Plant Syst Evol (2009) 282:201–228
DOI 10.1007/s00606-008-0099-7

ORIGINAL ARTICLE

Phylogeny of the eudicot order Malpighiales: analysis of a recalcitrant clade with sequences of the *pet*D group II intron

Nadja Korotkova · Julio V. Schneider · Dietmar Quandt · Andreas Worberg · Georg Zizka · Thomas Borsch

Received: 25 April 2007 / Accepted: 14 August 2008 / Published online: 23 January 2009

Abstract Malpighiales are one of the most diverse orders of angiosperms. Molecular phylogenetic studies based on combined sequences of coding genes allowed to identify major lineages but hitherto were unable to resolve relationships among most families. Spacers and introns of the chloroplast genome have recently been shown to provide strong signal for inferring relationships among major angiosperm lineages and within difficult clades. In this study, we employed sequence data of the *pet*D group II intron and the *pet*B-*pet*D spacer for a set of 64 Malpighiales taxa, representing all major lineages. Celastrales and Oxalidales served as outgroups. Sequence alignment was straightforward due to frequent microstructural changes with easily recognizable motifs (e.g., simple sequence repeats), and well defined mutational hotspots. The secondary structure of the complete *pet*D intron was calculated for *Idesia polycarpa* as an example. Domains I and IV are the most length variable parts of the intron. They contain terminal A/T-rich stem-loop elements that are suggested to elongate independently in different lineages with a slippage mechanism earlier reported from the P8 stem-loop of the *trn*L intron. Parsimony and Bayesian analyses of the *pet*D dataset yielded trees largely congruent with results from earlier multigene studies but statistical support of nodes was generally higher. For the first time a deep node of the Malpighiales backbone, a clade comprising Achariaceae, Violaceae, Malesherbiaceae, Turneraceae, Passifloraceae, and a Lacistemataceae–Salicaceae lineage received significant statistical support (83% JK, 1.00 PP) from plastid DNA sequences.

Keywords Malpighiales · Angiosperms · Molecular evolution · Group II introns · Non-coding DNA

Introduction

The Malpighiales are one of the largest and most diverse orders of flowering plants, containing about 8% of all eudicots and 6% of all angiosperms (Davis et al. 2005). In an expanded circumscription the order currently comprises 38 families (APG 2003; Barkman et al. 2004) and nearly 16,000 species (information taken from the Angiosperm Phylogeny Website, Stevens 2001 onwards). The order contains some well known families, such as Euphorbiaceae (spurges), Passifloraceae (passion fruits), Linaceae (flaxes), Salicaceae (poplars and willows), and Violaceae (violets).

N. Korotkova · A. Worberg · T. Borsch
Nees Institute for Biodiversity of Plants, University of Bonn,
Bonn, Germany

J. V. Schneider
Department of Systematic Botany, Biology I,
University of Leipzig, Leipzig, Germany

D. Quandt
Institute of Botany, Plant Phylogenetics and Phylogenomics
Group, Dresden University of Technology, Dresden, Germany

G. Zizka
Department of Botany and Molecular Evolution, Research
Institute Senckenberg and Johann Wolfgang Goethe-University,
Frankfurt am Main, Germany

Present Address:
D. Quandt
Nees Institute for Biodiversity of Plants, University of Bonn,
Bonn, Germany

Present Address:
T. Borsch (✉)
Botanischer Garten und Botanisches Museum Berlin-Dahlem
und Institut für Biologie, Freie Universität Berlin,
Königin Luise-Str. 6-8, 14195 Berlin, Germany
e-mail: t.borsch@bgbm.org

Many of the families are distributed in the tropics where they constitute an important element of the understory of tropical rain forests (Davis et al. 2005).

The first molecular study across angiosperms based on sequences of the plastid gene *rbc*L (Chase et al. 1993) already depicted a lineage of Chrysobalanaceae, Erythroxylaceae, Violaceae, Ochnaceae, Euphorbiaceae, Humiriaceae, Passifloraceae and Malpighiaceae within a rosid clade. Close relationships of these families had not been considered in pre-cladistic classification systems, e.g., those of Cronquist (1981) or Takhtajan (1997). The addition of morphological characters to the *rbc*L matrix (Nandi et al. 1998) also recovered this new clade and suggested morphological features such as a fibrous exotegmen, dry stigmas, trilacunar nodes and toothed leaf margin as possible synapomorphies. Subsequent analyses combining *rbc*L and *atp*B (Savolainen et al. 2000a) or *rbc*L, *atp*B and 18S rDNA sequences (Soltis et al. 2000) yielded 92% bootstrap (BS) and 100% jackknife (JK) support for the Malpighiales, respectively. The highest support from a single gene was obtained by the phylogenetic analysis of angiosperms of Hilu et al. (2003) based on partial sequences of the rapidly evolving plastid gene *mat*K.

Some major clades within Malpighiales have been identified so far, e.g., a clade uniting Elatinaceae and Malpighiaceae (Davis and Chase 2004), the clade of Ochnaceae, Quiinaceae and Medusagynaceae (Fay et al. 1997) or the grouping of Clusiaceae, Hypericaceae, Bonnetiaceae and Podostemaceae (Davis et al. 2005). Some of these families were merged into broadly defined families by APG II (2003), as for example Ochnaceae s.l. (including Medusagynaceae and Quiinaceae). Other families such as Flacourtiaceae were split up and partly transferred to Salicaceae, a family that now contains about 1,000 species (Chase et al. 2002). Euphorbiaceae s.l. are now viewed as several independent lineages (Euphorbiaceae, Phyllanthaceae, Picrodendraceae and Putranjivaceae (APG 1998; Savolainen et al. 2000b; Wurdack et al. 2005).

But even with large sets of data and the use of three (Soltis et al. 2000) or four genes (Davis et al. 2005) from all three plant genomes, the phylogeny of Malpighiales could not be resolved. The most recent study on Malpighiales (Tokuoka and Tobe 2006) combined sequences of *rbc*L, *atp*B, 18S rDNA, and *mat*K and yielded the best phylogenetic hypotheses of Malpighiales so far. Nevertheless, Malpighiales still remain the phylogenetically least understood angiosperm order.

Davis et al. (2005) provide evidence that the diversity in Malpighiales is the result of a rapid radiation that began in tropical rain forests in the late Aptian (114 mya), and that most lineages began to diversify shortly thereafter, with the Hypericaceae–Podostemaceae clade appearing as the youngest during the Campanian (76 mya). A relatively fast diversification into major lineages may serve as an explanation for the difficulty of resolving deep nodes in Malpighiales. Finding sequence characters that have changed at a sufficiently high rate to accumulate mutations between fast lineage branching events, and at the same time have not changed so fast that phylogenetic signal was obscured, appears as a solution. Introns are a promising tool since they are mosaics of conserved and variable elements and provide a greater range of variable sites evolving under different constraints (Kelchner 2002). Group II introns with their overall conserved secondary and tertiary structure and well characterized domains are especially suited for studying phylogenetic information content with respect to structure, function and molecular evolution of genomic regions.

The effectiveness of rapidly evolving and non-coding chloroplast regions as markers for deep nodes in angiosperms has already been demonstrated. For basal angiosperms, Borsch et al. (2003) sequenced the *trn*T–F region from the chloroplast genome consisting of two spacers and a group I intron, and Löhne et al. (2005) generated a dataset of sequences of the *pet*D group II intron and the *pet*B–*pet*D spacer. The resulting trees in both studies were highly resolved and well supported and congruent with the multigene and multigenome studies comprising a manifold higher number of sequenced nucleotides (Qiu et al. 2000; Zanis et al. 2002). Combined analyses of the rapidly evolving chloroplast regions *mat*K, *trn*T-F, and *pet*D for early branching angiosperms (Borsch et al. 2005) and for early branching eudicots (Worberg et al. 2007) showed that confidence into phylogenetic hypotheses still can be improved by including more sequence data from introns and spacers. Müller et al. (2006) have shown that the amount of informative sites as well as phylogenetic signal per informative character is higher in *mat*K and *trn*T-F as compared to the slowly evolving *rbc*L using a character resampling and statistical analysis pipe.

This study is part of an ongoing project to evaluate mutational dynamics of rapidly evolving and non-coding chloroplast DNA and their phylogenetic utility in eudicots. Aims of this study were first to generate a dataset of sequences of the *pet*B–*pet*D region for a representative taxon set of Malpighiales, and second to examine their alignability and potential for inferring relationships in a difficult to resolve clade. The third major aim was to evaluate the effects of microstructural mutations on the evolution of the different intron domains.

Materials and methods

Taxon sampling

The data set comprises 64 taxa from Malpighiales and eight representatives from Celastrales and Oxalidales as outgroup. All families of the order recognized by APG II (2003) are included except Bonnetiaceae, Euphroniaceae, Goupiaceae, Lophopyxidaceae and Putranjivaceae for which no material was available. For large families such as Euphorbiaceae or Salicaceae we selected representatives of major clades as retrieved in published phylogenetic analyses of these families. Most of the plants sampled were obtained from the living collection at the Botanical Gardens Bonn. A list of all sampled taxa, their origin and voucher information is given in Table 1.

Isolation of genomic DNA

Genomic DNA was isolated from silica-dried leaves or herbarium specimens following the modified CTAB extraction method with triple extractions described by Borsch et al. (2003). Fresh leaves were generally dried in silica gel before extraction. Dry tissue was ground to a fine powder using a mechanical homogenizer (Retsch MM200) with 5 mm beads at 30 Hz for 2 min. DNA from *Malesherbia ardens*, *Dichapetalum mossambicense*, *Chrysobalanus icaco*, *Picrodendron baccatum*, *Touroulia guianensis*, *Quiina integrifolia*, *Bergia suffruticosa*, *Ctenolophon englerianus*, *Phyllocosmus lemaireanus*, and *Microdesmis puberula* was isolated using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany).

Amplification and sequencing

The amplified fragment consisted of the *pet*B–*pet*D intergenic spacer, the *pet*D-5′-exon and the *pet*D intron. For practical reasons the *pet*B–*pet*D spacer was co-amplified using the universal forward primer pipetB1411F and the reverse primer pipetD738R designed by Löhne and Borsch (2005). Additional internal sequencing primers (OpetD897R: 5′-RATCCCTTSTTTCACTCCGATAG-3′; LIpetD878R: 5′-TGTAGTCATTTCCTCTGCATCGAC-3′; LAMpetD951R: 5′-CATACAAAGRATTTACTTGTTAC-3′; and SALpetD599F: 5′-GCAGGCTCCGTAAAATCC AGTA-3′) were designed in this study for specific groups of taxa because of pherograms not being readable downstream of long mononucleotide stretches.

PCR conditions followed Löhne and Borsch (2005). Reactions were performed in a T3 thermocycler (Biometra, Göttingen, Germany). In some cases where DNA had been isolated from herbarium specimens the universal primers were used in combination with the internal primers OpetD897R and SALpetD599F to amplify the *pet*D region in two overlapping halves. Fragments were visualized using the Flu-o-blu system (Biozym, Hamburg, Germany) and excised from the gel. The DNA was then purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. PCR products were directly sequenced using the DCTS Quick Start Kit (Beckman Coulter). The reaction mix contained 3 µl DCTS Quick Start Kit (Beckman Coulter), 0.5 µl primer (20 pm/µl), 0.5–6.5 µl DNA template and ultrapure water to obtain a total volume of 10 µl. The cycle sequencing temperature profile consisted of 30 cycles of 96°C for 020 min, 50°C for 020 min, 60°C for 0400 min, on a T3 thermocycler (Biometra, Göttingen, Germany). Samples were run on an automated capillary sequencer (CEQ 8000 Genetic Analysis System, Beckman Coulter). Pherograms were edited using the software PhyDE v0.97 (www.phyde.de).

Sequence alignment

Chloroplast introns and spacers exhibit a high number of microstructural mutations apart from substitutions. For correct primary homology assessment, the respective mutational events need to be identified and gaps have to be placed accordingly (e.g., Kelchner (2000)). The main alignment principle was therefore to search for sequence motifs, not overall sequence similarity. Sequences were aligned manually, using the alignment editor PhyDE v. 097 (www.phyde.de). The rules for manual alignment of non-coding chloroplast regions proposed by Löhne and Borsch (2005) were also followed here. Single-base indels that were identified during alignment were checked in the original pherograms to make sure that they were not reading errors. Mutational hotspots with uncertain homology assessment (Borsch et al. 2003) were excluded from phylogenetic analysis. The alignment is available from the corresponding author on request.

Sequence statistics and coding of length mutational events

The length ranges of the spacer and the structural partitions of the intron as well as GC content, transition/transversion ratio, and the number of informative and variable positions were calculated using SeqState v. 1.25 (Müller 2005b). Length mutations were coded according to the Simple Indel Coding method (Simmons and Ochoterena 2000) using the Indel Coder option in SeqState v. 1.25 and analysed in combination with the sequence data matrix.

Phylogenetic analysis

Parsimony tree search

All aligned positions were given equal weight and gaps were treated as missing data. The search for the shortest tree was performed using the parsimony ratchet approach

Table 1 Taxa used in this study, origin of the plant material, voucher information, herbarium acronyms, and GenBank accession numbers

Family	Genus/species	Country garden/field origin	Voucher	EMBL/GenBank accession number
Achariaceae	*Hydnocarpus annamensis* (Gagnep.) Lescot and Sleumer	BG Bonn 24705 [Kambodscha]	N. Korotkova 68 (BONN)	FM178044
Achariaceae	*Lindackeria paludosa* (Benth.) Gilg	Bolivia	St G. Beck, A. Zonta, L. Medina, G. Pardo, M. Puri 20410 (B,LPB)	FM178043
Balanopaceae	*Balanops balansae* Baill.	New Caledonia, Province du Sud	G. McPherson et al. 19244 (MO)	FM178067
Brunelliaceae	*Brunellia mexicana* Standl.	Mexico	Romero-Romero 2990 (MEXU)	FM178018
Caryocaraceae	*Anthodiscus amazonicus* Gleason and. A.C. Sm.	Peru	H. van der Werff, R. Vasquez 13889 (MO)	FM178083
Caryocaraceae	*Caryocar brasiliense* Cambess.	Bolivia	W. Hanagarth, C. Rosales 110 (B, LPB)	FM178084
Celastraceae	*Euonymus* cf. *europaea* L.	BG Bonn 3810	No voucher (photo)	FM178013
Celastraceae	*Salacia lehmbachii* Loes.	BG Bonn 05039	T. Borsch 3549 (BONN)	FM178015
Celastraceae	*Brexia madagascariensis* (Lam.) Ker Gawl.	BG Bonn	N. Korotkova 51 (BONN)	FM178014
Cephalotaceae	*Cephalotus follicularis* Labill.	BG Bonn 20402	No voucher (Photo)	FM178019
Chrysobalanaceae	*Chrysobalanus icaco* L.	Cuba	S. Dressler 164 (FR)	FM178064
Chrysobalanaceae	*Licania kunthiana* Hook.f.	Bolivia	St G. Beck, R. de Michel 20904 (B, LPB)	FM178063
Clusiaceae	*Clusia* spec.	BG Bonn 14150	N. Korotkova 61 (BONN)	FM178058
Clusiaceae	*Garcinia tinctoria* (DC.) Dunn	BG Bonn 1921	N. Korotkova 19 (BONN)	FM178059
Clusiaceae	*Calophyllum inophyllum* L.	BG Bonn, ex BG Osnabrück	N. Korotkova 52 (BONN)	FM178053
Ctenolophonaceae	*Ctenolophon englerianus* Mildbr.	Gabon	G. McPherson 16911 (UPS)	FM178028
Dichapetalaceae	*Dichapetalum mossambicense* Engl.	Tansania	Kayombo & Ntemi 2986 (MO)	FM178049
Elaeocarpaceae	*Crinodendron hookerianum* Gay	BG Bonn 16434	A. Worberg 29 (BONN)	FM178020
Elatinaceae	*Elatine hexandra* DC.	Germany, Westerwälder Seenplatte	N. Korotkova, K. Lewejohann & W. Lobin 1 (BONN)	FM178065
Elatinaceae	*Bergia suffruticosa* (Delile) Fenzl	Burkina Faso	J. Krohmer 1082 (FR)	FM178068
Erythroxylaceae	*Erythroxylum coca* Lam.	BG Bonn 19149	N. Korotkova 16 (BONN)	FM178024
Euphorbiaceae	*Acalypha hispida* Burm.f.	BG Bonn 1050	N. Korotkova 62 (BONN)	FM178072
Euphorbiaceae	*Aleurites fordii* Hemsl.	BG Bonn 19200	N. Korotkova 53 (BONN)	FM178077
Euphorbiaceae	*Croton tiglium* L.	BG Bonn 22270	N. Korotkova 14 (BONN)	FM178074
Euphorbiaceae	*Euphorbia milii* Des Moul.	BG Bonn 15789	A. Worberg 2 (BONN)	FM178080
Euphorbiaceae	*Manihot esculenta* Crantz	BG Bonn 19208	N. Korotkova 15 (BONN)	FM178079
Euphorbiaceae	*Sapium* spec..	Argentina, Salta	T. Borsch, T. Ortuno, R. P. Lopez 3745 (B, LPB)	FM178073
Humiriaceae	*Sacoglottis gabonensis* (Baill.) Urb.	Gabon	J. Stone, G. Walters, T. Nzabi & T. Mboumbore 3283 (MO)	FM178045
Humiriaceae	*Vantanea compacta* ssp. *macrocarpa* Cuatrec.	Bolivia	St. G. Beck 29502 (B, LPB)	FM178046
Hypericaceae	*Hypericum hookerianum* Wight & Arn.	BG Bonn 3113	No voucher (photo)	FM178054
Irvingiaceae	*Irvingia gabonensis* (Aubry-Lecomte ex O'Rorke) Baill.	Gabon	G. Mc. Pherson 16704 (MO, BR)	FM178030

