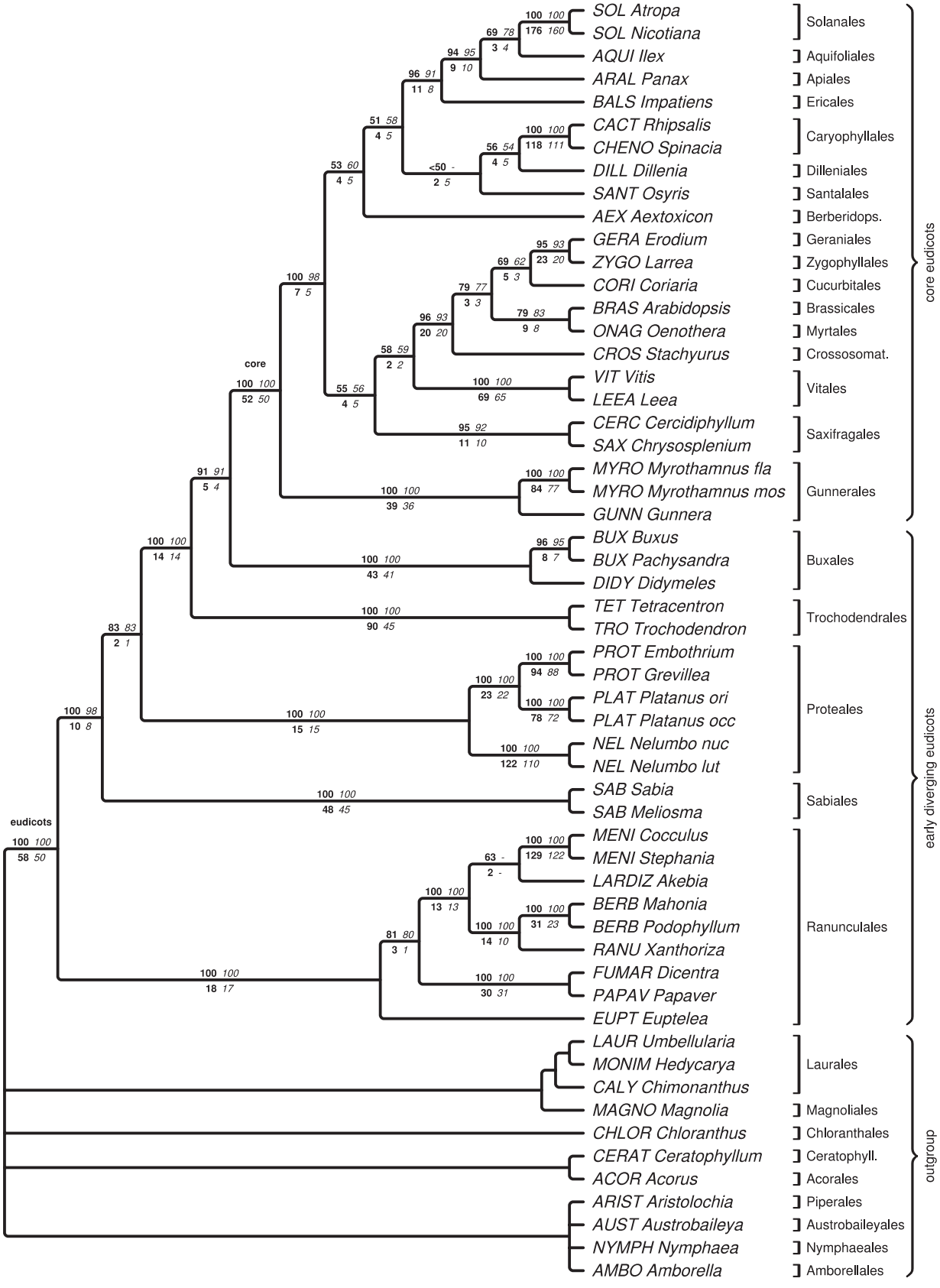

Fig. 3. Combined tree based on substitutions and indels of all five regions, inferred with MP. Values above branches are Jackknife percentages, and below are decay values. Bold figure refer to substitutions plus indels, italics to substitutions only.

monophyletic (100 JK, 48 DI). Proteales are strongly supported as monophyletic (100 JK, 15 DI), with Nelumbonaceae as sister to a Platanaceae + Proteaceae clade (100 JK, 23 DI). Trochodendrales (100 JK, 90 DI) and Buxales (100 JK, 43 DI) are monophyletic, with Didymelaceae resolved as sister to Buxaceae (96 JK, 8 DI) (Fig. 4).

BI resulted in one tree that is similar in topology to the MP tree, except for the branching of Sabiales after Proteales with no support (0.52 PP). The ML analysis

