Molecular phylogenetics and character evolution of Cannabaceae

Mei-Qing Yang,[1,2,3] Robin van Velzen,[4,5] Freek T. Bakker,[4] Ali Sattarian,[6] De-Zhu Li[1,2] & Ting-Shuang Yi[1,2]

1 *Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, P.R. China*
2 *Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, P.R. China*
3 *University of Chinese Academy of Sciences, Beijing 100093, P.R. China*
4 *Biosystematics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands*
5 *Laboratory of Molecular Biology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands*
6 *Department of Natural Resources, Gonbad University, Gonbad Kavous 4971799151, Iran*
Authors for correspondence: *Ting-Shuang Yi, tingshuangyi@mail.kib.ac.cn; De-Zhu Li, dzl@mail.kib.ac.cn*

Abstract Cannabaceae includes ten genera that are widely distributed in tropical to temperate regions of the world. Because of limited taxon and character sampling in previous studies, intergeneric phylogenetic relationships within this family have been poorly resolved. We conducted a molecular phylogenetic study based on four plastid loci (*atpB-rbcL*, *rbcL*, *rps16*, *trnL-trnF*) from 36 ingroup taxa, representing all ten recognized Cannabaceae genera, and six related taxa as outgroups. The molecular results strongly supported this expanded family to be a monophyletic group. All genera were monophyletic except for *Trema*, which was paraphyletic with respect to *Parasponia*. The *Aphananthe* clade was sister to all other Cannabaceae, and the other genera formed a strongly supported clade further resolved into a *Lozanella* clade, a *Gironniera* clade, and a trichotomy formed by the remaining genera. Morphological ancestral state reconstructions indicated the complex evolution pattern of most analyzed morphological characters, and it is difficult to identify morphological synapomorphies for most clades within Cannabaceae.

Keywords Cannabaceae; character evolution; classification; molecular phylogenetics

Supplementary Material The Electronic Supplement (Figs. S1–S2) is available in the Supplementary Data section of the online version of this article (http://www.ingentaconnect.com/content/iapt/tax).

Received: 31 May 2012; revision received: 12 Sept. 2012; accepted 16 Apr. 2013

INTRODUCTION

Cannabaceae was first separated from Moraceae by Rendle (1925) and contained only *Cannabis* L. and *Humulus* L. The circumscription of this family has been expanded significantly to include most members of Ulmaceae subfam. Celtidoideae sensu Engler & Prantl (1893) or Celtidaceae sensu Link (1829). *Ampelocera* Klotzsch has been excluded from the group and placed in Ulmaceae following a series of molecular phylogenetic studies (Ueda & al., 1997; Wiegrefe & al., 1998; Sytsma & al., 2002). Cannabaceae contains ten genera (Sytsma & al., 2002; Mabberley, 2008) and 109 accepted and 71 putative species (The Plant List, 2010) (Table 1).

Cannabaceae comprises taxa varying greatly in terms of habit and morphology. Most are trees and shrubs, but the family also includes herbs (*Cannabis*) and vines (*Humulus*). Fruits are usually drupes, but samaras occur in *Pteroceltis* Maxim. and achenes in *Cannabis* and *Humulus*. Leaves are usually alternate, but opposite in *Lozanella* Greenm., and both opposite and alternate in *Cannabis* and *Humulus*. Although the morphological synapomorphies of Cannabaceae are not clear, some morphological characters including usually unisexual and inconspicuous flowers, antitepalous stamens, the presence of stipules, diporate or triporate pollen, and free filaments slightly adnate to the tepals can be used to identify this family (Judd & al., 2008). Most genera have restricted distributions, although *Aphananthe* Planch., *Celtis* L. and *Trema* Lour. are widely distributed in tropical and temperate regions (Table 1).

The modern circumscription of Cannabaceae was first proposed by Wiegrefe & al. (1998), who carried out parsimony analyses of chloroplast DNA restriction site data. This conclusion has been supported by a subsequent parsimony analysis of the chloroplast gene *matK* (Song & al., 2001), parsimony analysis of the plastid regions *rbcL*, *trnL-trnF* and *ndhF* (Sytsma & al., 2002), and Bayesian and parsimony analyses of the plastid regions *rbcL* and *trnL-trnF* (Van Velzen & al., 2006).