Table 1 continued

Family	Genus/species	Country garden/field origin	Voucher	EMBL/GenBank accession number
Ixonanthaceae	*Phyllocosmus lemaireanus* (De Wild. & T. Durand) T. Durand & H.Durand	Sambia	D.K. Harder et al. 3140 (UPS)	FM178029
Lacistemataceae	*Lacistema aggregatum* (P.J. Bergius) Rusby	Bolivia	T. Miranda et al. 225 (LPB, MO)	FM178031
Lacistemataceae	*Lacistema nena* J.F. Macbr.	Bolivia	P. Espinoza 3 (B, LPB)	FM178032
Linaceae	*Linum catharticum* L.	Germany, Bavaria	T. Borsch 3826 (B)	FM178027
Linaceae	*Linum narbonense* L.	BG Bonn 8344	N. Korotkova 13 (BONN)	FM178026
Linaceae	*Reinwardtia cicanoba* (Buch.-Ham. ex D.Don) Hara	BG Bonn 24189	N. Korotkova 11 (BONN)	FM178025
Malesherbiaceae	*Malesherbia ardens* J.F. Macbr.	Peru	J. Schneider et al. 2799 (FR)	FM178033
Malpighiaceae	*Bunchosia nitida* (Jacq.) DC.	BG Bonn 11383	N. Korotkova 63 (BONN)	FM178062
Malpighiaceae	*Heteropterys chrysophylla* (Lam.) Kunth.	BG Bonn 5035	N. Korotkova 54 (BONN)	FM178061
Malpighiaceae	*Malpighia glabra* L.	BG Bonn 6013	No voucher	FM178060
Medusagynaceae	*Medusagyne oppositifolia* Baker	BG Edinburg 20030393-A [Seychelles]	E0021445 (E)	FM178069
Ochnaceae	*Discladium* spec.	BG Bonn 8126	T. Borsch 3395 (BONN)	FM178070
Ochnaceae	*Ochna serrulata* (Hochst.) Walp.	BG Bonn 117	A. Worberg 28 (BONN)	FM178071
Oxalidaceae	*Oxalis hedysaroides* Kunth	BG Bonn 14200	N. Korotkova 55 (BONN)	FM178017
Pandaceae	*Microdesmis puberula* Hook.f. ex Planch.	Ghana	D. K. Harder et al. 3302 (UPS)	FM178081
Pandaceae	*Panda oleosa* Pierre	Ghana	M. Merello, H. H. Schmidt, J. Amponsah, M. Chintoh, K. Baah 1626 (MO)	FM178082
Parnassiaceae	*Parnassia palustris* L.	Germany, Bavaria	T. Borsch 3783 (B, BONN)	FM178016
Passifloraceae	*Passiflora quadrangularis* L.	BG Bonn 1020	N. Korotkova 56 (BONN)	FM178035
Phyllanthaceae	*Andrachne colchica* Fisch. & C.A. Mey. ex Boiss.	BG Bonn 3738	A. Worberg 8 (BONN)	FM178078
Phyllanthaceae	*Phyllantus fluitans* Benth. ex Müll.Arg.	BG Bonn 1066	N. Korotkova 57 (BONN)	FM178051
Phyllanthaceae	*Securinega suffruticosa* (Pall.) Rehder	BG Bonn 3739	A. Worberg 7 (BONN)	FM178052
Picrodendraceae	*Picrodendron baccatum* (L.) Krug & Urb. Ex Urb.	Cuba	Kuba-Exkursion 83 (FR)	FM178050
Podostemaceae	*Dicraeanthus africanus* Engl.	Cameroon	J.-P. Ghogue 1413 (YA, GC, Z/ZT)	FM178055
Podostemaceae	*Djinga felicis* C. Cusset	Cameroon	J.-P. Ghogue, G. Ameka, R. Rutishauser 021021-09 (YA, GC, Z/ZT)	FM178056
Podostemaceae	*Tristichia trifaria* (Bory ex Willd.) Spreng.	Cameroon	J.-P. Ghogue, G. Ameka, R. Rutishauser 021023-14 (YA, GC, Z/ZT)	FM178057
Quiinaceae	*Quiina integrifolia* Pulle	French Guiana	M. F. Prévost & D. Sabatier 4162 (CAY)	FM178076
Quiinaceae	*Touroulia guianensis* Aubl.	French Guiana	M.F. Prévost & D. Sabatier 4164 (CAY)	FM178075
Rhizophoraceae	*Bruguiera gymnorhiza* (L.) Savigny	BG Bonn 17887	N. Korotkova 4 (BONN)	FM178048
Rhizophoraceae	*Rhizophora mangle* L.	BG Bonn 24763-3 [USA, Florida]	N. Korotkova 58 (BONN)	FM178047
Salicaceae	*Azara salicifolia* Griseb.	Bolivia	G. Torrico & C. Peca 204 (LPB)	FM178041

Table 1 continued

Family	Genus/species	Country garden/field origin	Voucher	EMBL/GenBank accession number
Salicaceae	*Dovyalis caffra* (Hook. f. & Harv.) Warb.	BG Bochum	T. Borsch (B)	FM178039
Salicaceae	*Flacourtia jangomas* (Lour.) Raeusch.	BG Bonn 12841	N. Korotkova 59 (BONN)	FM178042
Salicaceae	*Idesia polycarpa* Maxim.	BG Bonn 15364	N. Korotkova 12 (BONN)	FM178040
Salicaceae	*Populus alba* L.	BG Bonn 1295	A. Worberg 6 (BONN)	FM178036
Salicaceae	*Salix purpurea* L.	BG Bonn 17982	A. Worberg 30 (BONN)	FM178037
Salicaceae	*Salix reticulata* L.	Germany, Bavaria	T. Borsch 3825 (B)	FM178038
Trigoniaceae	*Trigonia nivea* Cambess.	Bolivia	St. G. Beck 17374 (B, LPB)	FM178066
Turneraceae	*Turnera grandidentata* (Urban) Arbo	BG Bonn 13932	N. Korotkova 60 (BONN)	FM178034
Violaceae	*Hybanthus anomalus* (Kunth) Melch.	BG Bonn 12796	T. Borsch 3897 (BONN)	FM178022
Violaceae	*Hybanthus concolor* (T.F. Forst.) Spreng.	USA, Missouri	T. Borsch (B)	FM178021
Violaceae	*Viola hederacea* Labill.	BG Bonn 14126	N. Korotkova 5 (BONN)	FM178023

Table 2 Characteristics of the *pet*B-*pet*D spacer and *pet*D intron sequences in Malpighiales

	*pet*B-*pet*D spacer	*pet*D intron	Domain I	domain II	domain III	domain IV	domain V	domain VI
Length range including hotspots	182–245	713–970	369–553	57–109	42–60	98–284	37–37	37–53
Mean length including hotspots (SD)	206.4 (13.6)	811.7 (46.4)	429.4 (33.3)	68.7 (9.3)	60 (3.1)	185.7 (29.5)	37 (0)	40 (1.7)
Position in the alignment (exhot)	1–315	316–1,548	316–830	831–955	956–1.042	1,043–1,458	1,459–1,495	1,496–1,548
Length range excluding hotspots	137–174	573–673	290–340	40–84	42–60	83–185	37–37	37–53
Mean length exhot (SD)	151.4 (6.2)	606.6 (14.3)	307.2 (7)	49.6 (6.5)	60 (3.1)	121.7 (13.7)	37 (0)	40 (1.7)
Mean length of all hotspots (nt)	55.08	205.14	122.15	18.99	0	64	0	0
Number of characters (exhot)	315	1,233	515	125	87	416	37	53
% variable charcters (exhot)	33.3	38.1	46.6	32	41.4	28.1	43.2	39.6
% informative characters (exhot)	26.7	29.3	34.6	27.2	34.5	21.2	35.1	34
Number of coded indels (exhot)	66	244	81	29	21	110	0	3
G/C content	27.2	34.3	34.4	32.3	44.1	28.8	44.3	39.9
Ti/Tv ratio	1.4	1.9	2.3	2.2	1.9	1.6	1.5	1.8

using the software PRAP (Müller 2004). Ratchet settings for this study were 200 iterations with 25% of the positions randomly upweighted (weight = 2) during each replicate and 10 random addition cycles. The matrix was run using only substitution information and then combined with the indel matrix. The number of steps for each tree and the consistency, retention, and rescaled consistency indices (CI, RI, and RC) were calculated by PAUP* v. 4.0b10 (Swofford 1998). Jackknifing was used to evaluate branch support. Jackknife parameters were chosen according to the optimal evaluation strategies described by (Müller 2005a). A total number of 10,000 jackknife replicates was performed using the TBR branch swapping algorithm with 36.788% of characters deleted in each replicate. One tree was held during each replicate.

Bayesian Inference

Bayesian Inference (BI) was performed using MrBayes 3.1 (Huelsenbeck and Ronquist 2001). Nucleotide substitution models for the dataset were evaluated using Modeltest 3.7 (Posada and Crandall 1998) with spacer and intron sequences analysed separately. The hierarchical likelihood ratio test (hLRT) suggested the GTR + I + Γ model as the best for both regions and, therefore, Bayesian analysis was run with the implementation of this model. Two separate BI analyses were run: one only with sequence data and another using sequence data combined with the indel matrix. For the latter, the dataset was partitioned into DNA and binary characters, the GTR + I + Γ model was employed for the sequences and the restriction model for the indel matrix.

Four simultaneous runs of Metropolis-coupled Markov Chain Monte Carlo (MCMCMC) analyses each with four parallel chains were performed for 1 million generations, saving one tree every 100th generation, starting with a random tree. Other MCMC parameters were left with the program's default settings. Likelihood values appeared stationary after 25,000 generations. From the 10,000 trees saved, the first 250 were discarded. The remaining trees were summarized in a majority rule consensus tree. All trees were drawn with TreeGraph v. 1.10 (Müller and Müller 2004).

Inference of RNA secondary structure

The complete intron structure was calculated from the sequence of *Idesia polycarpa* (Salicaceae). *Idesia* has a mid-sized intron where no large indels were observed and the extension of sequences in hotspots was moderate, and thus seemed a suitable model for Malpighiales. Apart from *Idesia*, structures of subdomain D2 of domain I and entire domains II–VI were calculated for additional taxa with deviating sequences. Secondary structures were determined using RNAstructure 4.3 (Mathews et al. 1996–2006). The respective algorithm is described in Mathews et al. (2004). Currently available algorithms on RNA secondary structure are not able to predict the structure of an entire group II intron (see Mathews et al. (2006) for discussion). Therefore, domains and subdomains of the intron were first identified by comparison with the annotated alignment of *pet*D intron sequences from maize, tobacco, spinach and *Marchantia* provided by Michel et al. (1989). Since the borders of structural partitions appear to be conserved, they could easily be identified. Then, secondary structures were individually calculated for each domain. Domain I had to be folded separately by each subdomain due to its large size. The DNA sequences were folded as RNA (allowing U–G pairing). Constraints for the two exon binding sites and the single stranded branch point A were defined. In cases where alternative foldings varying only slightly in their free energy were possible the choice of structures for illustration was based on both, free energy and comparison with the already known group II intron structures (Michel and Dujon 1983; Michel et al. 1989). Structures of each domain were later assembled using the software RNAViz 2.0 (De Rijk et al. 2003) to draw the entire intron.

Results

Sequence characteristics of the *pet*B–*pet*D region

The length of the entire fragment consisting of the *pet*B–*pet*D intergenic spacer, the *pet*D 5' exon and the *pet*D intron ranged from 912 to 1,094 nt in the taxa studied. No substitutions occurred in the *pet*D 5′-exon. The final matrix (only spacer and intron) contained 1548 characters after the exclusion of hotspots and the *pet*D 5′-exon. Positions excluded as hotspots in individual sequences are given in the "Appendix 1" (Table 3). The characteristics of the *pet*B–*pet*D-region, such as sequence length, GC-content, T_i/T_v-ratios, and the numbers of variable and informative characters are given in Table 2. A comparison of average GC content of the six intron domains revealed remarkable differences between them (Table 2). Domain I has a GC content slightly higher than domain II but lower than in domain III, although domain I is nearly as large as the other five domains together. The highest GC content is observed in domains III, V, and VI, which all are small.

Length variation in the *pet*B–*pet*D spacer was comparatively low. The shortest spacer was found in *Phyllanthus fluitans* (182 nt) and the longest in *Tristichia trifaria* (245 nt). Apart from larger indels of 5–10 nt that accounted for most of the length variability in the spacer, single nucleotide indels were frequent. Five hotspots in the spacer were excluded from the phylogenetic analyses. The first (H1) was the part at the beginning of the spacer, where several indels occurred, for which a sequence motif and a probable origin could not be determined. To avoid artifacts in the indel matrix, this part was excluded from analyses. The second (H2) hotspot was a poly-G stretch of 2–7 G's. The third hotspot (H3) was basically a poly-A stretch of 7–20 nt (containing individual substitutions). The largest hotspot (H4) was 10–54 nt long and an AT-rich satellite-like region. The fifth hotspot (H5) was again a poly-A stretch of 9–15 nt.

The *pet*D intron was shortest in *Brunellia mexicana* (713 nt) and longest in *Malpighia glabra* (970 nt). This length variability is mainly due to frequent microstructural changes in two large hotspots in the intron (see below). After exclusion of all hotspots, the number of base characters from the intron ranged from 573 to 673 in the matrix.

Secondary structure of the *pet*D intron

The proposed secondary structure of the *pet*D intron in *Idesia polycarpa* is shown in Fig. 1. Domain I is connected to the central core by a helical element of 20–24 nt. Domain I comprises the largest part of the intron, varying in length from 369 nt in *Brunellia mexicana* to 553 nt in *Malpighia glabra.* Subdomains A, B, and C are small stem-loop structures connected to each other by few interhelical nucleotides. A large helical element (D1), interrupted by several small bulges is the connecting part to subdomains D2 and D3 and forms the stem of the entire subdomain. Subdomain D2 is a large stem-loop element located between subdomain D3 containing the exon binding site 1 (EBS 1) and EBS 2. This stem-loop element corresponds to hotspot H6 and accounts for a large amount of the length variation in the *pet*D intron (Fig. 2). An alignment of the respective sequence parts is only feasible among closely related taxa within some of the families like Salicaceae, Ochnaceae–Quiinaceae, or Rhizophoraceae. Domain II and domain III are small stem-loop structures (Figs. 3, 4) separated by 10–13 interhelical nucleotides depending on the individual taxon. Domain II was approximately 70-nt long in most taxa without major variation between outgroups and Malpighiales. A small poly-T was excluded from the analyses as hotspot H7. Domain III was conserved in its length (Table 2). Short indels of 4–8 nt were present but not frequent and the domain was unambiguously alignable without exclusion of hotspots. Three interhelical nucleotides (ADT) separate domain III from domain IV. Domain IV is the second largest domain and another highly variable element of the intron. The helix that comprises the

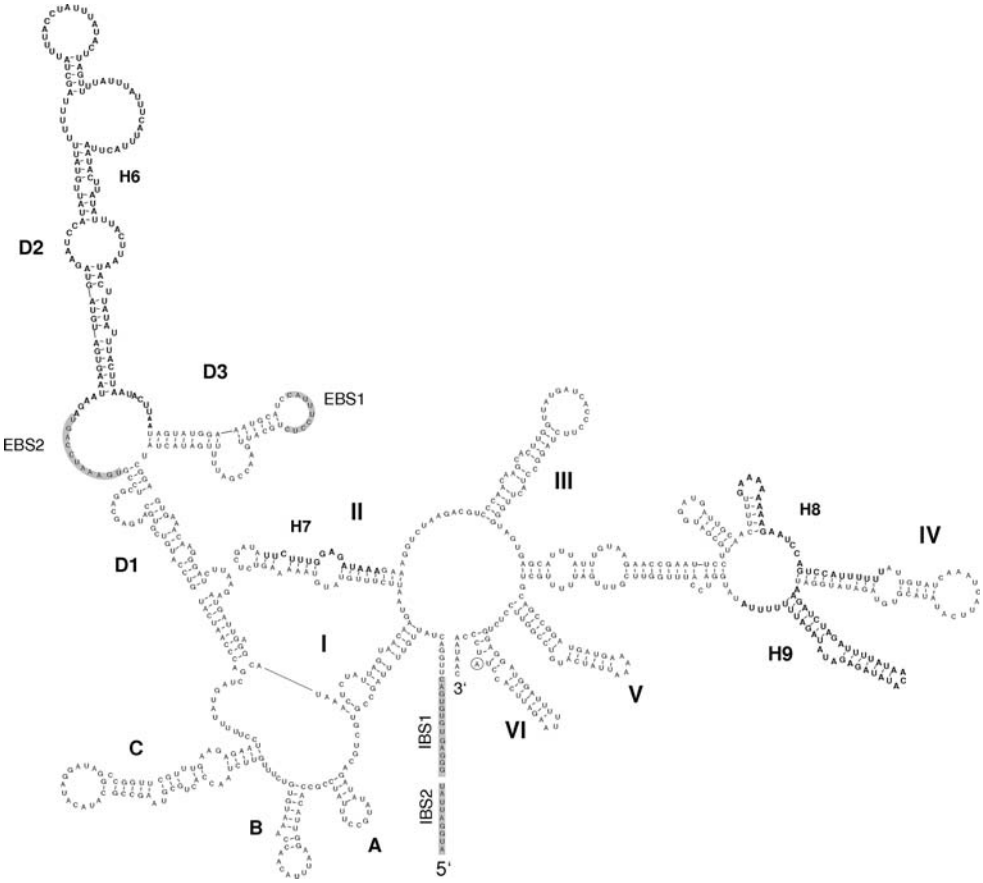


Fig. 1 Secondary structure of the *pet*D intron of *Idesia polycarpa* (Salicaceae). *Roman numbers I–IV* designate the six intron domains. Domain I is subdivided into subdomains *A–D*, with the latter being further subdivided into subdomains D1, D2 and D3. The encircled unpaired adenine in domain IV is the branch point A. Sequences falling in hotspots 6–9 are highlighted in *bold*. The exon binding sites (EBS 1 and EBS 2) and the intron binding sites (IBS 1 and IBS 2) are highlighted in *grey*