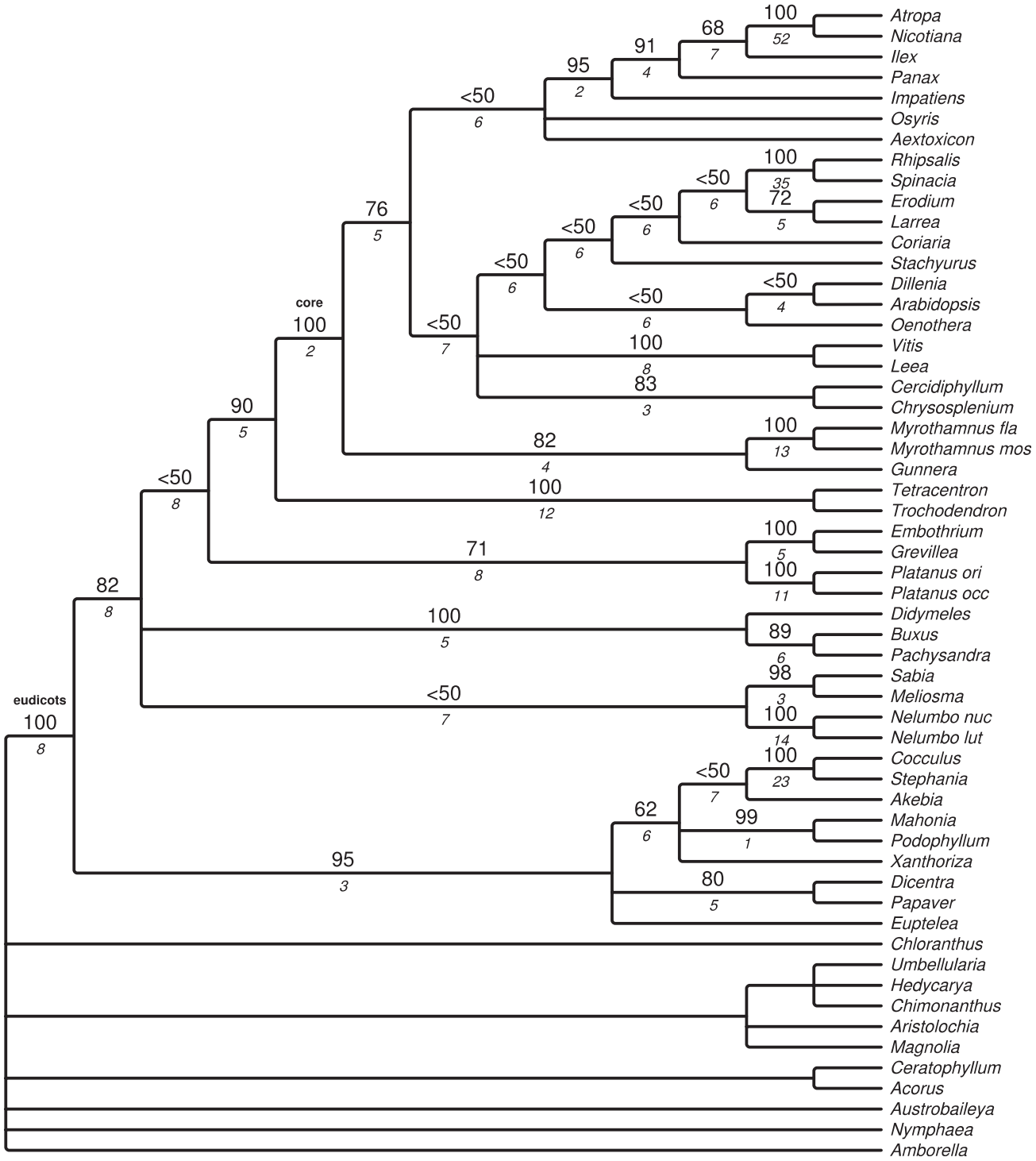

Fig. 4a. Phylogenetic utility of individual markers (substitutions and indels). Strict consensus tree inferred from *petD* (incl. *petB-D*; 128 shortest trees of 2983 steps, CI = 0.515, RI = 0,492).

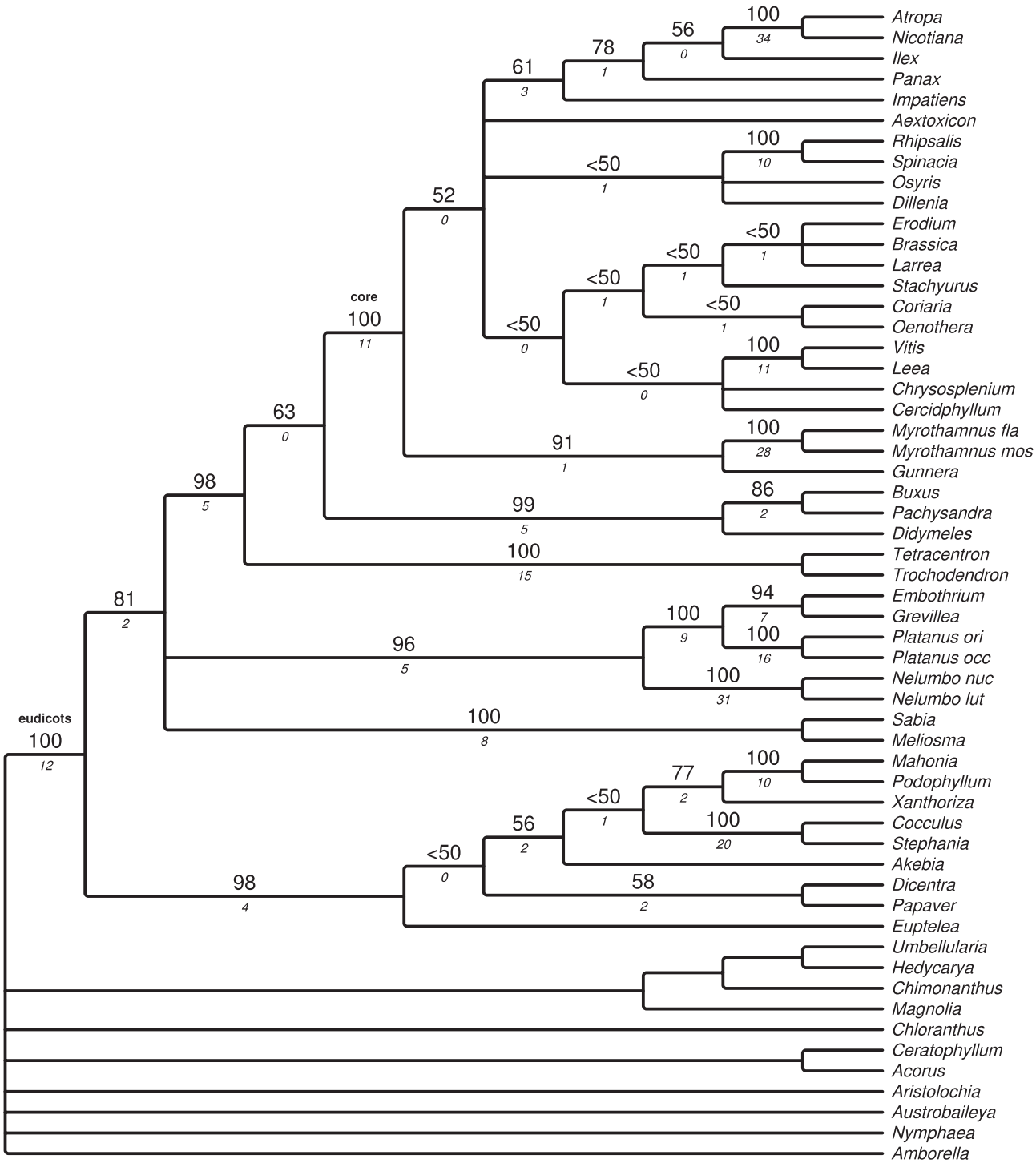

Fig. 4b. Strict consensus tree inferred from *trnL-F* (139 shortest trees of 3547 steps, CI = 0.521, RI = 0.456).

resulted in a tree (−ln 57250.56886) similar in topology to the MP and Bayesian analyses, but resolution for the respective positions of Sabiales and Proteales was lacking (Fig. 5). Most nodes that gained ML bootstrap values were significantly (PP < 0.90) supported in the Bayesian tree, too. Some major clades, such as Vitales as sister to rosids, lack bootstrap support in the ML analysis but are well supported with BI (highlighted by asterisks in Fig. 5).

Discussion

Relationships among first-branching eudicots

Phylogenetic analyses from recent years (Hoot et al. 1999; Hilu et al. 2003; Soltis et al. 2003; Kim et al. 2004) have provided a framework of relationships among the first-branching eudicots. Nevertheless, the