Cannabaceae is sister to Moraceae and Urticaceae (Sytsma & al., 2002; Van Velzen & al., 2006; Wang & al., 2009; Zhang & al., 2011). However, phylogenetic relationships within Cannabaceae are still largely unresolved. A phylogenetic analysis based on 33 morphological, palynological, biochemical, and cytogenetic characters suggested that *Broussonetia* L'Hér. ex Vent. of Moraceae is nested within Celtidaceae, a surprising phylogenetic relationship (Zavada & Kim, 1996). This study also suggested a sister-group relationship of *Chaetachme* Planch. and *Gironniera* Gaudich., which was surprising because the two genera show strong morphological divergence (Zavada & Kim, 1996). Some previous molecular studies

Table 1. Genera currently included in Cannabaceae with their species numbers and geographic range.

Genera	Recognized species (putative species)	Species sampled (individuals sampled)	Distribution
Aphananthe	5	3	Madagascar, southwestern China to Japan, Malaysia, Indonesia and east Australia, Mexico
Cannabis	1	1 (2)	Asia, cultivated worldwide
Celtis	73 (36)	6(7)	Tropical (most), temperate (Europe 4)
Chaetachme	1	1 (3)	Tropical Africa, southern Africa and Madagascar
Gironniera	6	2	Malaysia, Indonesia to Pacific
Humulus	3	3	North temperate
Lozanella	2	2	Tropical areas of America
Parasponia	5 (5)	2	Malaysia, west Pacific
Pteroceltis	1	1 (2)	North and central China
Trema	12 (30)	8 (10)	Worldwide in tropical and warm climates
Total	109 (71)	29 (36)	

suggested that *Cannabis* and *Humulus* are sister taxa (Song & al., 2001; Sytsma & al., 2002), that *Trema* is paraphyletic with respect to *Parasponia* (Sytsma & al., 2002; Yesson & al., 2004), and that *Aphahanthe* is sister to the remainder of the family (Song & al., 2001; Sytsma & al., 2002), but other intergeneric relationships in Cannabaceae have remained unresolved (or were weakly supported because of limited taxon and character sampling). Van Velzen & al. (2006) analysed the phylogeny of Cannabaceae based on two plastid markers (*rbcL*, *trnL-trnF*). They obtained stronger support for intergeneric relationships and concluded that *Chaetachme* and *Pteroceltis* were distinct genera, and recovered a large, strongly supported clade containing *Cannabis*, *Celtis*, *Chaetachme*, *Humulus*, *Parasponia*, *Pteroceltis* and *Trema*.

The current study included the three monotypic genera (*Cannabis*, *Chaetachme*, *Pteroceltis*) of the family and sampled multiple species of the other seven genera. Each sample was sequenced for four plastid loci to estimate phylogenetic relationships in Cannabaceae. The major goals of this study were to reconstruct intergeneric relationships in the family and to optimise selected morphological characters on the inferred phylogeny to identify potential synapomorphies for clades.

■ MATERIALS AND METHODS

Sampling. — Thirty-six individuals representing twenty-nine species of all ten recognized genera of Cannabaceae were included, which is 26.6% of the accepted species (Table 1). We sampled six species of *Celtis* (representing 8.2% of the accepted species) across its distribution range (Eurasia, Africa, North America, South America). Eight species of *Trema* (representing 66.7% of the accepted species) were included. For the monotypic genera *Cannabis*, *Pteroceltis* and *Chaetachme*, two to three samples each were included. Two to three species were sampled for the remaining genera (*Aphananthe*, *Gironniera*, *Humulus*, *Lozanella*, *Parasponia*). Six species from the closely related Moraceae, Urticaceae and Ulmaceae (all Rosales) were chosen as outgroups based on the recent phylogenetic study of Zhang & al. (2011). Genera currently included in Cannabaceae with their species numbers and geographic ranges are shown in Table 1.

DNA extraction, PCR and sequencing. — Total genomic DNA was extracted from silica gel-dried leaves or herbarium specimens following the CTAB protocol of Doyle & Doyle (1987).

Nucleotide sequences for the four chloroplast loci, i.e., *atpB-rbcL* spacer, *rbcL*, *rps16* and *trnL-trnF* spacer were generated using the following primers for both PCR amplification and sequencing: atpB-F and rbcL-R for the *atpB-rbcL* region (Chiang & al., 1998); rbcL-30F and rbcL-1400R for the *rbcL* region (Zhang & al., 2011); f and 2r for the *rps16* region (Oxelman & al., 1997); and c and f for the *trnL-trnF* region (Taberlet & al., 1991).