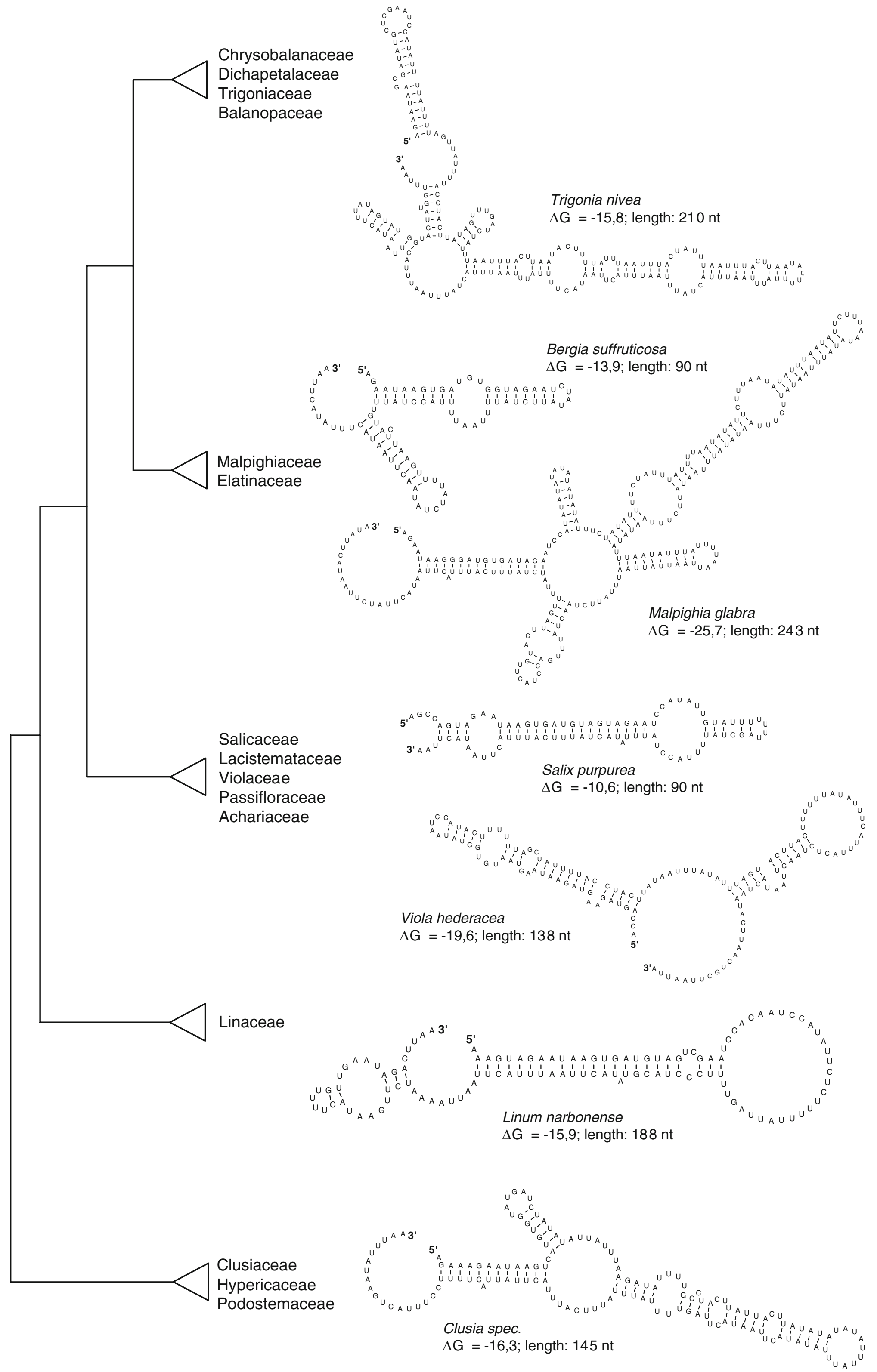


Fig. 2 Structures of the *pet*D group II intron subdomain D2 of domain I across Malpighiales plotted on a simplified phylogeny. Subdomain D2 corresponds to hotspot H6. Note the independent growth of AT-rich stem-loop elements in different lineages that is mainly the result of tandem repeats, e.g., the large size of D2 of *Malpighia glabra* is due to the 19-nt sequence motif "TTCTTTAATATATTTAATA" that is repeated four times

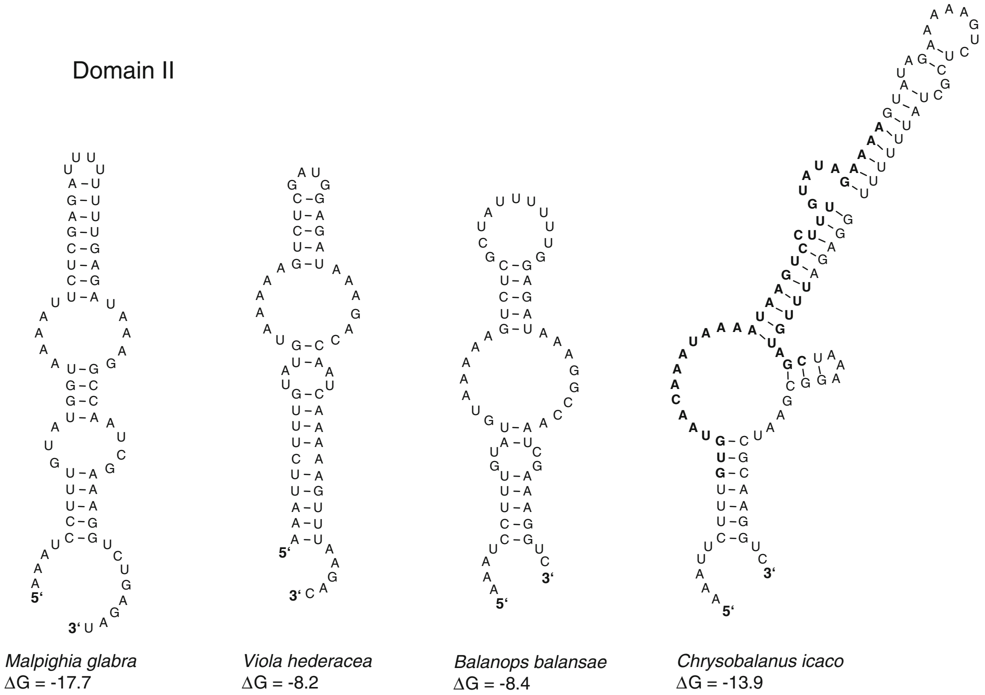


Fig. 3 Structural variability of domain II of the *pet*D intron in Malpighiales. *Chrysobalanus* possesses a derived structure due to two insertions (in *bold*)

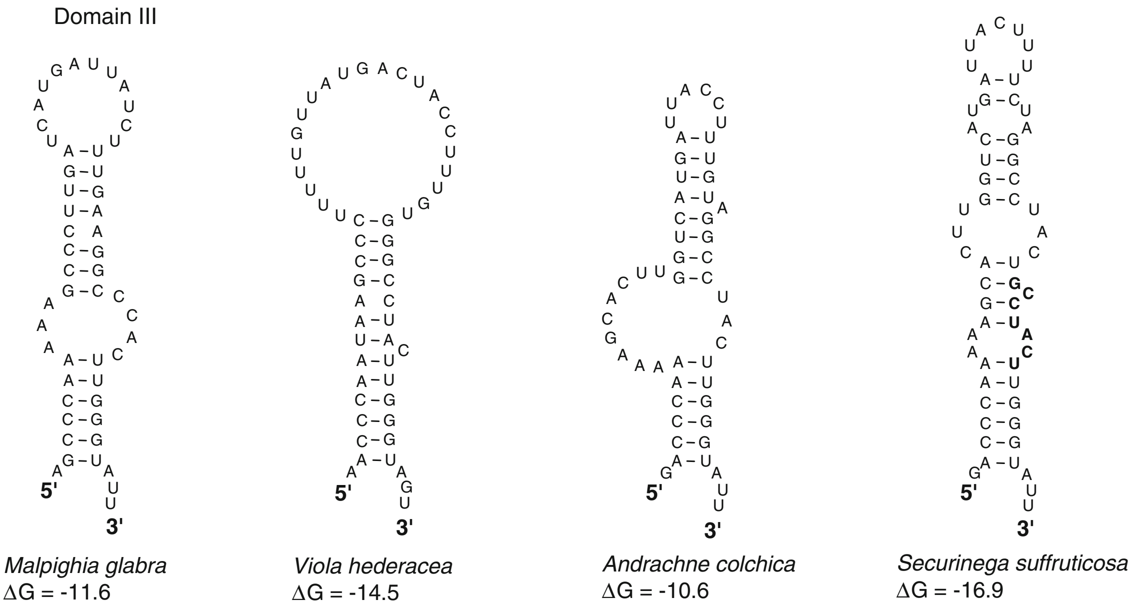


Fig. 4 Structural variability of domain III of the *pet*D intron in Malpighiales. *Securinega* possesses a derived structure relative to *Andrachne* due to an inserted simple sequence repeat (in *bold*)

stem of the domain is often only 4-nt long but substitutions can occur that lead to a larger interhelical part between domain III and IV. Domain IV (Fig. 5) was the most variable domain in terms of length, sequence and structural variability. Two hotspots (H8, H9) make up more than half of the domain and are composed of AT-rich elements and poly-A or poly-T stretches. Figure 6 depicts the secondary structure of the inferred inversion in *Djinga*. Unlike other inversions known (Kelchner and Wendel 1996) it is not associated with a hairpin. Domain IV and V are connected by usually only 1 nt. The structure of domain V (Fig. 7) reflects the conserved scheme known from other group II introns (Lehmann and Schmidt 2003; Michel and Dujon 1983; Pyle et al. 2007). Most parts of it are double-stranded with the exception of the bulge consisting of 2 nt and the small terminal loop of 4 nt. Domain V was the most

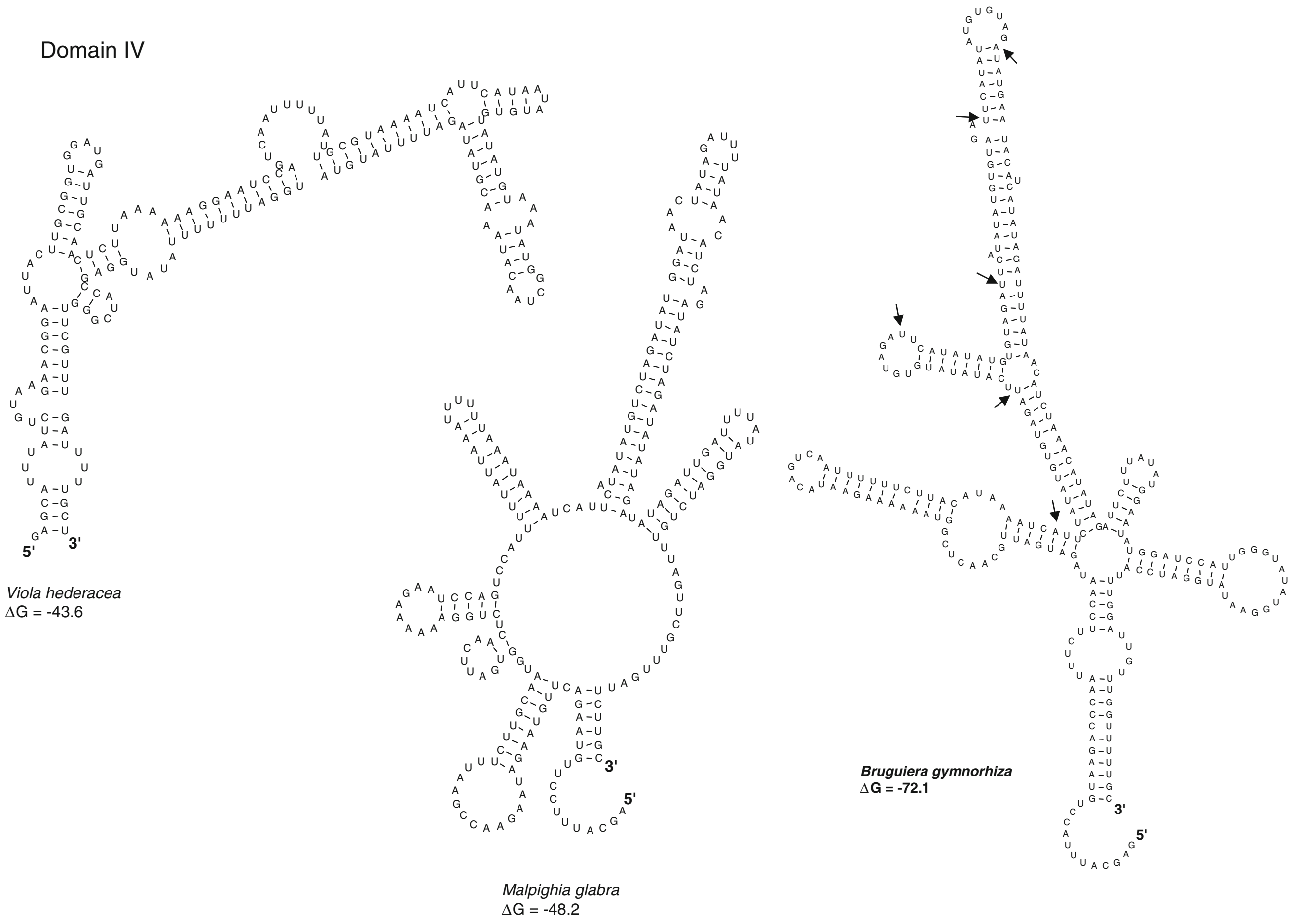


Fig. 5 Structural variability of domain IV of the *pet*D intron in Malpighiales. *Bruguiera gymnorhiza* possesses a strongly derived sequence due to a multiple simple sequence repeat of the 16-nt motif "TTCATATATGTGTAGA" (highlighted by *arrows*) that forms a stable stem-loop

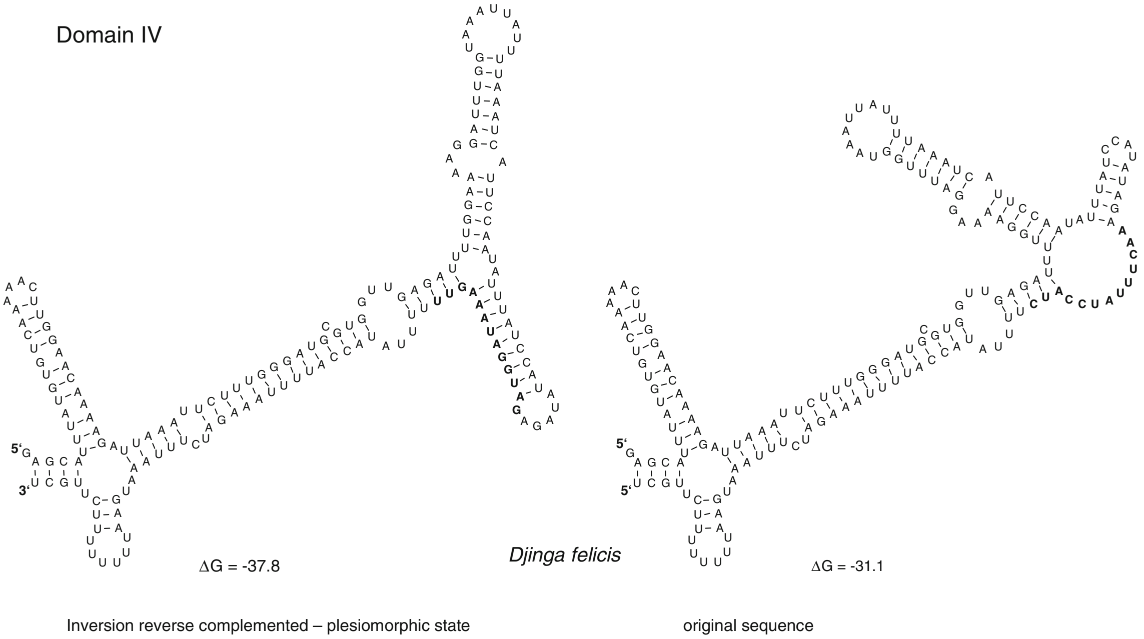


Fig. 6 Inversion in the *pet*D intron domain IV of *Djinga felicis* (Podostemaceae). The inverted motif is highlighted in *bold*

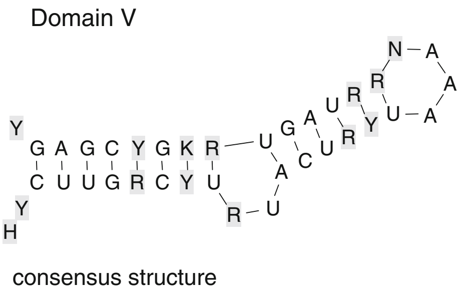


Fig. 7 Consensus structure of the highly conserved 37 nt long domain V of the *pet*D intron in Malpighiales. The 14 positions that were variable in the dataset are indicated by ambiguity *codes* and *highlighted*

conserved domain without any length mutations (Fig. 7). Four interhelical nucleotides, either Ts or Cs, separate the stems of domain V and VI. Domain VI was also strongly conserved around 40 nt and is largely helical with a small terminal loop of 3–8 nt (Fig. 8).