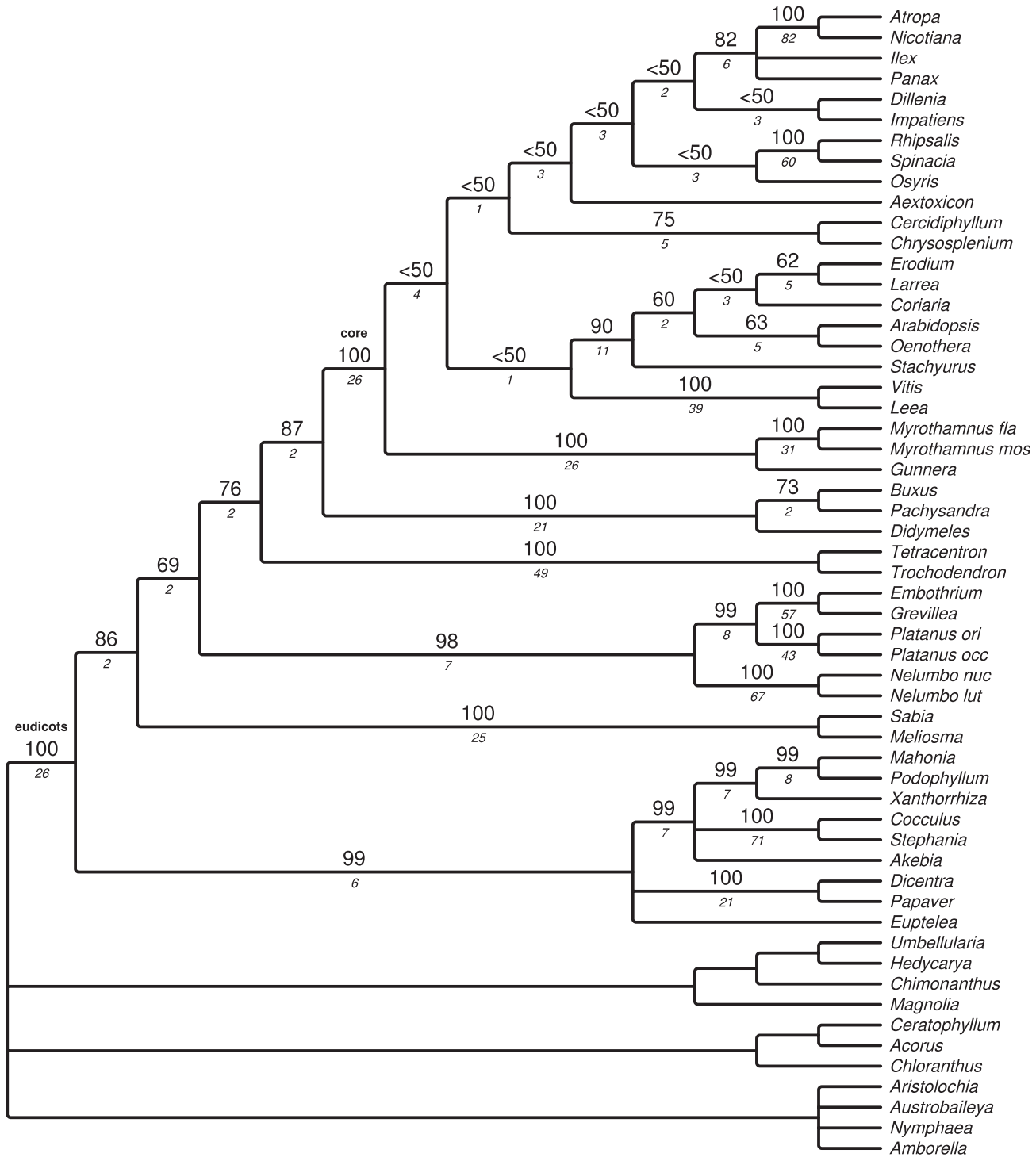

Fig. 4c. Strict consensus tree inferred from *matK* (6 shortest trees of 5800 steps, CI = 0.393, RI = 0,465).

exact branching order above Ranunculales remained to be substantiated. The analysis of partial *matK* sequences (Hilu et al. 2003) provided 91% JK for a clade of Buxaceae–Didymelaceae and core eudicots. Since none of the other analyses yielded good support for any position of this clade, the hypothesis of Buxaceae–Didymelaceae being sister to the core eudicots was among the prominent issues to be tested here.

In order to make sampling for the *matK* analysis more representative than in Hilu et al. (2003) and comparable to analyses of slowly evolving genes, the present study includes *Papaver* to complement *Dicentra* (Papaveraceae s.l.), *Sabia* to complement *Meliosma* (Sabiaceae), the second subsp. of *Nelumbo* (Nelumbonaceae), a second species of *Platanus* (Platanus), and the two species of *Myrothamnus* (Myrothamnaceae) to complement

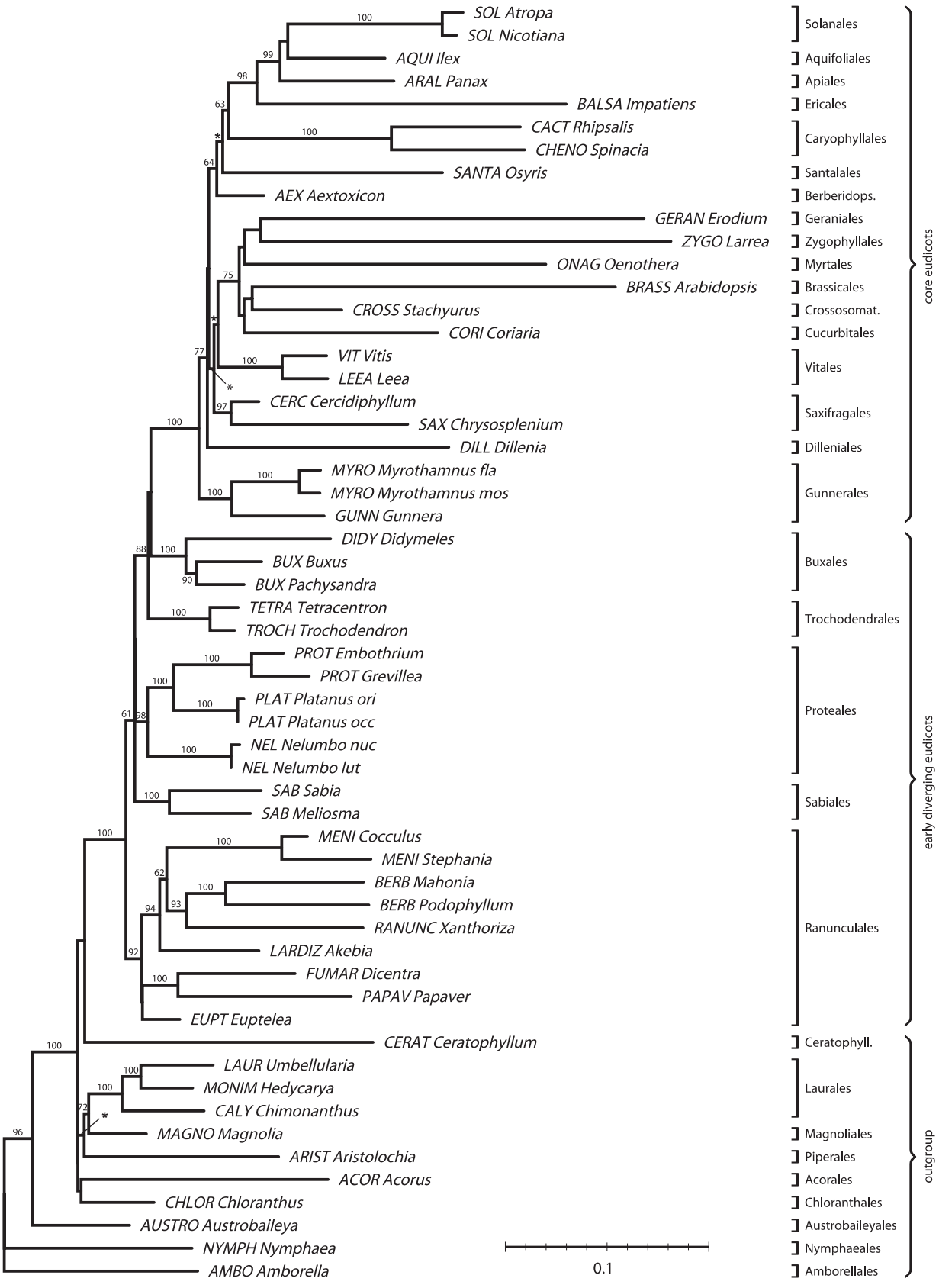

Fig. 5. Maximum likelihood phylogram (−ln 57250.56886) based on the combined *petD*+ *trnL-F*+*matK* matrix (substitutions only). Bootstrap values are depicted above branches. Clades that gained no ML bootstrap support but significant PP (>0.90) in Bayesian Inference are marked by an asterisk.