Polymerase chain reaction (PCR) amplifications were performed in a volume of 25 μl containing 10–50 ng of genomic DNA, 0.2 μmol of each primer, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μmol of each dNTP and 0.5 U *Taq* polymerase (Takara, Shanghai, China). The PCR cycling parameters for all regions were as follows: a 95°C initial hot start for 5 min, followed by 32 cycles of 94°C for 30 s, 50°C for 40 s and 72°C for 60 s, and a final extension of 72°C for 10 min. PCR products were isolated and purified using a commercial DNA purification kit (Sangon Inc., Shanghai, China) following the manufacturer's protocols. Cycle sequencing was carried out using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit with 5 ng of primer, 1.5 μl of sequencing dilution buffer, and 1 μl of cycle sequencing mix in a 10 μl reaction volume. Cycle sequencing conditions were as follows: 30 cycles of 30 s denaturation (96°C), 30 s annealing (50°C), and 4 min elongation (60°C). The samples were sequenced on an ABI 3700xl DNA analyzer. We sequenced

both strands of DNA with overlapping regions to ensure unambiguous base calls.

Sequences were initially assembled using Sequencher v.4.2 (GeneCodes Corporation, Ann Arbor, Michigan, U.S.A.) and aligned using Clustal X (Thompson & al., 1997), followed by manual adjustments using Se-Al v.2.0 (Rambaut, 2002). Potentially informative indels in regions of unambiguous alignment were scored following the simple indel coding method (Simmons & Ochoterena, 2000), which treats each gap as a single presence/absence character independent of its length. A total of 82 gap characters were coded (*trnL-trnF*: 24; *atpB-rbcL*: 27; *rps16*: 31) for inclusion in the parsimony analysis, 75 of which were phylogenetically informative. The matrix is available in TreeBASE (study accession number 13867) and all new sequences generated in this study have been deposited in GenBank under accession numbers JN040281 to JN040432 (Appendix 1).

Phylogenetic analyses. — Phylogenetic analyses were conducted on each separate gene region as well as the combined plastid datasets, with gaps treated as missing data and indels coded as binary characters (simple indel coding). Phylogenetic relationships were inferred using maximum parsimony (MP) as implemented in PAUP* v.4.0b10 (Swofford, 2003), Bayesian inference (BI) as implemented in MrBayes v.3.1.2 (Ronquist & Huelsenbeck, 2003), and maximum-likelihood (ML) as implemented in Garli v.0.96 (Zwickl, 2006).

The MP analyses used heuristic searches with 1000 random sequence addition replicates, tree bisection-reconnection (TBR) branch swapping, and MULTREES on. All character states were treated as unordered and equally weighted. To evaluate the relative robustness of clades in the MP trees, bootstrap analysis (Felsenstein, 1985) was performed with 1000 replicates using the same options as above except that a maximum of 100 trees were saved per random sequence addition replicate.

For BI and ML analyses, Modeltest v.3.7 (Posada & Crandall, 1998) was run for each dataset to select the best model of sequence evolution for each gene. The models were chosen by the Akaike information criterion (AIC) and determined by AIC scores. For coded indels, we modeled indels as evolving according to a stochastic binary model. Considering the importance of data partitioning (Brown & Lemmon, 2007), the combined plastid dataset was analyzed by applying separate models to each data partition, with all parameters unlinked across data partitions except for topology and branch length. For the Bayesian inference, one cold and three incrementally heated Markov chain Monte Carlo (MCMC) chains were run for 2,000,000 generations. Trees were sampled every 100 generations. MCMC runs were repeated twice to avoid spurious results. Stationarity of the Markov chain was ascertained by plotting and interpreting likelihood values against number of generations in Tracer v.1.3 (Rambaut & Drummond, 2004). The first 5000 trees were discarded as burn-in, and the remaining trees were used to construct majority-rule consensus trees. The average standard deviation of split frequencies between the two runs was 0.004, and ESS values as computed by Tracer v.1.3 (Rambaut & Drummond, 2004) were above 600 for all individual MCMC runs. Following Alfaro & al. (2003), we considered posterior probabilities (PP) greater or equal to 0.95 as significant probability for a clade. For maximum likelihood analyses, default parameters were used for the Garli searches except that "significant topochange" was set to 0.01 and a total of 100 ML bootstrap replicates (MLBS) were performed. The trees obtained from Garli were used to construct 50% majority-rule consensus trees using PAUP* v.4.0b10 (Swofford, 2003).