Length mutations

Length mutations were observed in the whole dataset but most of the length variability was found within the mutational hotspots. After excluding hotspots a total of 66 indels in the spacer and 244 in the intron were found (Table 4 in Appendix 2). Small indels were most frequent: 48 of 310 were indels of 1 nt and 130 were between 2 and 10-nt long. Only 23 indels were larger than 50 nt and still nine indels were larger than 100 nt, the largest indel in the dataset spanned 215 nt and was a deletion in domain IV shared by *Chrysobalanus icaco* and *Licania kunthiana* (both Chrysobalanaceae), resulting in the absence of nearly half of the domain. Nearly all the other large indels were also located in domain IV where also two inversions of 13 nt were detected in *Dicraeanthus* and *Djinga* (both Podostemaceae).

Phylogeny of Malpighiales

After the exclusion of hotspots the aligned matrix comprised 1,548 characters of which 973 were constant, 130 were variable but parsimony-uninformative, and 445 were parsimony-informative. Appending the 310 coded indels, the number of parsimony-informative characters was 554, whereas 331 were variable but parsimony-uninformative. The parsimony ratchet retained 624 shortest trees of 2,277 steps (CI: 0.44 RI: 0.59, RC: 0.26). Including the coded indels resulted in 483 shortest trees of 2,665 steps (CI: 0.49, RI: 0.60, RC: 0.29).

Results from the tree searches are shown in Figs. 9, 10, 11. Malpighiales were supported as monophyletic in all analyses (99% JK, 1.00 PP). The trees from Parsimony and Bayesian analyses differed only in the positioning of some terminals. Only one backbone node was recovered with confidence. Most of the terminal clades, however, received maximum support by jackknife values and posterior probabilities. The phylogram from Bayesian analysis

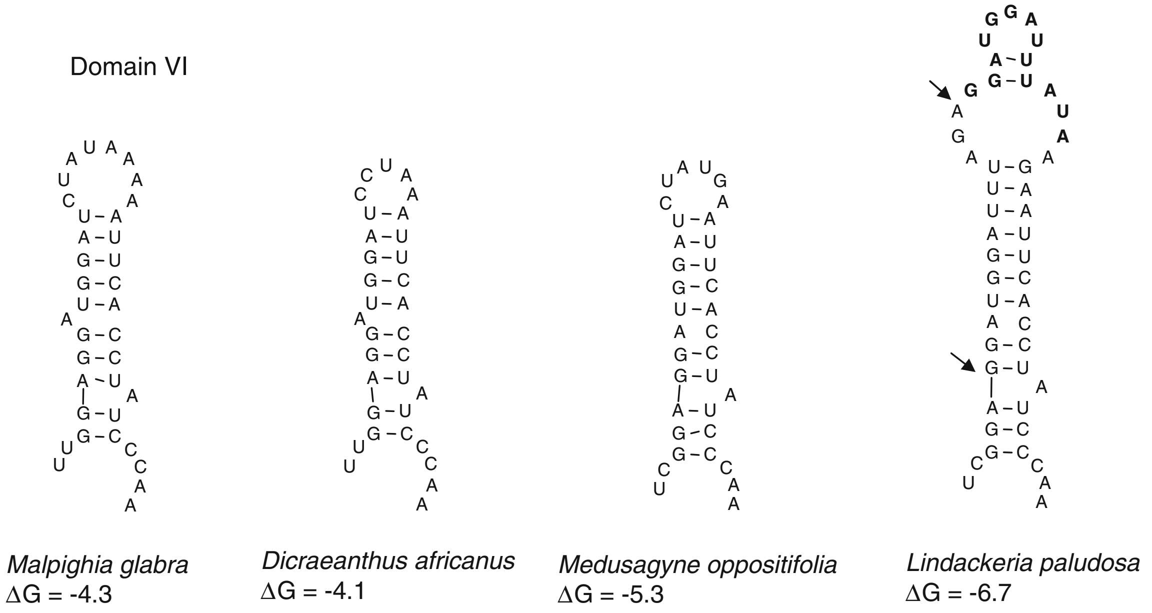


Fig. 8 Structural variability of domain VI of the *pet*D intron in Malpighiales. Three microstructural mutations are found, all of which affect the terminal loop. Structures of *Dicraeanthus* and *Medusagyne* deviate by obviously independent deletions of 2 or 3 nt, respectively. The phylogenetic context suggests that the two A's in *Medusagyne* were already deleted in the common ancestor of Quiinaceae, Medusagynaceae, and Ochnaceae. *Lindackeria* contains one 13 nt insertion that is a simple sequence repeat (indicated by *arrows*)

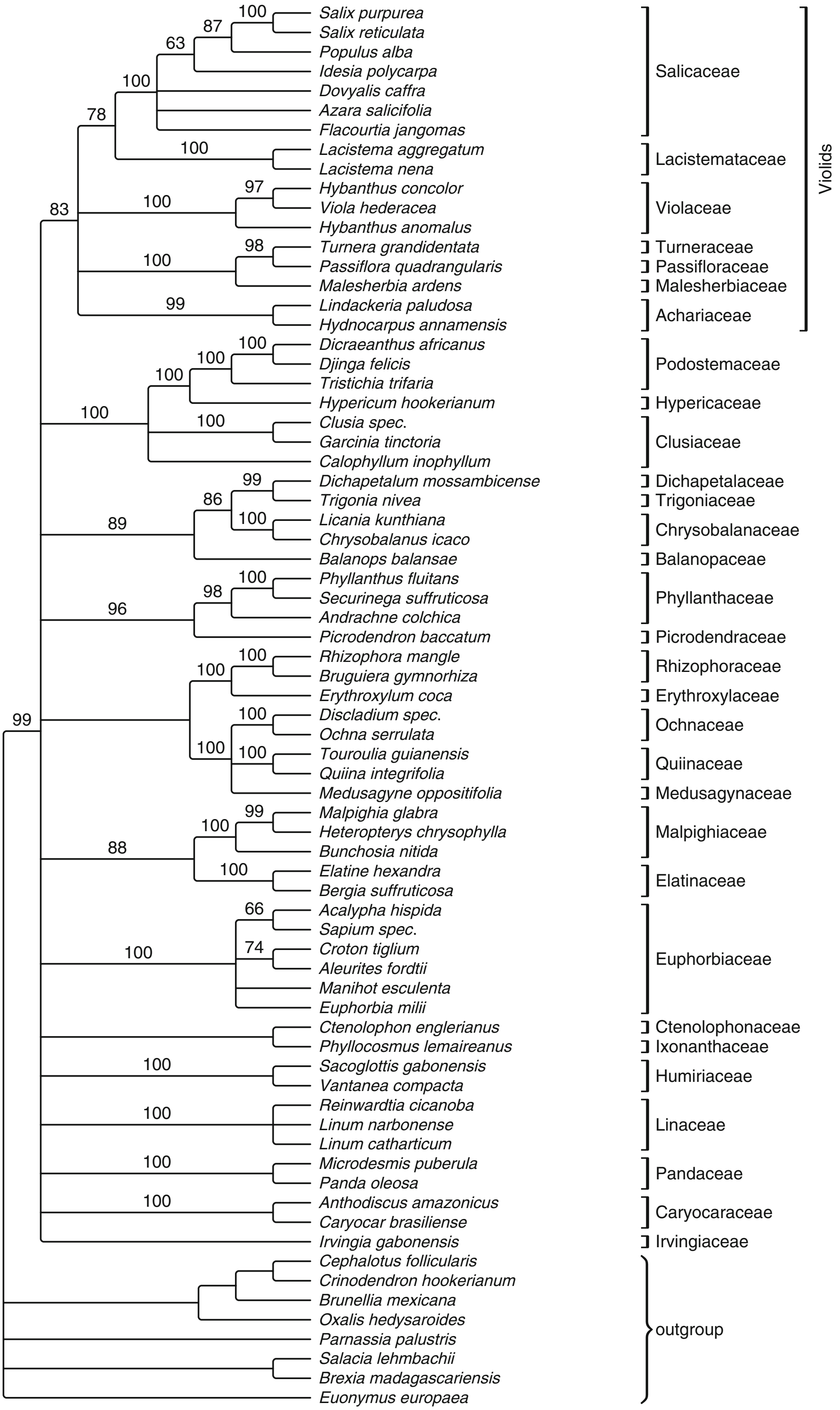


Fig. 9 Strict consensus tree of 483 shortest trees found by the parsimony ratchet based on the *pet*D dataset (excluding hotspots) combined with indels. Tree length: 2,665 steps (CI: 0.49, RI: 0.60, RC: 0.29). Numbers above branches are Jackknife support values (10,000 JK replicates)

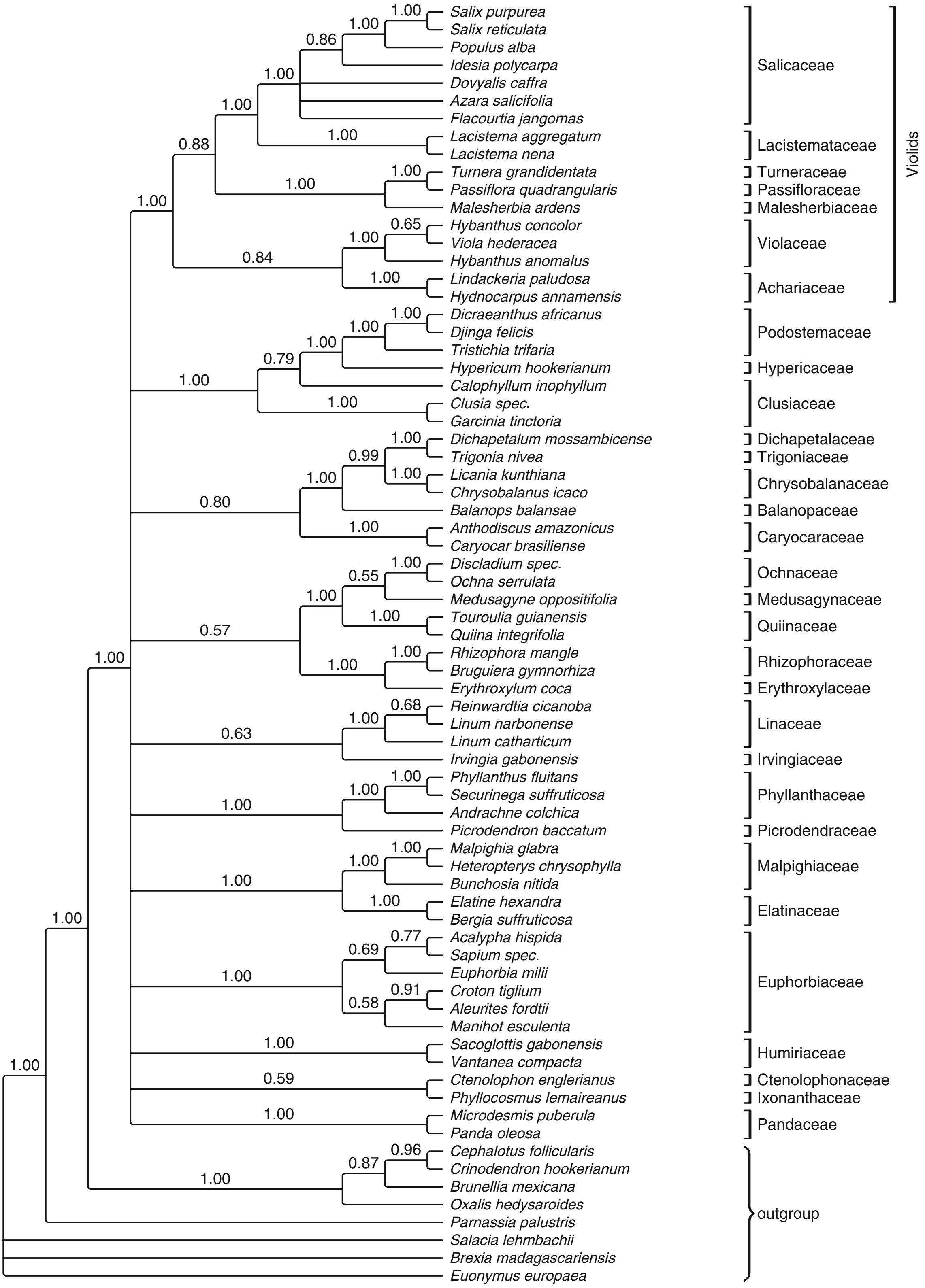
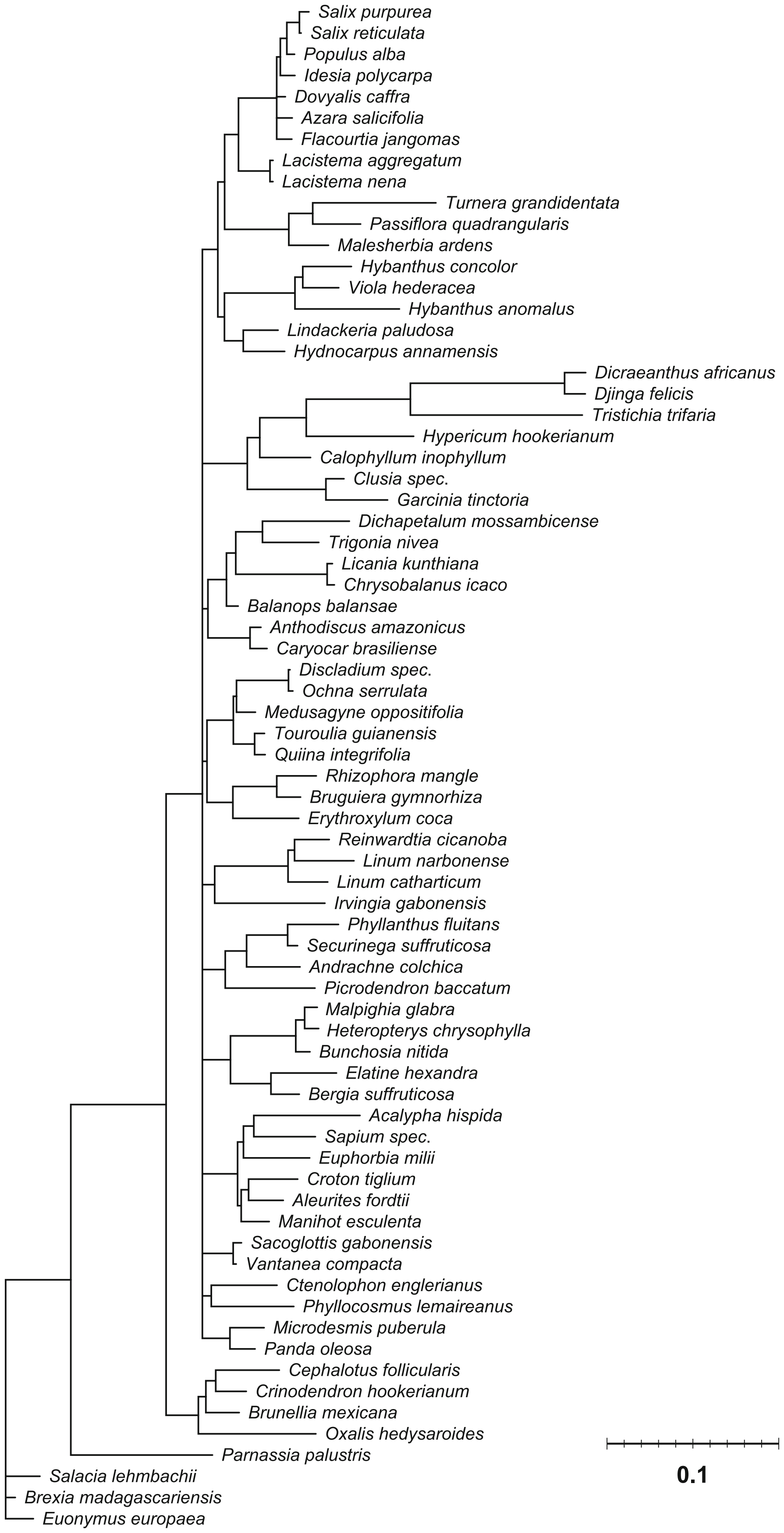


Fig. 10 The 50% majority-rule consensus tree obtained from Bayesian Inference based of the *pet*D dataset (excluding hotspots) combined with indels. Numbers above branches are Posterior Probabilities. Note the clade comprising Achariaceae, Violaceae, Malesherbiaceae, Turneraceae, Passifloraceae, and a Lacistemataceae–Salicaceae lineage (Violids) that is depicted with high posterior probability congruently to the parsimony tree

Fig. 11 Phylogram obtained from Bayesian Inference depicting long branches in the Hypericaceae-Podostemaceae-lineage