Gunnera (Gunneraceae) in Gunnerales. The tree based on complete *matK* sequences (Fig. 4c) shows resolution comparable to the tree inferred from partial *matK* (starting from about position 500 downstream of the start codon; Hilu et al. 2003), but using complete *matK* sequences significantly increased support for most nodes. The 5′ region of *matK* was considered as the most variable part of the gene (Hilu and Liang 1997). The only topological difference in the *matK* tree between this study (Fig. 4c) and Hilu et al. (2003) concerns the position of Vitaceae in core eudicots. They appear as sister to rosids in this study, in accordance with the three- and four-gene analyses of Soltis et al. (2000, 2003), whereas partial *matK* depicted them with low support as sister to Dilleniaceae.

Close relationships between Buxaceae and Didymelaceae are generally accepted based on molecular and morphological characters (Nandi et al. 1998; von Balthazar et al. 2000; von Balthazar and Endress 2002), although the endemic Madagascan, dioecious and nearly perianthless genus *Didymeles* has been considered as an isolated lineage (Cronquist 1981). The earlier observation that *matK* resolves Buxaceae as paraphyletic to *Didymeles* (Hilu et al. 2003) cannot be upheld. In the 2003 analysis, the *matK* fragment from *Didymeles* used was only 900 nt long. To test the resulting tree, we generated a complete *matK* sequence from a different individual, which required several *Didymeles*-specific internal sequencing primers due to a high number of autapomorphies. The new sequence resolves Didymelaceae as sister to Buxaceae (Fig. 4c); the latter are supported as monophyletic with 73% JK. Comparison of both sequences indicates that both are *Didymeles* (presence of characteristic autapomorphies), although they differ by 5 substitutions. Reanalysis using the earlier partial sequence in this dataset again resulted in paraphyly of Buxaceae. An explanation could be that high lineage-specific variability in the short *matK* fragment leads to a spurious position of *Didymeles*. In some angiosperms, translocated paralogous *trnK* intron copies have been found (*Nepenthes*, Meimberg et al. 2006; *Peperomia,* Wanke et al. 2007) that can be identified as non-functional, based on many indels not being multiples of three nucleotides within the *matK* coding region. We amplified the *trnK* intron in *Didymeles* in two halves, found only one amplification product through gel electrophoresis, and did not observe overlying signal after direct sequencing. Given that a correct reading frame is present in our *matK* sequence, we regard this sequence as orthologous and functional.

Combining *petD* + *trnL-F* + *matK* sequences, and including indel information, for the first time provides a fully resolved and well-supported topology of the basal eudicot grade using parsimony (Fig. 3). The position of Ranunculales as sister to all remaining eudicots gains maximum support (MP 100% JK; BI 1.00 PP). First-branching Ranunculales were congruently inferred by the 4-gene analyses of Soltis et al. (2003; 87% JK) and Kim et al. (2004), although the latter surprisingly found no support. Partial *matK* sequences (Hilu et al. 2003) yielded 82% JK, and similar values of medium support are also achieved by each of the three individual partitions, *petD*, *trnL-F* and *matK*, in the present study (Fig. 4a–c). Thus, maximum statistical support values in the combined analysis of this study can be explained as resulting from an additive effect, because individual partitions do not provide sufficient amounts of information although their phylogenetic signal favours the same nodes.

Early phylogenetic analyses (Drinnan et al. 1994; Hoot and Crane 1995; Loconte et al. 1995) indicated that most families of Ranunculales belong to a core clade from which Eupteleaceae and Papaveraceae sensu lato (incl. Fumariaceae, *Hypecoum*, *Pteridophyllum*; Kadereit et al. 1995) are excluded. Partial *matK* data in the study of Hilu et al. (2003) resolved Eupteleaceae at the base of a strongly supported Ranunculales clade, followed by Papaveraceae and the remaining Ranunculales. However, support for the positions of Eupteleaceae and Papaveraceae was lacking. Using four genes, Kim et al. (2004) could increase confidence in the hypothesis that Eupteleaceae are first-branching in Ranunculales (70% JK), followed by Papaveraceae (78% JK). The combined analysis of *petD* + *trnL-F* + *matK* data yields 81% and 100% JK for the respective nodes. An alternative hypothesis that assumes Papaveraceae as sister to all remaining Ranunculales (Soltis et al. 2000), can be rejected. Increased confidence in the first-branching position of Eupteleaceae is also relevant to inferring an ancestrally woody condition in Ranunculales and eudicots (Kim et al. 2004), contrary to Cronquist's (1981) hypothesis of Ranunculales as primitively herbaceous. In our study, Lardizabalaceae and Menispermaceae form a weakly supported clade, indicating that the climbing habit predominant in these two families arose only once. Results of Soltis et al. (2000) and Kim et al. (2004) differ in showing a Lardizabalaceae + *Sargentodoxa* clade in a tritomy with Circaeasteraceae and the other core Ranunculales families, or a Lardizabalaceae–Circaeasteraceae clade as sister to all remaining core Ranunculales. The respective nodes are only weakly or moderately supported both in our and the other studies. Sampling additional taxa in Ranunculales with fast-evolving and non-coding markers (e.g. *Circaeaster*, *Kingdonia*, *Sargentodoxa*) is needed.

Sabiaceae (*Meliosma* and *Sabia*) were inferred as monophyletic based on *rbcL* and morphological data (Nandi et al. 1998) although the third genus, *Ophiocaryon*, has never been included in any phylogenetic analysis. Monophyly of a *Meliosma–Sabia* lineage is substantiated by our study, whereas the earlier

broad-scale analysis of *matK* (Hilu et al. 2003) included only *Meliosma*. The combined parsimony tree of *petD* + *trnL-F* + *matK* reveals Sabiales as the second branch in the basal eudicot grade (83% JK), in accordance with the tree found by Hilu et al. (2003), albeit with this node unsupported. Signal for the second-branching position of Sabiales comes from complete *matK*, whereas the *petD* and *trnL-F* partitions are inconclusive (Fig. 4). The anticipated position of Sabiales near Trochodendraceae or Buxaceae as proposed by Kim et al. (2004) seems unlikely.

The traditional classification of *Nelumbo* (Nelumbonaceae) within Nymphaeales (water-lilies; Cronquist 1988) was challenged by analyses of *rbcL* sequences (Chase et al. 1993) that suggested a *Nelumbo–Platanus* sister group, and of epicuticular waxes (Barthlott et al. 1996) that also indicated affinities to basal eudicots. Since then, multi-gene studies have refined the hypothesis to *Nelumbo* being sister to Platanaceae–Proteaceae (Hoot et al. 1999; Soltis et al. 2000; Kim et al. 2004). This clade of three lineages was classified as Proteales by APG II (2003) but gained only 62% BS in three-gene analyses (Hoot et al. 1999; Soltis et al. 2000) and 65% JK in the four-gene analysis of Kim et al. (2004), whereas confidence in a Platanaceae–Proteaceae sister group was high. Partial *matK* data (Hilu et al. 2003) yielded 64% JK for Proteales. In the present study, the Proteales clade for the first time receives 96–100% JK support from the *trnL-F* and *matK* partitions (Figs. 4b and c) and the combined tree (Fig. 3; 100% JK, 1.00 PP). The affinities of the three families Platanaceae–Proteaceae and Nelumbonaceae are therefore substantiated despite the lack of clear morphological synapomorphies. Savolainen et al. (2000b), in analysing a large matrix of *rbcL* sequences from eudicots, resolved *Nelumbo* as sister to Sabiaceae, albeit without support – an inconsistent topology also found unsupported by the *petD* partition.