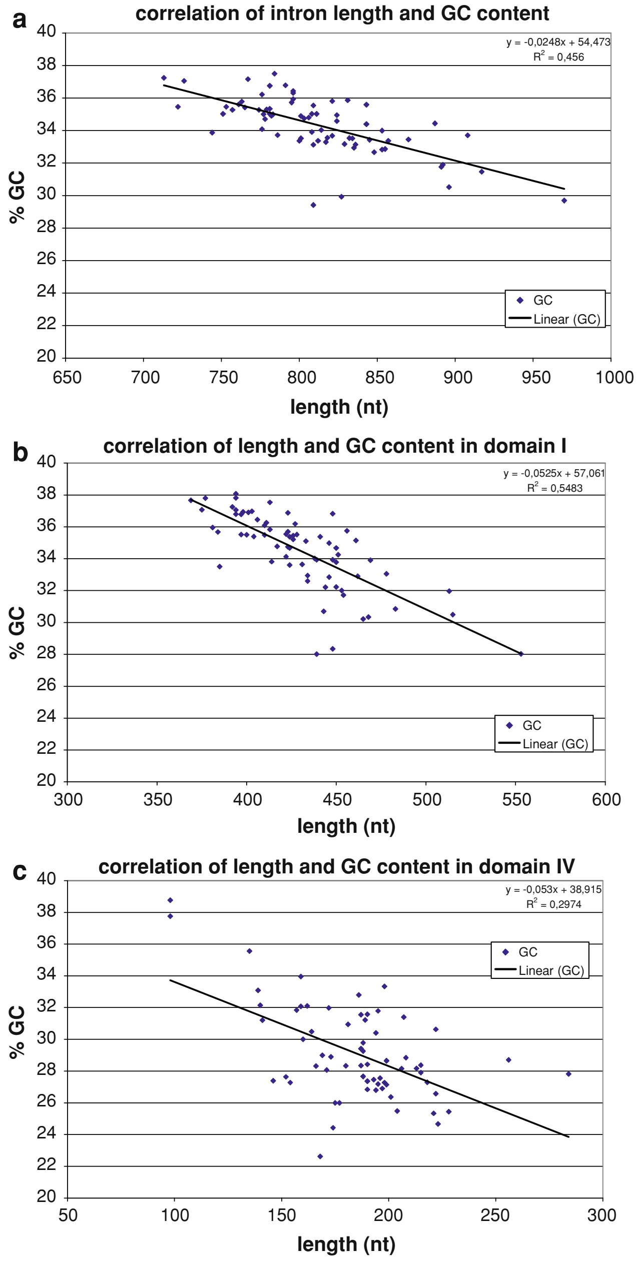


Fig. 12 Negative correlation of intron length and GC content for (**a**) the whole intron, (**b**) domain I, and (**c**) domain IV. The overall trend that increased size of the intron does not lead to a higher GC content is most prominent in the longest *pet*D intron sequence in the dataset (970 nt) that has one of the smallest GC contents (29.6%)

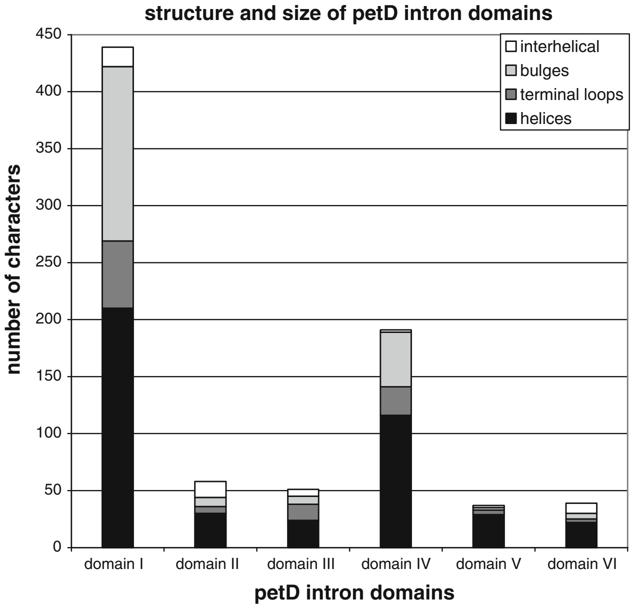


Fig. 13 Proportion of structural elements of the *pet*D intron of *Idesia*. Only those nucleotides that connect the intron domains are referred to as interhelical, single stranded parts or single unpaired nucleotides within domains are referred to as bulges

shows that most of the branches leading to the terminal clades of Malpighiales are short. However, branch lengths differ within terminal clades with the longest branches being observed in *Turnera grandidentata*, *Hypericum hookerianum*, *Hybanthus anomalus* and especially in the Podostemaceae.

A clade of Podostemaceae, Clusiaceae, and Hypericaceae is supported with 100% JK support and a PP of 1.00. *Hypericum* is sister to the Podostemaceae and the Clusiaceae. *Calophyllum* appears distant from other Clusiaceae genera *Clusia* and *Garcinia*. Euphorbiaceae are found as sister to the Hypericaceae/Podostemaceae/Clusiaceae clade in the parsimony tree, but there is no support for this grouping.

Linaceae are supported as monophyletic with maximal support, although the relationships within the clade are not resolved in the parsimony trees. *Irvingia* is depicted as sister to Linaceae but support for this grouping is low (0.62 PP). The sister family of Malpighiaceae are Elatinaceae with 83% JK support and a posterior probability of 1.00. Rhizophoraceae are found as sister to Erythroxylaceae with maximum support and both may be sister to the Ochnaceae s.l. clade but this grouping receives only 0.59 PP in the Bayesian tree. A clade comprising Chrysobalanaceae, Dichapetalaceae, Trigoniaceae, and Balanopaceae is supported with 83% JK and a PP of 1.00. Caryocaraceae are additionally found as sister to this clade in the Bayesian trees (0.76 PP). The two former Euphorbiaceae lineages Phyllanthaceae and Picrodendraceae were found to be sister to each other with 96% JK support and 1.00 PP. Pandaceae and Humiriaceae are supported as monophyletic, but their position within Malpighiales or their sister group is not resolved. Ochnaceae, Quiinaceae and Medusagynaceae form a clade that receives maximum support. The only backbone node that is supported as monophyletic with 81% JK and PP = 1.00 comprises Achariaceae, Violaceae, Passifloraceae, Turneraceae, Malesherbiaceae, Lacistemataceae and Salicaceae (including former Flacourtiaceae genera). Turneraceae, Malesherbiaceae, and Lacistemataceae appear in a clade. Moreover, Lacistemataceae are supported as sister to Salicaceae. The Bayesian tree further resolves Achariaceae as sister to Violaceae (0.84 PP) and the Achariaceae–Violaceae clade as sister to a Passifloraceae/Malesherbiaceae/Turneraceae plus Lacistemataceae plus Salicaceae clade.

Discussion

Molecular evolution of the *pet*D intron

The secondary structure calculated for the *pet*D intron of *Idesia* (Salicaceae) in this study fits very well into the known scheme of group II introns (Hausner et al. 2006; Michel et al. 1989; Qin and Pyle 1998; Toor et al. 2001). Alternative foldings are either energetically less favoured or violate structural constrains essential for correct splicing. Since subdomain D2 and domain IV are highly variable in terms of substitutions and sequence length, a common scheme for all *pet*D introns cannot be inferred. The calculated structures here reflect an optimization based on energy minimization that might only change slightly with advancing energy tables and algorithms. The first detailed study on the *pet*D intron evolution was conducted by Löhne and Borsch (2005). The author's analysis of frequency of structural partitions (stems, loops, bulges, interhelical single stranded sequence) in the different domains was an approximation based on the annotated consensus alignment by Michel et al. (1989) and visual examinations of the sequences with attention to complementary regions. To the contrary, this study shows the exact distribution of structural elements for the calculated intron structure of *Idesia*. In this study, all effectively paired nucleotides (Fig. 13) are considered helical. The need for understanding the effects of differential evolution of sequence partitions in phylogeny inference has clearly been pointed out by Kelchner (2002). Future work needs to recognize consensus helical elements by comparing secondary structures in order to group sequence characters that evolve under certain comparable constraints in a certain class.

Mutational hotspots are located in subdomain D2 of domain I, domain II and domain IV, which are the most variable parts of the intron. Already existing datasets for the *pet*D intron, i.e., those of Löhne and Borsch (2005) and the basal eudicots dataset of Worberg et al. (2007) allowed a comparison of hotspot locations. The hotspot in D2 is present in all datasets but is remarkably smaller in basal angiosperms or basal eudicots. Mutational dynamics as well as the AT content are increased in Malpighiales in D2. A hotspot in subdomain C of domain I was found in both studies, but not in the dataset analysed here. A hotspot in domain II is present in the alignment of Worberg et al. (2007) and in this dataset in about the same position. Alignments of different taxon sets basically show highly variable regions (hotspots H8/H9 in Malpighiales) in terminal parts of domain IV but these cannot be assigned to homologous sequence elements in different groups of angiosperms. Possible causes are in deviating mutational mechanisms that lead to insertion of AT-rich elements (see below).

Patterns of sequence conservation correspond to domain patterns of group II introns. Domain I is important for correct splicing and contains several tertiary interaction sites (Pyle and Lambowitz 2006). Besides domain I, Domain V is the only structural element that is essential for the catalytic function of the intron (Lehmann and Schmidt 2003; Pyle and Lambowitz 2006). It is the most conserved element with no length variability in this study. In domain I large parts apart from subdomain D2 are conserved. The percentage of variable characters (46%) is comparable to domain III (41%), but concerning the length of both domains, domain I is by far the more conserved one. Generally, domain IV is considered to be the most variable of all group II intron domains with respect to size and primary sequence (Lehmann and Schmidt 2003; Pyle and Lambowitz 2006). This can be confirmed for *pet*D in Malpighiales (Table 2). Sequence variation in the most conserved domains V and VI affects only their terminal parts. In domain V only one site located in the 4-nt long terminal loop seems freely substituted, exhibiting all four possible nucleotide states in Malpighiales (Fig. 6). In domain VI the branch point A that is essential for the transesterification during the splicing reaction along with many other positions is invariable. The only microstructural changes observed affect the terminal loop (Fig. 7).

The striking length variability of the subdomain D2 is the result of microstructural mutations happening independently in different lineages of Malpighiales (Fig. 2). Observation of sequence motifs revealed that length variability is caused mostly by multiple tandem repeats and poly-T-stretches. As suggested by Levinson and Gutman (1987), sequence motifs once repeated are prone to further duplication. Additional duplications might then involve the template motif and earlier duplicated elements at once, so that multiple repeats can be explained by few steps. Such a pattern is most prominent in the sequence of *Malpighia* (Fig. 2). To explain the evolution of terminal stem-loop elements in the P8 loop that is part of the *trn*L group I intron (Quandt et al. 2004) suggested slippage mediated growth of A/T rich sequence elements to have led to independent elongations of P8 in different land plant lineages. This process appears to have led to the stepwise insertion of up to 250 nt. It was further hypothesized that hairpin formation of complementary AT-rich sequence elements results in the stabilization of structure. We believe that similar mechanisms of sequence evolution also occur in subdomain D2 of domain I (Fig. 2) and possibly in domain IV. Figure 5 shows domain IV of *Bruguiera gymnorhiza* with a multiple tandem repeat of 19 nt. The repeat motif is pairing either with itself or is complementary to other sequence parts of the domain.

In *pet*D of Malpighiales a negative correlation of G/C content and sequence length is evident in domain I and in domain IV, affecting the whole intron (Fig. 12).

Microstructural changes are now widely accepted to provide useful phylogenetic information with a low degree of homoplasy, e.g., (Graham et al. 2000; Müller and Borsch 2005; Simmons and Ochoterena 2000). Nevertheless, the mutational mechanisms leading to microstructural changes are far from clear. We have analyzed the effects of a number of larger microstructural mutations (inserted or deleted motifs > 3 nt) on secondary structure. There seem to be two groups of such mutations. One group (Fig. 5) are those in AT-rich terminal stem-loops as discussed above. The other group (Figs. 3, 4) are length mutations that do not occur in terminal loops where their impact on the overall structure would be lowest. In the latter group the inserted repeats lead to the formation of helical secondary structural elements that are GC-rich and therefore stable. In addition, reverse complementary sequence elements to the inserted motif are present in other parts of a domain. Figure 4 illustrates a SSR in domain III that is synapomorphic for Phyllanthaceae (*Phyllanthus* and *Securinega*). Compared to the sister taxon *Andrachne* (Fig. 4; plesiomorphic state without SSR) the inserted motif "GCCTACT" has a complementary $5'$ part and leads to an elongated stable stem in *Securinega*. A similar situation is found in domain II (Fig. 3). The still insufficient resolution of the tree of Malpighiales limits the analysis of the evolutionary history of microstructural changes to unambiguous cases as the ones discussed. The mechanisms that lead to the insertion of long G/C rich, repeated sequence elements may differ from those acting in A/T rich stem-loops, the latter of which are usually compared with slipped strand mispairing (Quandt et al. 2004). Slipped strand mispairing (Levinson and Gutman 1987) seems to be an insufficient explanation for the insertion of rather long (sometimes 20 nt and more) G/C-rich elements because patterns of homoplasy differ between GC-rich domain elements and AT-rich stem-loops. (Borsch et al. 2007) found a strong insertion bias of SSRs in the evolution of the *trn*T-*trn*F region in Nymphaeales. However, slipped strand mispairing as it is also considered to occur in satellite sequences (Levinson and Gutman 1987) is expected to result in a stochastic distribution of deletions and insertions of short motifs. Considering our observation of long insertions that lead to stable helical elements in the intron's secondary structure appears to be in line with this because stable RNA foldings might be less likely affected by negative selection. Further structural comparisons of length variable sequences in a phylogenetic context are likely to provide insights into patterns and mechanisms of intron evolution.

Phylogenetic utility of the *pet*B-*pet*D region at ordinal level and the backbone of Malpighiales

The best so far existing phylogenetic hypotheses for Malpighiales are trees inferred from the multi-gene datasets of Davis et al. (2005), Soltis et al. (2000) and Tokuoka and Tobe (2006). The *pet*D trees also recovered all major lineages inferred by the multigene studies and even resolved additional nodes. The application of *pet*D sequence data in this study provides yet another example that non-coding and rapidly evolving genomic regions entail the same or even more phylogenetic structure than manifold bigger datasets of sequences of coding genes.

The fact that for the first time a backbone node (a clade comprising the seven families Passifloraceae, Malesherbiaceae, Turneraceae, Violaceae, Salicaceae, Lacistemataceae, and Achariaceae) receives significant Jackknife support with plastid DNA data can be taken as further evidence for the phylogenetic utility of *pet*D in Malpighiales. Well supported trees have been inferred based on *pet*D sequence data across angiosperms. Löhne and Borsch (2005) found trees for early diverging angiosperms, comparable to gene trees of *mat*K and *trn*T-*trn*F. Worberg et al. (2007) depicted a similar picture for resolving the basal grade of eudicots. One of the so far most comprehensive datasets for different chloroplast spacers, introns and *mat*K with identical taxon sampling is the Nymphaeales dataset of Löhne et al. (2007). A comparison of variability, homoplasy and phylogenetic structure of different group II introns in Nymphaeales revealed the highest values of phylogenetic structure *R* (Müller et al. 2006) for the *rpl*16 and the *trn*K intron, whereas the *pet*D intron had the lowest *R* value. The *pet*D intron seems to be one of the most conserved group II introns in the chloroplast single copy region. Thus, it will be promising to employ other group II introns, such as those residing in *rpl*16 or *trn*K for phylogeny reconstruction in Malpighiales.

The alignment of *pet*D sequences in Malpighiales was straightforward, as experienced in other datasets of angiosperms. Mutational hotspots are well defined (see also discussion above) although not much smaller as compared to those delimited in alignments across basal angiosperms (Löhne and Borsch 2005) or basal eudicots (Worberg et al. 2007). When only a single clade of angiosperms is sampled such as the Malpighiales, it could be expected that overall distances of sequences are smaller, and that accordingly, the hotspots are smaller. However, our data show that this is not necessarily true because of lineage specific effects. Mutational dynamics seems to be increased within hotspot regions in several Malpighiales families, including the above described lineage-specific insertions of A/T-rich sequence elements. In groups of closely related taxa where the respective regions in domains I and IV have a common evolutionary history, additional *pet*D characters can be used at lower taxonomic level.

Relationships within Malpighiales

This study is the first to use non-coding spacer and intron sequences for phylogeny inference of the Malpighiales. Most of the interfamilial relationships found in previous studies were also recovered in our analysis, and several clades received even higher support. An important outcome is that our analysis corroborated the close relationship of Salicaceae, Lacistemataceae, Turneraceae, Passifloraceae, Malesherbiaceae, Violaceae, and Achariaceae which received 83% JK and a PP of 1.0. This group is here called Violids (Figs. 10, 11) to facilitate further discussion. The clade has been previously hypothesized by a combined analysis of *ndh*F and *rbc*L data (Davis and Chase 2004) and in the four-gene study of Tokuoka and Tobe (2006) but only with 57% BS and 59% BS, respectively.

Passiflora, *Turnera* and *Malesherbia* form a clade that corresponds to Passifloraceae *sensu lato* of APG II (2003), where an inclusion of Turneraceae and Malesherbiaceae into Passifloraceae was suggested. Passifloraceae and Turneraceae are tropical herbs, shrubs vines, or rarely trees, Malesherbiaceae are a small family of xerophytes native to the Andes and to the arid parts of coastal Chile and Peru. These families formed a clade with 100% support in (Chase et al. 2002; Davis and Chase 2004), as well as in the three-gene study of (Soltis et al. 2000). Chase et al. (2002) found Turneraceae and Malesherbiaceae being sister to Passifloraceae, whereas our *pet*D data provide evidence that Turneraceae and Passifloraceae are sister groups (98% JK, 1.0 PP). The relationship of these three families in respect of floral morphology was discussed recently by Krosnick et al. (2006).