Trochodendraceae and Tetracentraceae were considered as close relatives based on morphological characters (Endress 1986; Drinnan et al. 1994; Endress and Igersheim 1999), and resolved as sister groups in all previous phylogenetic analyses, with varying confidence (Hilu et al. 2003; Soltis et al. 2003; Kim et al. 2004). Trochodendraceae and Tetracentraceae both share the lack of vessels as a prominent synapomorphy, a feature which now is understood as secondarily derived (Doyle and Endress 2000). In this study, the *Tetracentron–Trochodendron* clade stands out by a high number of synapomorphic indels (Figs. 3 and 4). This is evidenced by the increase of decay values from 45 to 90 after adding the indel matrix (Fig. 3).

Analysis of partial *matK* sequences (Hilu et al. 2003) yielded 91% JK for Buxaceae–Didymeleaceae as sister to core eudicots, whereas previous analyses were inconclusive about the respective positions of this lineage and the *Tetracentron–Trochodendron* clade. An early, morphology-based cladistic analysis inferred the *Tetracentron–Trochodendron* clade as sister to the remaining eudicots (Hufford and Crane 1989). Signal from complete *matK* and *trnL-F* agrees on Buxales as sister to core eudicots (87% and 63% JK, respectively), whereas the *petD* partition is incongruent (90% JK for *Tetracentron–Trochodendron* as sister to core eudicots). Nevertheless, the combined analysis gave 87% JK for this position. Further sequence data are needed to clarify the situation.

The divergence of Gunnerales next after the basal eudicot grade (as sister to all remaining core eudicots) was recently hypothesized (84% JK) by adding nr26S sequences to a *rbcL* + *atpB* + 18S dataset (Soltis et al. 2003). Chloroplast data (Savolainen et al. 2000a; Hilu et al. 2003) have not provided significant support for this position of Gunnerales. Adding *trnL-F* and *petD* sequences to a complete *matK* dataset clearly substantiates Gunnerales as sister to the remaining core eudicots (100% JK and 1.00 PP). As the first branch of core eudicots, the Gunnerales play an important role in understanding eudicot floral diversification. The perianth of *Gunnera* is dimerous, the perianth of *Myrothamnus* dimerous or labile (Endress 1989; Drinnan et al. 1994). Using their phylogenetic hypothesis to reconstruct perianth merosity, Soltis et al. (2003) demonstrated that the pentamerous condition characteristic of core eudicots must have originated after the divergence of the Gunnerales lineage. More recently, Wanntorp and De Craene (2005) argued that *Gunnera* floral morphology is reduced in response to wind pollination. A well-resolved and supported phylogeny of basal eudicots is important for all studies aiming at understanding the evolution of floral characters, because merosity in fact is highly variable among early-diverging eudicots (Endress 1996; Drinnan et al. 1994). It ranges from dimerous (most Papaveraceae, *Glaucidium, Hydrastis, Sanguinaria*) through trimerous (Berberidaceae, Lardizabalaceae, *Sargentodoxa*, *Circaeaster*, Menispermaceae), tetramerous (Platanaceae), and pentamerous (*Meliosma* and *Sabia*, Ranunculaceae) to absence of a perianth in *Trochodendron* (Endress 1986).

An improved understanding of basal eudicot relationships will also help to clarify classification at the ordinal level. Unresolved or unsupported relationships among families will leave open alternative hypotheses for possible sister groups, and thus hinder decisions to classify more than one family into a monophyletic order. Although Takhtajan (1997) proposed the order Sabiales, it was difficult to apply because a possible sister-group relationship of Sabiaceae to Proteales could not be excluded, leaving the option of including Sabiaceae into Proteales. The family Sabiaceae was not classified in any order by APG II (2003). Increased evidence for Proteales branching next after *Sabia* + *Meliosma* in a

grade based on combined *matK*+*trnL-F*+*petD* may justify recognition of Sabiales (Fig. 3). Moreover, we recognize Trochodendrales, first proposed by Cronquist (1981), because the branching order of *Tetracentron*+*Trochodendron* prior to Buxaceae+Didymelaceae in the eudicot basal grade is well supported.

Molecular evolution and phylogenetic utility of genomic regions studied

The three partitions, *petD*, *trnL-F*, and *matK*, provide congruent signal for hypotheses on basal eudicot relationships. The only topological differences are in parts of the tree that are weakly supported and regarded as inconsistent rather than incongruent. It is worth noting that small regions such as *trnL-F* or *petD*, with average sequence lengths excluding mutational hotspots of 755 or 840 nt, respectively (Table 3), are resolving most of the eudicot topology. This compares to the *rbcL* gene that is comprised of roughly 1400 nt. Congruence of trees obtained from the three character partitions may be a further indication that analysis of non-coding regions does not show spurious relationships.

Comparing the three partitions, highest length variability occurs in the *trnL-F* spacer and the *trnL* intron, both in absolute terms and with respect to size and frequency of indels present in the alignment (Table 3). This trend to high sequence variation within the *trnL-F* region, with the *trnL-F* spacer being the most dynamic in terms of length mutations in eudicots, is in accordance with observations already made on basal angiosperms (Borsch et al. 2003). The *trnT-L* spacer was not included in the present study because of large insertions in several taxa (within hotspot H1 of basal angiosperms; Borsch et al. 2003) that required additional internal primers for complete sequencing, resulting in comparatively high laboratory effort. As in the case of basal angiosperms (Löhne and Borsch 2005), the *petD* intron could be amplified and completely sequenced easily using universal primers in eudicots.

Mutational hotspots in non-coding regions have been shown to correspond to certain stem-loop elements of the secondary structures where constraints are expected to be lowest (Borsch et al. 2003; Quandt et al. 2004; Löhne and Borsch 2005). This raises the question whether similar hotspots can be found in more derived angiosperms like eudicots, too. In this study (Fig. 1) we numbered mutational hotspots (HS) individually for each spacer or intron to facilitate future comparisons across angiosperms. In the *petB-D* spacer, HS1 is a microsatellite that extends up to 20 As in individual taxa. A length-variable satellite was not present in basal angiosperms (Löhne and Borsch 2005). Examination of the basal angiosperm alignment shows that nucleotide substitutions in eudicots must have led to longer A/T-homonucleotide stretches, with increased probability for slipped strand mispairing. Such patterns were assumed for the emergence of microsatellites (Levinson and Gutman 1987), as mutational rates in satellite regions generally increase with their length. Microsatellites are well defined and can be excluded easily from phylogenetic analyses of more distant sequences. The first hotspot in the *petD* intron recognized in this basal eudicot dataset is located in the D2 loop of domain I, and was also found in basal angiosperms (HS2; Löhne and Borsch 2005). Due to increased variability in asterids, this hotspot is extended around 30 positions downstream in eudicots. HS2 (Fig. 1) is located in the domain II stem loop, which is not highly length-variable in basal angiosperms. Hotspot HS3 of the eudicot dataset corresponds to HS4 in basal angiosperms, and is located in the terminal stem loop of domain IV.