Our analysis recovered Lacistemataceae as sister to Salicaceae with 78% JK and a PP of 1.0. This confirms the findings from two to four-gene studies (Davis et al. 2005; Tokuoka and Tobe 2006) and an analysis using *mat*R sequences (Davis and Wurdack 2004). Salicaceae is here used in its recent and broad definition (APG II 2003) including Flacourtiaceae p.p. The woody pantropical family Flacourtiaceae has been shown to be polyphyletic in all previous molecular analyses. The morphology of Flacourtiaceae is very heterogeneous and the circumscription of the family has always been controversial. Based on a detailed molecular analysis using *rbc*L, Chase et al. (2002) proposed a splitting of the family: one part was transferred to Salicaceae; the other part was placed in the newly accepted Achariaceae (APG II 2003). Not surprisingly, representatives of the former Flacourtiaceae were retrieved in our analysis in Salicaceae s.l. and Achariaceae, respectively. Since both families are not sister to each other, the separation of Achariaceae as proposed by Chase et al. (2002) is supported by our *pet*D data.

It is noteworthy that the families of the Violid clade were all assigned to the order Violales sensu Cronquist (1981) except Salicaceae s.str. A feature that could be considered a synapomorphy for this clade is parietal placentation. In Cronquist's system, Flacourtiaceae were supposed to stand “basal” within Violales with supposed affinities to Lacistemataceae. Turneraceae, Passifloraceae, and Malesherbiaceae were considered to be related to each other, but as distinct families that probably have originated in or near Flacourtiaceae. Achariaceae (circumscribed including only the genera *Acharia*, *Ceratiosicos* and *Guthriea*) were also considered as related to Passifloraceae (Cronquist 1981). Salicaceae, consisting only of the genera *Salix* and *Populus* were treated as the separate monofamilial order Salicales. However, Cronquist also mentioned that Salicales share many morphological features (such as the numerous stamens, parietal placentation, separate styles and the occurrence of salicin in *Salix*, *Populus* and *Idesia*) with Flacourtiaceae and could be possibly placed near them. Thus, there is as well support from non-molecular characters for the clade of members of the former Violales (plus Salicaceae and Lacistemataceae) depicted in the *pet*D trees.

Clusiaceae and Hypericaceae were always considered as related to each other but were treated differently regarding their taxonomic rank. Some authors, e.g., Takhtajan (1997) and the most recent classification system of APG II (2003) maintained Clusiaceae and Hypericaceae as own families. Other authors considered them as subfamilies within Clusiaceae (e.g., Cronquist 1981). Applying a broad circumscription of the family, Clusiaceae was paraphyletic in a study using *rbc*L sequences (Gustafsson et al. 2002). The phylogeny presented therein recovered the subfamilies Clusioideae and Kielmeyeroideae as well supported clades, but subfamily Hypericoideae formed a clade with Podostemaceae. A sister group relationship between Hypericaceae/Hypericoideae and Podostemaceae was also recovered by our *pet*D data (100% JK, PP 1.0) as well as in the four-gene studies of Davis et al. (2005) and Tokuoka and Tobe (2006). Since *Calophyllum* does not appear in the same clade than *Clusia* and *Garcinia*, *pet*D data suggest that Clusiaceae might also be paraphyletic to the Hypericaceae–Podostemaceae-clade (Figs. 11, 12, 13) but this requires further testing with additional sequence data and increased taxon sampling. Davis et al. (2005) found that not Clusiaceae but Bonnetiaceae—a family not included in our study—are sister to Hypericaceae/Podostemaceae (with 80% BS). Due to the odd morphology of Podostemaceae it has long been problematic to place them within angiosperms (Soltis et al. 1999) and they seem to have little in common with Hypericaceae. However, a closer look reveals that Hypericaceae and Podostemaceae share also a number of non-molecular characters (Gustafsson et al. 2002). For Podostemaceae our *pet*D data corroborate the close relationship of *Dicraeanthus* and *Djinga* (Podostemoideae), whereas *Tristichia* (subfam. Tristichoideae) is distantly related (Kita and Kato 2001; Moline et al. 2007).

The monophyly of Malpighiaceae is well supported by *rbc*L and *mat*K (Cameron et al. 2001) as well as *ndh*F and *trn*L-F data (Davis et al. 2001). The floral morphology of Malpighiaceae is unique and distinguishes them from other rosids. Assumptions about the sister group of Malpighiaceae were difficult because of their morphological uniqueness (Cronquist 1981). A first hypothesis based on molecular data came from Davis and Chase (2004), who sampled a broad range of taxa from Malpighiales to establish the sister family of Malpighiaceae that turned out to be the small cosmopolitan family Elatinaceae. Elatinaceae and especially the genus *Elatine* are mostly aquatic herbs or semi-aquatic shrubs and were formerly placed near Clusiaceae and Hypericaceae (Cronquist 1981; Takhtajan 1997) because of morphological similarities, such as opposite leaves, seed and stem anatomy. However, since the morphological features of Elatinaceae were difficult to interpret, they were also treated as an own order Elatinales by Takhtajan (1997). Our study provides again evidence (88% JK, PP 1.00) that Elatinaceae are sister to Malpighiaceae. There are indeed some morphological and cytological features that link Malpighiaceae and Elatinaceae, as discussed in detail by Davis and Chase (2004). Most notable is the shared chromosome base number of $X = 6$ (shared only with byrsonimoids), opposite or whorled leaves with stipules, the presence of unicellular hairs and multicellular leaf glands.

Erythroxylaceae and Rhizophoraceae are families of tropical shrubs or trees with simple leaves and cymose inflorescences. Common features are tropane alkaloids and the presence of sieve-element plastids containing protein crystals (Nandi et al. 1998; Setoguchi et al. 1999). Both families may be treated together as Rhizophoraceae s.l. (APG II 2003). This study recovers both families as sisters in line with results of (Savolainen et al. 2000b; Schwarzbach and Ricklefs 2000; Setoguchi et al. 1999) and the three-gene study of Soltis et al. (2000), each with >90% bootstrap support, respectively.

There is evidence for a close relationship between the monogeneric family Medusagynaceae, an endemic family of the Seychelles, and the tropical families Quiinaceae and Ochnaceae. APG II (2003) suggested the inclusion of Quiinaceae and Medusagynaceae into a more widely circumscribed Ochnaceae sensu lato. Ochnaceae s.l. are recovered as a strongly supported (100% JK, PP 1.00) monophyletic group by the *pet*D data as already suggested by all studies that sampled taxa from these families (Chase et al. 2002; Fay et al. 1997; Savolainen et al. 2000b; Soltis et al. 2000). Quiinaceae are probably sister to Medusagynaceae and Ochnaceae, although only Soltis et al. (2000) provided some statistical support (60% JK) for this hypothesis. The most recent study with a broad taxon sampling on these families of Schneider et al. (2006) recovers Ochnaceae, Quiinaceae and Medusagynaceae as monophyletic groups and the authors suggest maintaining them as separate families. The three families were considered to be closely related by Cronquist (1981), who assigned them to the order Theales but without making assumptions about a direct relationship between them. Some morphological features that are common to all three families can be found, such as multilacunar nodes, mucilage cells/cavities, dentate leaves, and bitegmic ovules (Fay et al. 1997).

Euphorbiaceae are a large and highly diverse family of mainly tropical herbs, trees and shrubs. The genus *Euphorbia* is also very diverse in the Mediterranean Basin, South Africa and East Africa, where it is often succulent and cactus-like. First molecular evidence for the polyphyly of Euphorbiaceae was found by Chase et al. (1993), where *Euphorbia* appeared as sister to *Passiflora* and *Drypetes* as sister to *Ochna*. Subsequent studies confirmed the assumption that Euphorbiaceae were polyphyletic in their previous circumscription, since they appeared scattered among Malpighiales (Chase et al. 2002; Savolainen et al. 2000b; Soltis et al. 2000). Consequently, two former sublineages of Euphorbiaceae have been segregated as the new families Pandaceae (the former tribe Galearieae) and Putranjivaceae (the former tribe Drypeteae) in the system of APG I (1998). Pandaceae were treated as a separate family related to Euphorbiaceae already in the system of Cronquist (1981). Savolainen et al. (2000b) proposed the additional separation of the subfamilies Phyllanthoideae and Oldfieldioideae that were classified as Phyllanthaceae and Picrodendraceae in APG II (2003). Kathriarachchi et al. (2005) further clarified relationships within Phyllanthaceae and the circumscription of the family. The remaining Euphorbiaceae *sensu stricto* have been verified to be monophyletic (Wurdack et al. 2005). Most recently, Davis et al. (2007) depicted the parasitic Rafflesiaceae as one of the three major clades within Euphorbiaceae s.str.

A close relationship of Phyllanthaceae and Picrodendraceae was already suggested by Davis and Chase (2004) but only with 53% BS support. *Pet*D data resolve the

Phyllanthaceae-Picodendraceae clade with high confidence (96% JK; PP 1.00). Further support comes from morphology with shared features like unisexual, apetalous trimerous flowers, crassinucellar ovules with a nucellar beak, a large obturatur, and explosive fruits with carunculate seeds, which unites both families also with Euphorbiaceae (Merino Sutter et al. 2006).

Our study retrieved a well-supported clade of the small tropical families Balanopaceae, Chrysobalanaceae, Dichapetalaceae, and Trigoniaceae (89% JK, PP 1.00) with Balanopaceae being sister to the rest (89% JK, PP 1.0). This finding is congruent with what was found by Soltis et al. (2000) and Savolainen et al. (2000b). Balanopaceae appeared as sister to the other four families in both studies and APG II (2003) suggests an inclusion of Trigoniaceae, Dichapetalaceae, and Euphroniaceae into an expanded Chrysobalanaceae.

Conclusion

Single non-coding and rapidly evolving plastid genomic regions entail phylogenetic structure that is comparable to the information content of much larger datasets of sequences of coding genes with a manifold higher number of nucleotides sequenced per taxon. As such chloroplast introns and spacers are promising markers to resolve the tree of Malpighiales and other recalcitrant clades. Selecting highly informative genomic regions to be combined in phylogenetic analyses may be more effective than total evidence approaches that combine any kind of sequence data available.

Because of frequent microstructural mutations occurring during the evolution of intron sequences, analytical approaches need to be more complex as compared to sets of length conserved sequences. Secondary structure analyses are helpful to understand patterns and mechanisms underlying microstructural mutations. Intron sequences evolve differently in different domains and levels of sequence conservation vary considerably with respect to different structural partitions. Considering these patterns of intron evolution is essential for homology assessment. Most importantly, hypervariable AT-rich terminal stem-loop elements within domains I and IV may evolve independently in different lineages, and thus have to be excluded from phylogeny inference in matrices comprising distant taxa. Nevertheless, when an alignment principle that is based on recognizing sequence motifs is applied, the recognition of such mutational hotspots is straightforward.

Acknowledgments This study is part of the project “Mutational dynamics of non-coding genomic regions and their potential for reconstructing evolutionary relationships in eudicots” supported by the Deutsche Forschungsgemeinschaft (grants BO1815/2 to T.B. and QU153/2 to D.Q.). The funding of this project is greatly acknowledged. T.B. extends his thanks to the Deutsche Forschungsgemeinschaft for a Heisenberg Scholarship. We appreciate the support of Wilhelm Barthlott (Bonn) and Christoph Neinhuis (Dresden). The most important sources of material were the living plant collections at the Botanical Gardens of Bonn University. Their curator, Wolfram Lobin, and staff, especially Michael Neumann and Bernd Reinken helped wherever needed. This research received support from the SYNTHESYS Project http://www.synthesys.info/ which is financed by European Community Research Infrastructure Action under the FP6 “Structuring the European Research Area” Programme (grant to J.V.S., SE-TAF 794) We are also grateful to a number of other institutions and persons who kindly provided plant material: Françoise Prévost (French Guiana), the directors of the herbaria FR, MO, UPS, Edinburgh Botanical Garden (Scotland, UK) for material of *Medusagyne oppositifolia,* Bochum University Botanical Garden (Germany) for *Dovyalis caffra* and Osnabrück University Botanical Garden (Germany) for *Calophyllum inophyllum*; Rolf Rutishauser, University of Zürich (Switzerland) for material of Podostemaceae; Herbario Nacional de Bolivia, Stephan Beck, for duplicate sets of specimens of Caryocaraceae, Chrysobalanaceae, and Trigoniaceae; the Missouri Botanical Garden silica material collection; Elmar Robbrecht of the National Herbarium of Belgium (BR) for material of *Irvingia*; Kim Govers, Nees-Institute Bonn helped much with the sequencing on the CEQ8000 and the analysis of the sequences. Peter F. Stevens (Missouri Botanical Garden) provided helpful comments on an earlier version of the manuscript.

Appendix 1

Table 3

Table 3 Position of hotspots in individual sequences

Taxon	Pos. H1	Pos. H2	Pos. H3	Pos. H4	Pos. H5	Pos. H6	Pos. H7	Pos. H8	Pos. H9
Euonymus europaea	1–9	51–58	94–101	135–159	189–198	414–488	639–656	777–803	832–840
Salacia lehmbachii	1–29	70–77	113–120	154–176	196–205	421–525	676–693	817–835	864–888
Brexia madagascariensis	1–41	82–89	125–132	166–190	220–229	445–549	694–711	833–859	888–916
Parnassia palustris	1–11	48–55	91–98	136–155	193–202	421–539	677–694	819–846	883–906
Oxalis hedysaroides	–	39–46	86–96	130–149	179–193	406–483	620–641	768–794	826–847
Brunellia mexicana	–	41–48	88–99	133–152	182–193	407–467	596–613	735–761	794–811
Cephalotus follicularis	–	44–51	91–106	140–159	189–201	415–483	620–637	754–780	812–829
Crinodendron hookerianum	–	41–48	88–98	132–151	181–192	406–520	657–675	797–804	836–854

Table 3 continued

Taxon	Pos. H1	Pos. H2	Pos. H3	Pos. H4	Pos. H5	Pos. H6	Pos. H7	Pos. H8	Pos. H9
Hybanthus concolor	1–12	52–59	99–106	140–155	185–195	398–562	699–710	824–850	886–886
Hybanthus anomalus	1–4	44–51	91–98	131–152	176–186	398–606	743–754	877–893	931–983
Viola hederacea	1–30	70–77	117–134	168–184	214–224	437–575	712–723	847–872	911–953
Erythroxylum coca	1–7	48–55	95–104	138–157	187–197	400–512	650–668	794–831	881–926
Reinwardtia cicanoba	–	39–47	87–95	129–147	177–187	398–516	654–673	796–821	863–903
Linum narbonense	–	38–45	85–92	126–144	174–184	393–511	649–680	797–827	860–897
Linum catharticum	–	34–41	81–88	122–140	170–180	389–531	669–694	816–841	874–930
Ctenolophon englerianus	–	39–46	86–106	140–158	188–198	407–529	667–684	806–832	870–871
Phyllocosmus lemaireanus	1–15	59–65	105–112	146–167	197–208	417–523	662–679	801–836	875–916
Irvingia	–	36–43	83–98	132–156	186–196	405–543	688–699	821–847	873–913
Lacistema aggregatum	1–8	49–56	96–103	137–159	189–199	408–554	692–709	831–868	906–947
Lacistema nena	1–8	49–56	96–103	137–159	189–199	408–554	692–709	831–868	906–947
Malesherbia ardens	1–8	49–56	96–103	137–165	189–199	413–528	666–683	800–826	864–905
Turnera grandidentata	1–8	49–55	95–102	136–173	190–201	418–554	692–709	831–857	895–937
Passiflora quadrangularis	–	41–49	89–96	130–153	177–187	406–543	681–698	819–842	880–926
Populus alba	1–28	69–76	116–123	157–178	208–218	428–529	667–684	806–845	883–924
Salix purpurea	1–28	69–76	116–123	157–176	206–216	425–515	653–670	792–825	863–904
Salix reticulata	1–28	69–76	116–123	157–176	206–216	425–515	653–670	792–823	861–902
Dovyalis cafrra	1–22	63–70	110–117	151–171	201–211	420–514	652–669	789–810	848–881
Idesia polycarpa	1–22	63–70	110–117	151–175	205–215	430–558	696–711	827–855	893–934
Azara salicifolia	1–22	63–70	110–117	151–171	201–210	419–512	650–667	789–815	853–894
Flacourtia jangomas	1–22	63–70	110–117	151–172	202–212	421–521	659–676	798–824	862–895
Lindackeria paludosa	1–8	49–56	96–103	137–157	187–197	406–511	663–684	806–832	870–910
Hydnocarpus annamensis	1–8	49–56	40–47	81–106	128–138	347–512	650–667	789–815	853–900
Sacoglottis gabonensis	–	40–47	40–47	81–100	130–140	341–456	594–611	712–717	755–796
Vantanea compacta	–	40–47	40–47	81–100	130–140	341–469	607–624	746–772	810–851
Rhizophora mangle	–	37–45	40–55	101–135	165–175	391–503	651–668	811–837	883–935
Bruguiera gymnorhiza	–	40–49	40–47	93–114	156–166	382–494	632–656	778–804	906–979
Dichapetalum mossambicense	–	45–51	40–47	81–100	130–140	348–513	657–681	802–821	859–881
Picrodendron baccatum	1–4	45–52	40–47	81–100	137–147	358–452	590–607	724–750	788–805
Phyllanthus fluitans	–	32–35	40–49	83–101	131–141	350–430	568–585	702–719	757–822
Securinega suffruticosa	–	41–48	40–50	84–102	132–142	351–422	560–577	706–732	770–810
Calophyllum inophyllum	–	34–41	40–47	81–103	133–143	341–455	603–621	748–774	812–831
Hypericum hookerianum	–	38–45	40–47	81–123	153–165	359–495	628–647	798–829	869–914
Dicraeanthus africanus	1–12	56–63	40–47	81–101	131–143	349–486	625–657	782–797	841–871
Djinga felicis	1–12	56–63	40–47	81–100	130–143	348–495	634–666	801–816	860–890
Tristichia trifaria	–	40–47	40–47	91–145	177–192	397–549	689–728	873–901	939–1014
Clusia spec.	–	41–48	40–47	81–92	122–132	329–474	619–639	766–794	832–871
Garcinia tinctoria	–	41–48	40–46	88–98	128–138	336–450	595–626	750–774	817–857
Malpighia glabra	–	41–48	40–47	81–128	160–170	384–627	766–784	918–952	990–1045
Heteropterys chrysophylla	–	40–47	40–47	81–125	155–165	379–552	691–709	843–877	915–961
Bunchosia nitida	–	41–48	40–47	81–126	156–166	380–504	643–661	795–821	859–906
Licania kunthiana	–	41–47	40–47	81–99	129–140	356–506	679–704	826–826	826–841
Chrysobalanus icaco	–	41–47	40–47	81–100	130–141	357–502	675–700	822–836	836–837
Elatine hexandra	–	61–68	40–47	81–102	133–143	357–495	640–659	780–814	852–869
Trigonia nivea	–	41–47	40–47	81–94	124–134	342–552	698–723	845–864	896–926
Balanops balansae	–	41–47	40–47	81–100	130–140	349–446	584–601	723–750	782–812
Bergia suffruticosa	–	41–48	40–47	77–89	119–129	341–431	572–589	710–745	788–851

Table 3 continued

Taxon	Pos. H1	Pos. H2	Pos. H3	Pos. H4	Pos. H5	Pos. H6	Pos. H7	Pos. H8	Pos. H9
Medusagyne oppositifolia	–	41–48	40–48	82–105	135–145	361–463	598–612	733–759	797–836
Discladium spec.	–	46–53	41–50	84–120	150–160	373–459	597–618	731–757	808–863
Ochna serrulata	–	46–53	41–52	86–123	153–163	376–462	600–621	734–760	811–866
Acalypha hispida	1–14	44–51	40–47	81–122	156–166	406–531	681–703	813–848	908–967
Sapium spec.	–	40–47	40–51	85–104	134–144	342–486	630–647	777–794	854–902
Croton tiglium	–	39–46	40–47	81–101	131–141	350–519	661–679	799–825	869–916
Touroulia guianensis	–	41–48	40–55	89–114	144–154	370–462	600–617	738–764	802–842
Quiina intergrifolia	–	41–48	40–55	89–111	141–151	362–454	592–609	730–756	794–834
Aleurites fordtii	1–5	39–46	40–47	81–95	125–135	344–488	631–657	781–807	845–893
Andrachne colchica	–	41–48	40–47	81–99	134–144	357–462	602–619	737–763	801–857
Manihot esculenta	–	60–67	40–47	81–94	124–134	343–466	608–625	745–779	817–871
Euphorbia milii	1–14	55–62	43–50	84–122	152–162	369–437	586–603	728–757	790–841
Microdesmis puberula	1–11	52–59	40–47	81–100	130–140	349–495	633–650	772–798	857–902
Panda oleosa	1–22	63–70	40–47	81–100	130–140	349–479	617–634	749–775	831–872
Anthodiscus amazonicus	–	41–48	40–47	81–111	141–151	360–497	635–652	774–791	811–851
Caryocar brasiliense	–	41–48	40–47	81–112	142–152	361–484	622–639	761–788	826–866

H1 is not present in all taxa

Appendix 2

Table 4

Table 4 List of indels found in the *pet*D dataset

Indel No.	Position	Length	Sequence motif
petB–petD spacer			
1	4–7	4	"ATTT"— insertion in *Phyllocosmus*
2	11–47	37	Deletion in *Aleurites*
3	12–38	27	Gap in most taxa
4	12–46	35	Gap in *Linum catharticum*
5	31–38	8	"TACATTTA"—insertion in *Tristichia*
6	39–53	15	Gap in *Tristichia*
7	42–44	3	"TTC"—insertion in *Cephalotus*
8	42–52	11	Gap in *Calophyllum*
9	42–54	13	Gap in *Phyllanthus*
10	46–46	1	1 nt gap in *Linum narbonense*
11	47–47	1	1 nt gap in *Heteropterys*
12	48–52	5	Gap in *Hypericum*
13	51–51	1	"T"—insertion in *Euonymus*
14	54–77	24	Gap in *Parnassia*
15	57–79	23	Gap in *Rhizophora*
16	58–77	20	"AATATAGATCACAGACATTT"—insetion in *Elatine*
17	58–94	37	Gap in *Acalypha*
18	82–82	1	"A"—insertion in *Hypericum*
19	85–90	6	"AGGTGT"—insertion in *Tristichia*
20	98–98	1	"A" insertion in *Discladium* and *Ochna*
21	98–102	5	Gap in Podostemaceae
22	98–105	8	Gap in *Parnassia*
23	98–112	15	Gap in most Malpighiales
24	98–113	16	Gap in Violaceae
25	98–114	17	Gap in *Oxalis*
26	98–117	20	Gap in *Irvingia*
27	103–112	10	Gap in *Dichapetalum*
28	106–112	7	Gap in Podostemaceae
29	114–114	1	Gap in *Phyllocosmus*
30	114–115	2	Gap in Linaceae and some Euphorbiaceae
31	115–115	1	Gap in *Sacoglottis* and *Vantanea*
32	118–118	1	Gap in Rhizophoraceae
33	122–129	8	Gap in Celastrales and *Parnassia*
34	126–128	3	"AAA"—insertion in *Euphorbia*
35	126–129	4	Gap in most Malpighiales
36	129–129	1	"T"—insertion in *Discladium* and *Ochna*
37	174–174	1	"C"—insertion in *Garcinia*
38	187–198	12	Gap in *Hybanthus*
39	188–194	7	"CTTCAAC"—SSR in *Garcinia*
40	188–198	11	Gap in most Malpighiales
41	195–198	4	"TTTC"-SSR in *Parnassia*
42	205–214	10	"CGACCTCAAA"—insertion in *Tristichia*
43	205–226	22	Gap in most Malpighiales

Table 4 continued

Indel No.	Position	Length	Sequence motif
44	205–230	26	Gap in *Bergia*
45	215–226	12	"CTTTCATTTCAA"—insertion in Rhizophoraceae, probably multiple SSR
46	231–237	7	Gap in *Malesherbia*, *Turnera*, *Passiflora* and *Hydnocarpus*
47	231–244	14	Gap in *Acalypha*
48	231–245	15	Gap in *Salacia*
49	237–237	1	"T"—insertion in *Tristichia*
50	237–252	16	Gap in *Hybanthus anomalus*
51	241–241	1	"G" insertion in *Tristichia*
52	241–242	2	Gap in Malpighia
53	241–244	4	Gap in most Malpighiales
54	241–251	11	Gap in *Hydnocarpus*
55	242–244	3	Gap in *Tristichia*
56	243–244	2	Gap in *Elatine*
57	247–251	5	"TGAAT" insertion in *Andrachne*
58	255–258	4	Gap in *Parnassia*
59	255–266	12	"GCCATGAATAGT"—insertion in *Bruguiera*
60	269–275	7	"ATGGTTG" insertion in *Picrodendron*
61	283–309	27	Gap in *Turnera*
62	290–302	13	"AAAAAAAAAAATG"—insertion in *Acalypha*
63	290–307	18	Gap in Salicaceae
64	290–309	20	Gap in most taxa
65	303–309	7	Gap in *Acalypha*
66	308–309	2	"TG"—insertion in Salicaceae
petD intron			
67	351–352	2	2 nt—deletion in *Ochna* and *Discladium*
68	355–357	3	"ATG"—SSR in *Turnera*
69	355–362	8	"ATATG"—SSR in *Malesherbia* and *Passiflora*, "ATATA"—SSR in *Turnera*
70	355–364	10	Gap in most taxa
71	355–371	17	Gap in *Erythroxylum*
72	355–389	35	Gap in *Sacoglittis* and *Vantanea*
73	363–364	2	"TT"—insertion in *Picrodendron*
74	371–371	1	"A"—insertion in *Populus alba*
75	374–377	4	"TTAT"—insertion in Violaceae
76	374–383	10	Gap in *Sapium*
77	374–389	16	Gap in most Malpighiales
78	378–389	12	Gap in Violaceae
79	384–389	6	"TTTATC"—SSR in *Sapium*
80	405–405	1	1 nt—gap in *Djinga*
81	410–416	7	"TTCATAA"—SSR in Rhizophoraceae
82	410–421	12	Gap in most taxa
83	410–422	13	Gap in *Clusia* and *Garcinia*
84	417–421	5	"AATAA"—SSR in *Acalypha*
85	454–464	11	Gap in *Dichapetalum*
86	460–464	5	"TACTC"—insertion in *Passiflora*
87	468–472	5	"AGAAC"—insertion in Oxalidales
88	468–479	12	Gap in *Dicraeanthus* and *Djinga*
89	468–485	18	Gap in *Tristichia*
90	468–487	20	Gap in *Medusagyne*, *Discladium* and *Ochna*
91	468–492	25	Gap in *Parnassia*
92	468–494	27	Gap in most taxa
93	473–494	22	Gap in Oxalidaceae
94	480–494	15	Gap in Celastrales
95	493–494	2	Gap in Podostemaceae, Ochnaceae
96	506–507	2	"TG"—SSR in *Reinwardtia*
97	514–518	5	Gap in *Dicraeanthus* and *Djinga*
98	515–518	4	"TTCA"—SSR in *Andrachne*
99	524–530	7	SSR with motif "TGTTTGA" in *Licania* and "TGCTTGA" in *Chrysobalanus*
100	534–534	1	"A" insertion, probably from duplication in *Euphorbia*
101	537–542	6	Gap in *Trigonia*
102	538–542	5	"AAAAA"—insertion in *Acalypha*
103	538–543	6	Gap in *Euonymus*, *Brexia*, *Salacia* and *Oxalis*, probably deletion
104	557–581	25	Gap in *Hybanthus*, probably deletion
105	560–561	2	"GA"—SSR in *Irvingia*
106	560–566	7	Gap in most taxa
107	560–589	30	Gap in *Sapium*, probably deletion
108	560–622	63	Gap in *Hypericum*
109	562–566	5	"TCAGG"—SSR in Malpighiaceae
110	568–571	4	Gap in *Irvingia*
111	569–571	3	Gap in Podostemaceae and *Clusia*
112	570–571	2	"C" duplication in *Medusagyne* and Quiinaceae
113	571–571	1	"C" duplication in *Trigonia*, *Discladium* and *Ochna*
114	573–622	50	Gap in *Calyphyllum*, probably deletion
115	574–622	49	Gap in *Tristichia*, *Clusia* and *Garcinia*, probably deletion
116	575–580	6	Gap in most taxa
117	575–627	53	Gap in *Dicraeanthus* and *Djinga*, probably deletion
118	578–580	3	"AAT"—SSR in *Bergia*
119	580–580	1	Gap in *Elatine*
120	587–628	42	Gap in *Euphorbia*, probably deletion
121	590–593	4	Gap in *Acalypha*
122	590–614	25	Gap in *Parnassia*
123	590–622	33	Gap in most taxa

Table 4 continued

Indel No.	Position	Length	Sequence motif
124	590–623	34	Gap in *Hybanthus*
125	594–628	35	Gap in *Sapium*
126	615–622	8	Gap in *Acalypha*
127	629–629	1	1 nt gap in *Trigonia*
128	633–633	1	"G"—duplication in *Euonymus*, *Salacia* and *Brexia*
129	656–657	2	Gap in *Calophyllum*
130	656–662	7	"ATAGTAT"—SSR in *Irvingia*
131	665–668	4	"ATAG"—SSR in *Rhizophora*
132	674–687	14	"ATAGTATGCAAATG" insertion in *Lindackeria*
133	711–715	5	"CAATT" insertion in *Calophyllum*
134	717–723	7	SSR—like insertion "YWTTTAT" in *Euonymus*, *Bexia*, *Salacia*, *Parnassia*. No clear template motif for this repeat
135	733–738	6	"TATTAA"—SSR in *Salacia*
136	733–744	12	Gap in most taxa
137	739–744	6	"TTATGA"—SSR in *Rhizophora*
138	751–751	1	"A" insertion in *Elatine*
139	760–764	5	"AGTGA"—SSR in Euphorbiaceae
140	774–779	6	"TCTAGA"—SSR in *Euphorbia*
141	791–797	7	"AAGAATG"—SSR in *Clusia* and *Garcinia*
142	791–798	8	Gap in most taxa
143	798–798	1	"T" insertion in Malpighiaceae
144	803–803	1	1 nt gap in *Manihot*
145	814–815	2	"CA"—insertion in *Tristichia*
146	815–815	1	"C" duplication in *Dicraeantus* and *Djinga*
147	820–820	1	"T" duplication in *Hypericum*, *Sapium* and *Irvingia*
148	835–835	1	"G" insertion in *Phyllocosmus*
149	841–841	1	"T" duplication in *Dichapetalum* and *Trigonia*
150	841–847	7	Gap in *Elatine* and *Bergia*
151	841–853	13	Gap in Chrysobalanaceae
152	841–888	48	Gap in *Euonymus*
153	841–894	54	Gap in most taxa
154	842–894	53	Gap in *Dichapetalum* and *Trigonia*
155	848–916	69	Gap in *Clusia* and *Garcinia*
156	854–894	41	Gap in *Elatine* and *Bergia*
157	889–894	6	"GTATGT"—SSR in *Euonymus*
158	901–907	7	"AAAATAA"—insertion in *Trigonia*
159	901–909	9	Gap in *Acalypha*
160	901–916	16	Gap in most taxa
161	901–917	17	Gap in *Croton*
162	901–923	23	Gap in *Parnassia*
163	901–925	25	Gap in *Brunellia*
164	906–916	11	Gap in *Dichapetalum*

Table 4 continued

Indel No.	Position	Length	Sequence motif
165	908–916	9	Gap in *Trigonia*
166	910–916	7	Gap in *Andrachne*
167	919–924	6	Deletion in *Hypericum*
168	924–924	1	1 nt—deletion in *Oxalis*, *Cephalotus* and *Crinodendron*
169	928–930	3	2 nt—deletion in *Bergia* and *Medusagyne*
170	930–930	1	1 nt—deletion in Violaceae and *Irvingia*
171	931–936	6	Deletion in *Idesia*
172	933–937	5	Deletion in *Malesherbia*, *Turnera* and *Passiflora*
173	934–936	3	Deletion in *Garcinia*
174	935–935	1	1 nt—deletion in *Quiina* and *Touroulia*
175	937–937	1	1 nt—deletion in *Medusagyne*
176	942–946	5	"AAAAG"—insertion in *Dicraeanthus* and *Djinga*
177	966–966	1	1 nt—gap in *Dichapetalum*
178	968–979	12	Deletion in *Discladium* and *Ochna*
179	970–972	3	"TTT"—insertion in Violaceae
180	970–981	12	Deletion in *Tristichia*
181	976–976	1	1 nt—gap in Violaceae
182	983–987	5	"TATCA"—insertion in *Oxalis*
183	983–995	13	Gap, probably deletion in *Hybanthus*
184	983–1010	28	Deletion in in *Dicraeanthus* and *Djinga*
185	992–995	4	"TTTT"—insertion in *Turnera*
186	997–997	1	"C"—duplication in *Rhenwardtia*
187	997–1008	12	Gap in *Croton*
188	999–999	1	"T"—duplication in *Salacia*
189	999–1004	6	Gap in *Acalypha*
190	999–1008	10	Gap in most taxa
191	1000–1008	9	Gap in *Salacia*
192	1011–1014	4	Gap, probably deletion in *Linum narbonense*
193	1013–1014	2	Deletion in in *Dicraeanthus* and *Djinga*
194	1014–1014	1	"A"—duplication in *Erythroxylum*
195	1016–1022	7	"GCCTACT"—insertion in *Phyllanthus* and *Securinega*
196	1024–1028	5	"ATTGG"—SSR in *Hypericum*
197	1030–1030	1	1 nt—gap in *Euonymus*
198	1052–1053	2	"TA"—SSR in *Parnassia*
199	1052–1057	6	Gap in most taxa
200	1054–1057	4	Gap in *Aleurites*
201	1061–1062	2	Gap in *Parnassia*
202	1062–1062	1	"A"—duplication in *Turnera*
203	1062–1127	66	Gap in *Acalypha*
204	1067–1080	14	"ACTTGTAAGATAAG"—SSR in Malpighiaceae
205	1067–1089	23	Gap in *Tristichia*
206	1067–1105	39	Gap in *Hypericum*