In the *trnL* intron, HS1 is a microsatellite similar to HS1 of the *petB-D* spacer, and it was not variable in basal angiosperms; consequently, it was not recognized in Borsch et al. (2003). Hotspots HS2 and HS3 correspond to the terminal stem-loop parts of the P6 and P8 elements in the group I intron's secondary structure that are generally least constrained (Borsch et al. 2003; Quandt et al. 2004; Quandt and Stech 2005). Extraordinary lengths of the *trnL* intron in some taxa, such as in *Rhipsalis* (Cactaceae; 643 nt), are due to big inserts in the P8 stem loop. These seem to be the result of a mechanism by which slippage leads to the accumulation of small repeats in a satellite-like manner, which later become more GC-rich due to subsequent substitutions (Quandt et al. 2004). Because these terminal parts of the P8 stem loop are not homologous across angiosperms they need to be excluded as a mutational hotspot (HS3 in Fig. 3). Sequence divergence (numerous overlapping duplication and deletion events) at the 5′ end of the *trnL-F* spacer is high in eudicots; thus a comparatively large hotspot (HS1) needs to be recognized. On the contrary, HS2 and HS3 are microsatellites (poly-A/Ts); the former is also present in basal angiosperms (Borsch et al. 2003). In *Chrysosplenium* almost the whole spacer is deleted, with the deletion ending approximately 80 bp upstream of the −35 promoter element in front of *trnF*. By comparing slowly evolving protein-coding and non-coding cp regions, it has been shown recently that the rapidly evolving regions may not only exhibit a higher proportion of parsimony informative sites but also more phylogenetic structure per informative site (Müller et al. 2006).

Conclusions

Resolving the branching order among basal eudicots has remained one of the major challenges in angiosperm phylogeny, despite the existence of analyses using

multiple coding genes with more than 8 kb of sequence. The analysis of five non-coding and fast-evolving genomic regions now provides high resolution and statistical support for the basal eudicot grade. In addition to basal angiosperms, for which similar observations have been made (Borsch et al. 2003; Hilu et al. 2003; Löhne and Borsch 2005; Müller et al. 2006), this is another example of the phylogenetic utility of rapidly evolving non-coding genomic regions for deep-level phylogenetics. It is worth noting that signal from the individual partitions is congruent, rather than randomized through saturation as was frequently assumed earlier (e.g. Moritz et al. 1987; Simon et al. 1994; Graham and Olmstead 2000). Moreover, it is important to note that *trnL-F* and *petD* contribute significant information, and that combining these non-coding sequence data with *matK* can lead to further advances in phylogeny reconstruction. Combining molecular datasets has been well established since earlier works, e.g. Soltis et al. (1998). Nevertheless, current data suggest that addition of any available marker with the aim to just raise the amount of sequence characters may not be the most efficient solution. Combining markers selected for their levels of phylogenetic structure may be a perspective for other difficult-to-resolve nodes in angiosperms, and in particular where dense taxon sampling is required. Molecular evolution of spacers and introns of groups I and II seems to follow general patterns in angiosperms, as indicated by the presence of highly dynamic sequence parts (mutational hotspots) in always similar, obviously structurally caused locations. Complementary analyses of additional group II introns and spacers, which are present more frequently in the plastid genome, will thus be rewarding, in order to generalize ideas on their mutational patterns.

Acknowledgements

We acknowledge support from the Deutsche Forschungsgemeinschaft (DFG) for the project "Mutational dynamics of non-coding genomic regions and their potential for reconstructing eudicot evolution" (grants BO1815/2 to T.B. and QU153/2 to D.Q.). Thanks are due to the DFG for a Heisenberg scholarship to T.B. that has greatly facilitated this work. K.W.H. acknowledges the US National Science Foundation for award 0431266. We are grateful to the CEMBIO network (Bonn) for supporting research in molecular systematics. Most of the material was provided by the living collections of the Botanical Gardens of Bonn University and Dresden Technical University. We thank the Missouri Botanical Garden for providing material of *Didymeles integrifolia*, and the Botanical Gardens of the Ruhr-University Bochum for *Meliosma cuneifolia*. Yin-Long Qiu (Ann Arbor) supplied DNA of *Sabia japonica*, Keiko Kosuge (Kobe University) an unpublished *matK* sequence of *Podophyllum*. For various kinds of support and helpful discussions we are grateful to W. Barthlott (Bonn), K. Govers (Bonn), K. Müller (Bonn), C. Neinhuis (Dresden), B. Schäferhoff (Bonn), and to P. and D. Soltis (Gainesville). Helpful comments of two anonymous reviewers are gratefully acknowledged. This study fulfils part of the requirements for the first author obtaining a doctoral degree from Bonn University.

Appendix A. Supplementary Materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ode.2006.08.001.

References

Aagesen, L., 2004. The information content of an ambiguously alignable region, a case study of the *trnL* intron from the Rhamnaceae. Org. Divers. Evol. 4, 35–49.

Anderson, C.L., Bremer, K., Friis, E.M., 2005. Dating phylogenetically basal eudicots using *rbcL* sequences and multiple fossil reference points. Am. J. Bot. 92, 1737–1748.

APG II, 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Bot. J. Linn. Soc. 141, 399–436.

Azuma, H., Thien, L.B., Kawano, S., 1999. Molecular phylogeny of *Magnolia* (Magnoliaceae) inferred from cpDNA sequence and evolutionary divergence of the floral scents. J. Plant Res. 112, 291–306.

von Balthazar, M., Endress, P.K., 2002. Development of inflorescences and flowers in Buxaceae and the problem of perianth interpretation. Int. J. Plant Sci. 163, 847–876.

von Balthazar, M., Endress, P.K., Qiu, Y.L., 2000. Phylogenetic relationships in Buxaceae based on nuclear internal transcribed spacers and plastid *ndhF* sequences. Int. J. Plant Sci. 161, 785–792.

Barthlott, W., Neinhuis, C., Jetter, R., Bourauel, T., Riederer, M., 1996. Waterlily, poppy, or sycamore: on the systematic position of *Nelumbo*. Flora 191, 169–174.

Benson, G., 1997. Sequence alignment with tandem duplication. J. Comp. Biol. 4, 351–367.

Borsch, T., Hilu, W., Quandt, D., Wilde, V., Neinhuis, C., Barthlott, W., 2003. Non-coding plastid *trn*T-*trn*F sequences reveal a well resolved phylogeny of basal angiosperms. J. Evol. Biol. 16, 558–576.

Borsch, T., Löhne, C., Müller, K., Hilu, K.W., Wanke, S., Worberg, A., Barthlott, W., Neinhuis, C., Quandt, D., 2005. Towards understanding basal angiosperm diversification: recent insights using rapidly evolving genomic regions. Nova Acta Leopoldina NF 92/342, 85–110.

Borsch, T., Hilu, K., Wiersema, J., Löhne, C., Barthlott, W., Wilde, V., in press. Phylogeny of *Nymphaea* (Nymphaea-

ceae): evidence from substitutions and microstructural changes in the chloroplast *tmT-trnF* region. Int. J. Plant Sci.

Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., Duvall, M.R., Price, R.A., Hillis, H.G., Qiu, G.Y.Y., Kron, K.A., Rettig, J.H., Conti, E., Palmer, J., Manhart, J.R., Sytsma, K.J., Michaels, H.J., Kress, W.J., Karol, K.G., Clark, W.D., Hedren, M., Gaut, B.S., Jansen, R.K., Kim, K.J., Wimpee, C.F., Smith, J.F., Furnier, G.R., Strauss, S.H., Xiang, Q.Y., Plunkett, G.M., Soltis, P.S., Swensen, S.M., Williams, S.E., Gadek, P.A., Quinn, C.J., Eguiarte, L.E., Golenberg, E., Learn, G.H.J., Graham, S.W., Barret, S.C.H., Dayanandan, S., Albert, V.A., 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Ann. Mo. Bot. Gard. 80, 528–580.

Cronquist, A., 1981. An Integrated System of Classification of Flowering Plants. Columbia University Press, New York.

Cronquist, A., 1988. The Evolution and Classification of Flowering Plants, 2nd ed. New York Botanical Gardens, New York.

Donoghue, M.J., Doyle, J.A., 1989. Phylogenetic analysis of angiosperms and the relationships of the Hamamelidae. In: Crane, P.R., Blackmore, S. (Eds.), Evolution, Systematics and Phylogeny of the Hamamelidae. Clarendon Press, Oxford, pp. 17–45.

Doyle, J., Hotton, C.L., 1991. Diversification of early angiosperm pollen in a cladistic context. In: Blackmore, S., Barnes, R.W. (Eds.), Pollen and Spores. Systematics Association Special volume 44. Clarendon Press, Oxford, pp. 169–195.

Doyle, J.A., Endress, P.K., 2000. Morphological phylogenetic analysis of basal angiosperms: comparison and combination with molecular data. Int. J. Plant Sci. 161, S121–S153.

Drinnan, A.N., Crane, P.R., Hoot, S.B., 1994. Patterns of floral evolution in the early diversification of non-magnoliid dicotyledons (eudicots). Plant Syst. Evol. Suppl. 8, 93–122.

Endress, P.K., 1986. Floral structure, systematics, and phylogeny in Trochodendrales. Ann. Mo. Bot. Gard. 73, 297–324.

Endress, P.K., 1989. Chaotic floral phyllotaxis and reduced perianth in *Achlys* (Berberidaceae). Bot. Acta 102, 159–163.

Endress, P.K., 1996. Homoplasy in angiosperm flowers. In: Sanderson, M.J., Hufford, L. (Eds.), Homoplasy: The Recurrence of Similarity in Evolution. Academic Press, San Diego, CA, pp. 303–325.

Endress, P.K., Igersheim, A., 1999. Gynoecium diversity and systematics of the basal eudicots. Bot. J. Linn. Soc. 130, 305–393.

Freudenstein, J.V., Simmons, M.P., 2004. Relative effects of increasing genetic distance on alignment and phylogenetic analysis. Cladistics 20, 83.

Golenberg, E.M., Clegg, M.T., Durbin, M.L., Doebly, J., Ma, D.P., 1993. Evolution of a non-coding region of the chloroplast genome. Mol. Phylogenet. Evol. 2, 52–64.

Graham, S.W., Olmstead, R.G., 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. Am. J. Bot. 87, 1712–1730.

Graham, S.W., Reeves, P.A., Burns, A.C.E., Olmstead, R.G., 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. Int. J. Plant Sci. 161, S83–S96.

Hilu, K.W., Liang, H., 1997. The *matK* gene: sequence variation and application in plant systematics. Am. J. Bot. 84, 830–839.

Hilu, K.W., Borsch, T., Müller, K., Soltis, D.E., Soltis, P.S., Savolainen, V., Chase, M., Powell, M., Alice, L.A., Evans, R., Sauquet, H., Neinhuis, C., Slotta, T.A., Rohwer, J.G., Campbell, C.S., Chatrou, L., 2003. Angiosperm phylogeny based on *matK* sequence information. Am. J. Bot. 90, 1758–1776.

Hoot, S.B., Crane, P.R., 1995. Inter-familial relationships in the Ranunculidae based on molecular systematics. Plant Syst. Evol. Suppl. 9, 119–131.

Hoot, S.B., Magallón, S., Crane, P.R., 1999. Phylogeny of basal eudicots based on three molecular datasets: *atpB*, *rbcL*, and *18S* nuclear ribosomal DNA sequences. Ann. Mo. Bot. Gard. 86, 1–32.

Hufford, L., Crane, P.R., 1989. A preliminary phylogenetic analysis of lower Hamamelidae. In: Crane, P.R., Blackmore, S. (Eds.), Evolution, Systematics and Fossil History of the Hamamelidae. Clarendon Press, Oxford, pp. 175–192.

Hupfer, H., Swiatek, M., Hornung, S., Herrmann, R.G., Maier, R.M., Chiu, W.L., Sears, B., 2000. Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable Euoenothera plastomes. Mol. Gen. Genet. 263, 581–585.

Johnson, L.A., Soltis, D.E., 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *matK* sequences. Ann. Mo. Bot. Gard. 82, 149–175.

Kadereit, J.W., Blattner, F.R., Jork, K.B., Schwarzbach, A., 1995. The phylogeny of the Papaveraceae sensu lato: morphological, geographical and ecological implications. Plant Syst. Evol. Suppl. 9, 133–145.

Kelchner, S.A., 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. Ann. Mo. Bot. Gard. 87, 482–498.

Kelchner, S.A., Clark, L.G., 1997. Molecular evolution and phylogenetic utility of the chloroplast *rpl*16 intron in *Chusquea* and the Bambusoideae (Poaceae). Mol. Phylogenet. Evol. 8, 385–397.

Kim, K.J., Lee, H.L., 2004. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res. 11, 247–261.

Kim, S., Soltis, D.E., Soltis, P.S., Zanis, M.J., Suh, Y., 2004. Phylogenetic relationships among early-diverging eudicots based on four genes: were the eudicots ancestrally woody? Mol. Phylogenet. Evol. 31, 16–30.

Kita, Y., Kato, M., 2001. Infrafamilial phylogeny of the aquatic angiosperm Podostemaceae inferred from the nucleotide sequences of the *matK* gene. Plant Biol. 3, 156–163.

Kramer, E.M., Jaramillo, M.A., Di Stilio, V.S., 2004. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. Genetics 166, 1011–1023.

Leebens-Mack, J., Soltis, D.E., Soltis, P.S., 2005. Plant reproductive genomics at the Plant and Animal Genome Conference. Comp. Funct. Genom. 6, 159–169.

Levinson, G., Gutman, A., 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4, 203–221.

Litt, A., Irish, V.F., 2003. Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: implications for the evolution of floral development. Genetics 165, 821–833.

Loconte, H., Campbell, L.M., Stevenson, D.W., 1995. Ordinal and familial relationships of ranunculid genera. Plant Syst. Evol. Suppl. 9, 99–118.

Löhne, C., Borsch, T., 2005. Phylogenetic utility and molecular evolution of the *petD* group II intron in basal angiosperms. Mol. Biol. Evol. 22, 317–332.

Lutzoni, F., Wagner, P., Reeb, V., Zoller, S., 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. Syst. Biol. 49, 628–651.

Magallon, S., Crane, P.R., Herendeen, P.S., 1999. Phylogenetic pattern, diversity and diversification of eudicots. Ann. Mo. Bot. Gard. 86, 297–372.

Meimberg, H., Thalhammer, S., Brachmann, A., Heubl, G., 2006. Comparative analysis of a translocated copy of the *trnK* intron in carnivorous family Nepenthaceae. Mol. Phylogenet. Evol. 39, 478–490.

Morgenstern, B., 1999. DIALIGN 2: improvement of the segment-to-segment approach to mutliple sequence alignment. Bioinformatics 15, 211–218.