Table 4 continued

Indel No.	Position	Length	Sequence motif
207	1067–1106	40	Gap in *Dicraeanthus* and *Djinga*
208	1067–1112	46	Gap in *Calophyllum*
209	1067–1117	51	Gap in *Garcinia* and *Clusia*
210	1067–1122	56	Gap in most taxa
211	1081–1122	42	Gap in Malpighiaceae
212	1086–1122	37	Gap in *Hybanthus anomalus*
213	1090–1122	33	Gap in *Erythroxylum*
214	1111–1111	1	Deletion in *Tristichia*
215	1118–1122	5	Gap in *Calophyllum*
216	1124–1124	1	Deletion in *Dicraeanthus* and *Djinga*
217	1124–1139	16	Deletion in *Hybanthus*
218	1126–1126	1	Deletion in some Euphorbiaceae
219	1126–1127	2	Deletion in *Linum narbonense*
220	1128–1137	10	Gap in *Picrodendron*
221	1130–1134	5	“TTTCA”—insertion in *Rhizophora*
222	1130–1146	17	Gap in *Parnassia*
223	1141–1141	1	1 nt—deletion in *Hybanthus anomalus*
224	1141–1188	48	Deletion in *Phyllanthus*
225	1141–1210	70	Deletion in *Sacoglottis*
226	1142–1183	42	Gap in *Dicraenthus*
227	1147–1162	16	Gap in *Rhizophora*
228	1147–1178	32	Gap in *Dicraeanthus* and *Djinga*
229	1147–1187	41	Gap in most taxa
230	1163–1190	28	Gap in Parnassia
231	1179–1187	9	Gap in *Rhizophora*
232	1189–1190	2	Gap in Malpighiaceae
233	1189–1203	15	Gap in *Panda*
234	1190–1190	1	“A”—duplication in *Salacia*
235	1190–1203	14	Deletion in *Erythroxylum*
236	1190–1206	17	Deletion in *Tristichia*
237	1196–1203	8	Deletion in Malpighiaceae and Elatinaceae
238	1197–1203	7	“ATGGTTG”—insertion in *Hypericum*
239	1197–1210	14	Gap in *Dicraeanthus* and *Djinga*
240	1199–1203	5	Gap in *Aleurites*
241	1206–1210	5	Deletion in *Cephalotus*
242	1207–1210	4	Deletion in *Andrachne*
243	1209–1210	2	Deletion in *Dovyalis*
244	1211–1292	82	Gap in *Anthodiscus*
245	1211–1425	215	Deletion in *Licania* and *Chrysobalanus*
246	1214–1214	1	“T”—duplication in *Brunellia*
247	1222–1243	22	Gap in *Dichapetalum*
248	1223–1226	4	Gap in *Discladium*, *Ochna* and *Acalypha*
249	1223–1231	9	Gap in *Rhizophora*
250	1223–1234	12	Gap in *Bergia*
251	1223–1243	21	Gap in most taxa
252	1223–1251	29	Gap in *Tristichia*
253	1224–1243	20	Gap in *Phyllocosmus*
254	1227–1243	17	Gap in *Croton*
255	1240–1243	4	“TGAT”—insertion in *Acalypha*
256	1247–1256	10	Gap in most taxa
257	1247–1393	147	Gap in *Irvingia*
258	1249–1256	8	Gap in *Croton*
259	1252–1256	5	Gap in *Acalypha*
260	1253–1256	4	Gap in *Parnassia*
261	1261–1265	5	“ATTCA”—SSR in *Sapium*
262	1261–1269	9	Gap in *Erythroxylum*
263	1261–1281	21	Gap in *Dicraeanthus* and *Djinga*
264	1261–1287	27	Gap in *Garcinia*
265	1261–1292	32	Gap in most taxa
266	1261–1295	35	Gap in *Cephalotus* and *Crinodendron*
267	1261–1298	38	Gap in *Euphorbia*
268	1261–1397	137	Gap in *Oxalis*
269	1266–1292	27	Gap in *Sapium*
270	1270–1292	23	Gap in *Panda*
271	1282–1292	11	Gap in *Erythroxylum*
272	1288–1292	5	Gap in *Dicraeanthus* and *Djinga*
273	1290–1292	3	Gap in *Brunellia*
274	1294–1299	6	Deletion in Linaceae
275	1295–1295	1	“A”—duplication in *Viola*
276	1297–1417	121	Gap in Celastrales
277	1299–1308	10	Gap in *Clusia*
278	1299–1417	119	Gap in *Parnassia*
279	1302–1308	7	“ATATGTG”—SSR in *Microdesmis*
280	1302–1417	116	Gap in *Cephalotus* and *Crinodendron*
281	1302–1419	118	Gap in *Trigonia* and *Balanops*
282	1302–1421	120	Gap *Brunellia*
283	1310–1311	2	“AT”—SSR in *Hypericum*
284	1310–1312	3	Gap in most taxa
285	1312–1312	1	“G”—insertion in *Tristichia*
286	1316–1379	64	Multiple SSR with sequence motif “TTCATATATGTGTAGA” in *Bruguiera*
287	1316–1384	69	Gap in *Reinwardtia*
288	1316–1403	88	Gap in *Microdesmis* and *Panda*
289	1316–1417	102	Gap in most taxa
290	1380–1417	38	Gap in *Bruguiera*
291	1385–1421	37	Gap in *Dichapetalum*
292	1394–1417	24	Gap in *Reinwardtia*
293	1398–1417	20	Gap in *Irvingia*
294	1404–1417	14	Gap in *Oxalis*
295	1424–1425	2	Deletion in *Hybanthus concolor*
296	1426–1431	6	Deletion in *Oxalis*
297	1426–1433	8	Deletion in *Ctenolophon*
298	1426–1437	12	Deletion in *Euonymus*

Table 4 continued

Indel No.	Position	Length	Sequence motif
299	1430–1430	1	"G"—dulpication in *Parnassia*
300	1430–1431	2	Deletion in *Cephalotus*
301	1433–1441	9	Deletion in *Microdesmis* and *Panda*
302	1434–1438	5	Deletion in *Garcinia* and *Clusia*
303	1439–1450	12	Deletion in *Djinga*
304	1440–1440	1	1 nt deletion in *Hypericum*
305	1440–1450	11	Deletion in *Dicraeanthus*
306	1442–1443	2	Deletion in *Calophyllum*
307	1443–1443	1	1 nt deletion in *Salacia*
308	1511–1526	16	Gap in *Dicraeanthus* and *Djinga*
309	1512–1524	13	"GAGGATGGATTTA"—SSR in *Lindackeria*
310	1526–1527	2	Deletion in *Medusagyne*, *Ochnaceae* and *Quiinaceae*

References

APG (1998) An ordinal classification for the families of flowering plants. Ann Missouri Bot Gard 85:531–553

APG (2003) An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. Bot J Linn Soc 141:339–436

Barkman TJ, Lim S-H, Nais J (2004) Mitochondrial DNA sequences reveal the photosynthetic relatives of *Rafflesia*, the world's largest flower. Proc Natl Acad Sci USA 101:787–792

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

Borsch T, Hilu KW, Wiersema JH, Lohne C, Barthlott W, Wilde V (2007) Phylogeny of *Nymphaea* (Nymphaeaceae): evidence from substitutions and microstructural changes in the chloroplast *trn*T-*trn*F region. Int J Pl Sci 168:639–671

Borsch T, Löhne C, Müller K, Hilu KW, 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 Leopold NF 92:85–110

Cameron KM, Chase MW, Anderson WR, Hillis HG (2001) Molecular systematics of Malpighiaceae: evidence from plastid *rbc*L and *mat*K sequences. Amer J Bot 88:1847–1862

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedrén M, Gaut BS, Jansen RK, Kim KJ, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang QY, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn GH, Graham SW, Barrett SCH, Dayanandan S, Albert VA (1993) Phylogenetics of seed plants—an analysis of nucleotide-sequences from the plastid gene *rbc*L. Ann Missouri Bot Gard 80:528–580

Chase MW, Zmartzy S, Lledó MD, Wurdack KJ, Swensen SM, Fay MF (2002) When in doubt, put it in Flacourtiaceae: a molecular analysis based on plastid *rbc*L sequences. Kew Bull 57:141–181

Cronquist A (1981) An integrated system of clasification of flowering plants. Columbia University Press, New York

Davis C, Wurdack KJ (2004) Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. Science 305:676–678

Davis CC, Anderson WR, Donoghue MJ (2001) Phylogeny of Malpighiaceae: evidence from chloroplast *ndh*F and *trn*L-F nucleotide sequences. Amer J Bot 88:1830–1846

Davis CC, Chase MW (2004) Elatinaceae are sister to Malpighiaceae; Peridiscaceae belong to Saxifragales. Amer J Bot 91:262–273

Davis CC, Latvis M, Nickrent DL, Wurdack KJ, Baum DA (2007) Floral gigantism in Rafflesiaceae. Science 315:1812–1812

Davis CC, Webb CO, Wurdack KJ, Jaramillo CA, Donoghue MJ (2005) Explosive radiation of Malpighiales supports a mid-Cretaceous origin of modern tropical rain forests. Amer Naturalist 165:E36–E65

De Rijk P, Wuyts J, De Wachter R (2003) RnaViz2: an improved representation of RNA secondary structure. Bioinformatics 19:299–300

Fay MF, Swensen SM, Chase MW (1997) Taxonomical affinities of *Medusagyne oppositifolia* (Medusagynaceae). Kew Bull 52:111–120

Graham SW, Reeves PA, Burns ACE, Olmstead RG (2000) Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. Int J Pl Sci 161:S83–S96

Gustafsson MHG, Bittrich V, Stevens PF (2002) Phylogeny of Clusiaceae based on *rbc*L sequences. Int J Pl Sci 163:1045–1054

Hausner G, Olsen R, Johnson I, Simone D, Sanders ER, Karol KG, McCourt RM, Zimmerly S (2006) Origin and evolution of the chloroplast *trn*K (*mat*K) intron: a model for evolution of group II intron RNA structures. Molec Biol Evol 23:380–391

Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R, Sauquet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS, Chatrou LW (2003) Angiosperm phylogeny based on *mat*K sequence information. Amer J Bot 90:1758–1776

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Kathriarachchi H, Hoffmann P, Samuel R, Wurdack KJ, Chase MW (2005) Molecular phylogenetics of Phyllanthaceae inferred from five genes (plastid *atp*B, *mat*K, 3 '*ndh*F, *rbc*L, and nuclear *PHYC*). Molec Phylogenet Evol 36:112–134

Kelchner SA (2000) The evolution of non-coding chloroplast DNA and its application in plant systematics. Ann Missouri Bot Gard 87:482–498

Kelchner SA (2002) Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. Amer J Bot 89:1651–1669

Kelchner SA, Wendel JF (1996) Hairpins create minute inversions in non-coding regions of chloroplast DNA. Curr Genet 30:259–262

Kita Y, Kato M (2001) Infrafamilial phylogeny of the aquatic angiosperm Podostemaceae inferred from the nucleotide sequences of the *mat*K gene. Pl Biol 3:156–163

Krosnick SE, Harris EM, Freudenstein JV (2006) Patterns of anomalous floral development in the Asian *Passiflora* (subgenus Decaloba: supersection Disemma). Amer J Bot 93:620–636

Lehmann K, Schmidt U (2003) Group II introns: structure and catalytic versatility of large natural ribozymes. Crit Rev Biochem Mol 38:249–303

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

Löhne C, Borsch T (2005) Molecular evolution and phylogenetic utility of the *pet*D group II intron: a case study in basal angiosperms. Molec Biol Evol 22:317–332

Löhne C, Borsch T, Wiersema JH (2007) Phylogenetic analysis of Nymphaeales using fast-evolving and noncoding chloroplast markers. Bot J Linn Soc 154:141–163

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci USA 101:7287–7292

Mathews DH, Schroeder SJ, Turner DH, Zuker M (2006) Predicting RNA secondary structure. In: Gesteland RF, Cech TR, Atkins JF (eds) The RNA World. The nature of modern RNA suggests a prebiotic RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 631–656

Mathews DH, Zuker M, Turner DH (1996–2006) RNAstructure 4.3

Merino Sutter D, Forster PI, Endress PK (2006) Female flowers and systematic position of Picrodendraceae (Euphorbiaceae s.l., Malpighiales). Pl Syst Evol 261:187–215

Michel F, Dujon B (1983) Conservation of RNA secondary structures in two intron families including mitochondrial-encoded, chloroplast-encoded and nuclear-encoded members. EMBO J 2:33–38

Michel F, Umesono K, Ozeki H (1989) Comparative and functional anatomy of group II catalytic introns—a review. Gene 82:5–30

Moline P, Thiv M, Ameka GK, Ghogue JP, Pfeifer E, Rutishauser R (2007) Comparative morphology and molecular systematics of African Podostemaceae-Podostemoideae, with emphasis on *Dicraeanthus* and *Ledermanniella* from Cameroon. Int J Pl Sci 168:159–180

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 K (2004) PRAP-computation of Bremer support for large data sets. Molec 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 datasets. Appl Bioinformatics 4:65–69

Müller K, Borsch T (2005) Phylogenetics of *Utricularia* (Lentibulariaceae) and molecular evolution of the *trn*K intron in a lineage with high substitutional rates. Pl Syst Evol 250:39–67

Müller K, Borsch T, Hilu KW (2006) Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: Contrasting *mat*, *trn*T-F, and *rbc*L in basal angiosperms. Molec Phylogenet Evol 41:99–117

Nandi OI, Chase MW, Endress PK (1998) A combined cladistic analysis of angiosperms using *rbc*L and non-molecular data sets. Ann Missouri Bot Gard 85:137–212

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818

Pyle AM, Fedorova O, Waldsich C (2007) Folding of group II introns: a model system for large, multidomain RNAs? Trends Biochem Sci 32:138–145

Pyle AM, Lambowitz AM (2006) Group II introns: ribozymes that splice RNA and invade DNA. In: Gesteland RF, Cech TR, Atkins JF (eds) The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 449–505

Qin PZ, Pyle AM (1998) The architectural organization and mechanistic function of group II intron structural elements. Curr Opin Struct Biol 8:301–308

Qiu Y-L, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M (2000) Phylogeny of basal angiosperms: analyses of five genes from three genomes. Int J Pl Sci 161(6 Suppl):3–27

Quandt D, Müller K, Stech M, Frahm J-P, Frey W, Hilu KW, Borsch T (2004) Molecular evolution of the chloroplast *trn*L-F region in land plants. In: Goffinet B, Hollowell V, Magill R (eds) Molecular systematics of bryophytes, vol 98. Missouri Botanical Garden, St Louis, pp 13–37

Savolainen V, Chase MW, Hoot SB, Morton CM, Soltis DE, Bayer C, Fay MF, De Bruijn AY, Sullivan S, Qiu YL (2000a) Phylogenetics of flowering plants based on combined analysis of plastid *atp*B and *rbc*L gene sequences. Syst Biol 49:306–362

Savolainen V, Fay MF, Albach DC, Backlund A, Van der Bank M, Cameron KM, Johnson LA, Lledó MD, Pintaud J-C, Powell M, Sheaham MC, Soltis DE, Soltis PS, Weston P, Whitten WM, Wurdack KJ, Chase MW (2000b) Phylogeny of the eudicots: a nearly complete familial analysis based on *rbc*L gene sequences. Kew Bull 55:257–309

Schneider JV, Swenson U, Ramuel R, Stuessy T, Zizka G (2006) Phylogenetics of Quiinaceae (Malpighiales): evidence from *trn*L-*trn*F sequence data and morphology. Pl Syst Evol 257:189–203

Schwarzbach AE, Ricklefs RE (2000) Systematic affinities of Rhizophoraceae and Anisophyllaceae, and intergenic relationships within Rhizophoraceae, based on chloroplast DNA, nuclear ribosomal DNA, and morphology. Amer J Bot 87:547–564

Setoguchi H, Kosuge K, Tobe H (1999) Molecular phylogeny of Rhizophoraceae based on *rbc*L sequences. J Pl Res 112:443–455

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. Syst Biol 49:369–381

Soltis DE, Mort ME, Soltis PS, Hibsch-Jetter C, Zimmer EA, Morgan D (1999) Phylogenetic relationships of the enigmatic angiosperm family Podostemaceae inferred from 185 rDNA and *rbc*L sequence data. Molec Phylogenet Evol 11:261–272

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS (2000) Angiosperm phylogeny inferred from 18S rDNA, *rbc*L, and *atp*B sequences. Bot J Linn Soc 133:381–461

Stevens PF (2001 onwards) Angiosperm Phylogeny Website. Version 7, May 2006 http://www.mobot.org/MOBOT/research/APweb/

Swofford DL (1998) PAUP*. Phylogenetic Analysis Using Parsimony (*and other Methods). Sinauer Associates, Sunderland

Takhtajan A (1997) Diversity and classification of flowering plants. Columbia University Press, New York

Tokuoka T, Tobe H (2006) Phylogenetic analyses of Malpighiales using plastid and nuclear DNA sequences, with particular reference to the embryology of Euphorbiaceae s. str. J Pl Res 119:599–616

Toor N, Hausner G, Zimmerly S (2001) Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. RNA 7:1142–1152

Worberg A, Quandt D, Barniske A-M, Löhne C, Hilu KW, Borsch T (2007) Phylogeny of basal eudicots: Insights from non-coding and rapidly evolving DNA. Org Divers Evol 7:55–77

Wurdack KJ, Hoffmann P, Chase MW (2005) Molecular phylogenetic analysis of uniovulate Euphorbiaceae (Euphorbiaceae sensu stricto) using plastid *rbc*L and *trn*L-F DNA sequences. Amer J Bot 92:1397–1420

Zanis MJ, Soltis DE, Soltis P, Mathews S, Donoghue MJ (2002) The root of the angiosperms revisited. Proc Natl Acad Sci USA 99:6848–6853