Moritz, C., Dowling, T.E., Brown, W.M., 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. 18, 269–292.

Müller, J., Müller, K., 2004. TreeGraph: automated drawing of complex tree figures using an extensible tree description format. Mol. Ecol. Notes 4, 786–788.

Müller, J., Müller, K., Quandt, D., Neinhuis, C., 2005. PhyDE 〈www.phyde.de〉.

Müller, K., 2004. PRAP – computation of Bremer support for large data sets. Mol. Phylogenet. Evol. 31, 780–782.

Müller, K., 2005a. The efficiency of different search strategies in estimating parsimony jackknife, bootstrap, and Bremer support. BMC Evol. Biol. 5, 58.

Müller, K., 2005b. SeqState – primer design and sequence statistics for phylogenetic DNA data sets. Appl. Bioinform. 4, 65–69.

Müller, K., Borsch, T., 2005. Phylogenetics of Amaranthaceae based on *matK*/*trnK* sequence data – evidence from parsimony, likelihood, and Bayesian analyses. Ann. Mo. Bot. Gard. 92, 96–102.

Müller, K., Borsch, T., Hilu, K.W., 2006. Phylogenetic utility of rapidly evolving DNA on high taxonomical levels: comparing three cpDNA datasets of basal angiosperms. Mol. Phylogenet. Evol. 41, 99–117.

Nandi, O.I., Chase, M., Endress, P.K., 1998. Combined cladistic analysis of angiosperms using *rbcL* and non-molecular datasets. Ann. Mo. Bot. Gard. 85, 137–212.

Nixon, K.C., 1999. The parsimony ratchet: a rapid means for analyzing large data sets. Cladistics 15, 407–414.

Nuin, P.A.S., 2005. MTgui – a simple interface to ModelTest. Program distributed by the author, University of Toronto.

Nyffeler, R., 2002. Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from *trnK*/*matK* and *trnL-trnF* sequences. Am. J. Bot. 89, 312–326.

Ohi, T., Wakabayashi, M., Wu, S., Murata, J., 2003. Phylogeography of *Stachyurus praecox* (Stachyuraceae) in the Japanese Archipelago based on chloroplast DNA haplotypes. J. Jpn. Bot. 78, 1–14.

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

Qiu, Y.-L., Lee, J., Bernasconi-Quadroni, F., Soltis, D.E., Soltis, P.S., Zanis, M., Zimmer, E.A., Chen, Z., Savolainen, V., Chase, M.W., 2000. Phylogeny of basal angiosperms: analyses of five genes from three genomes. Int. J. Plant Sci. 161, S3–S27.

Quandt, D., in press. Universal primers for amplification of the *trnK*/*matK-psbA* region in land plants. An. Jard. Bot. Madrid 63.

Quandt, D., Stech, M., 2005. Molecular evolution of the *trnL*(UAA) intron in bryophytes. Mol. Phylogenet. Evol. 36, 429–443.

Quandt, D., Müller, K., Stech, M., Hilu, K.W., Frey, W., Frahm, J.-P., Borsch, T., 2004. Molecular evolution of the chloroplast *trnL-F* region in land plants. In: Goffinet, B., Hollowell, V., Magill, R. (Eds.), Molecular Systematics of Bryophytes. Monogr. Syst. Bot., Mo. Bot. Garden Press, pp. 13–37.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Tabata, S., 1999. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res. 6, 283–290.

Savolainen, V., Chase, M.W., Hoot, S.B., Morton, C.M., Soltis, D.E., Bayer, C., Fay, M.F., De Bruijn, A.Y., Sullivan, S., Qiu, Y.-L., 2000a. Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences. Syst. Biol. 49, 306–362.

Savolainen, V., Fay, M.F., Albach, D.C., Backlund, M., Van der Bank, M., Cameron, K.M., Johnson, S.A., Lledo, L., Pintaud, J.C., Powell, M., Sheanan, M.C., Soltis, D.E., Soltis, P.S., Weston, P., Whitten, W.M., Wurdack, K.J., Chase, M.W., 2000b. Phylogeny of the eudicots: a nearly complete familial analysis of the *rbcL* gene sequences. Kew Bull. 55, 257–309.

Schmitz-Linneweber, C., Maier, R.M., Alcaraz, J.P., Cottet, A., Herrmann, R.G., Mache, R., 2001. The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. Plant Mol. Biol. 45, 307–315.

Schmitz-Linneweber, C., Regel, R., Du, T.G., Hupfer, H., Herrmann, R.G., Maier, R.M., 2002. The plastid chromosome of *Atropa belladonna* and its comparison with that of *Nicotiana tabacum*: the role of RNA editing in generating divergence in the process of plant speciation. Mol. Biol. Evol. 19, 1602–1612.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C.,

Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., Sugiura, M., 1986. The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. EMBO J. 5, 2043–2049.

Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49, 369–381.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, B., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 651–701.

Soltis, D.E., Soltis, P.S., Mort, M.E., Chase, M.W., Savolainen, V., Hoot, S.B., Morton, C.M., 1998. Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. Syst. Biol. 47, 32–42.

Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M.E., Albach, D.C., Zanis, M., Savolainen, V., Hahn, W.H., Hoot, S.B., Fay, M.F., Axtell, M., Swensen, S.M., Prince, L.M., Kress, J.W., Nixon, K.C., Farris, J.S., 2000. Angiosperm phylogeny inferred from *18S* rDNA, *rbcL*, and *atpB* sequences. Bot. J. Linn. Soc. 133, 381–461.

Soltis, D.E., Senters, A.E., Zanis, M.J., Kim, S., Thompson, J.D., Soltis, P.S., De Craene, L.P.R., Endress, P.K., Farris, J.S., 2003. Gunnerales are sister to other core eudicots: Implications for the evolution of pentamery. Am. J. Bot. 90, 461–470.

Stech, M., Quandt, D., 2006. Molecular evolution and phylogenetic utility of the chloroplast atpB-rbcL spacer in bryophytes. In: Sharma, A.K., Sharma, A. (Eds.), Plant Genome: Biodiversity and Evolution, vol. 2B. Enfield, Science Publishers, pp. 409–431.

Swofford, D.L., 2001. PAUP*. Phylogenetic Analysis Using Parsimony (*and other Methods). Sinauer Associates, Sunderland, MA.

Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of the chloroplast DNA. Plant Mol. Biol. 17, 1105–1109.

Takhtajan, A., 1997. Diversity and Classification of Flowering Plants. Columbia University Press, New York.

Wanke, S., Jaramillo, M.A., Borsch, T., Samain, M.-S., Quandt, D., Neinhuis, C., 2007. Evolution of Piperales – combined *matK* gene and *trnK* intron sequence data reveal lineage specific resolution contrast. Mol. Phylogenet. Evol. 42, 477–497.

Wanntorp, L., De Craene, L.P.R., 2005. The *Gunnera* flower: key to eudicot diversification or response to pollination mode? Int. J. Plant Sci. 166, 945–953.

Wheeler, W.C., Gladstein, D.S., De Laet, J., 1996–2003. POY. Version 3.0., New York.

Yang, Y.W., Tai, P.Y., Chen, Y., Li, W.H., 2002. A study of the phylogeny of *Brassica rapa*, *B. nigra*, *Raphanus sativus*, and their related genera using noncoding regions of chloroplast DNA. Mol. Phylogenet. Evol. 23, 268–275.