In order to better understand phylogenetic relationships in the *Parasponia*-*Trema* complex and *Celtis*, we also expanded our sampling by including sequences of more taxa and two more plastid loci (*matK*, *ndhF*) from GenBank (Appendix 2). These sequences were combined with our own sequences of four plastid gene loci, and unavailable sequences were treated as missing data. The phylogeny was reconstructed from this combined data matrix applying the same methods as described above.

Approximately unbiased test. — The approximately unbiased (AU) test (Shimodaira, 2002), as implemented in CONSEL v.0.1i (Shimodaira & Hasegawa, 2001) with default scaling and replicate values (1000 bootstrap replicates), was used to estimate relative support for all possible topologies among *Lozanella*, *Gironniera* and clade E.

Ancestral character state reconstructions. — The aim of reconstructing ancestral character states was to evaluate the evolution of morphological characters and identify potential synapomorphies for clades. We chose eight morphological characters that have been widely used in the classification of genera in this family (Table 2). These morphological data were obtained from specimen observations and the literature (Killip & Morton, 1931; Zavada & Dilcher, 1986; Takahashi, 1989; Takaso & Tobe, 1990; Todzia, 1993; Zavada & Kim, 1996;

Table 2. Morphological characters and character states (for data matrix see Appendix 3).

Morphological characters	States
Sexual system	0, monoecious; 1, dioecious; 2, andromonoecious; 3, monoecious or dioecious; 4, polygamous
Leaf arrangement	0, opposite; 1, alternate; 2, alternate and opposite
Pollen aperture number	0, triporate; 1, diporate; 2, pentaporate
Aestivation	0, valvate; 1, imbricate
Fruit type	0, drupe; 1, achene; 2, samara
Seed coat morphology	0, with holes; 1, without holes
Perianth at fruiting time	0, deciduous; 1, persistent
Stipule arrangement	0, intrapetiolar; 1, extrapetiolar; 2, interpetiolar

Sytsma & al., 2002; Fu & al., 2003; Zhou & Bartholomew, 2003; Sattarian & Maesen, 2006; Sattarian & al., 2006), and all characters were discrete and coded as binary or multistate (Table 2; Appendix 3). Ancestral state reconstructions were performed in Mesquite v.2.74 (Maddison & Maddison, 2010) using parsimony for tracing character evolution.

The Bayesian 50% majority-rule consensus tree of the combined chloroplast regions was used for ancestral state analysis, including outgroup species. In order to account for uncertainty in phylogeny inference, ancestral state reconstructions were also performed using the Bayesian stochastic mapping (Huelsenbeck & al., 2003) approach as implemented in SIMMAP v.1.5 (Bollback, 2006). Posterior probabilities for ancestral character states were calculated using 15,000 trees sampled from the trees (excluding burn-in) obtained in the Bayesian analysis (see above). Morphology priors were calculated using a two-step approach (http://www.simmap.com/pgs/priors.html). In the first step, we performed an MCMC analysis to sample the overall rate values (Gamma prior) and bias values (Beta prior for two-state characters). In the second step, we used the samples from the posterior distribution of these parameters from the first step and selected the best fitting Gamma and Beta distribution in R v.2.14.1 (R Development Core Team, 2011).

RESULTS

Phylogenetic analyses of individual chloroplast genes. — Sequences of *rbcL* and *trnL-trnF* were obtained for all sampled accessions. However, sequences of *rps16* for *Chaetachme aristata* 3 and *Gironniera subaequalis* were unavailable. We failed to amplify *atpB-rbcL* of *Aphananthe monoica*, *A. philippinensis*, *Chaetachme aristata* 1, *C. aristata* 2, *C. aristata* 3, *Humulus scandens*, *Parasponia melastomatifolia*, and *Trema tomentosa* 1, although multiple attempts were made.

The four chloroplast datasets showed different levels of sequence variation. Characteristics of each gene and the combined dataset are presented in Table 3. The MP analysis for *rbcL* resulted in five equally parsimonious trees of 425 steps: this gene provided the lowest percentage of informative sites of the four markers examined (12.0%). The MP analysis for *atpB-rbcL* resulted in two equally parsimonious trees of 416 steps (15.1% of characters potentially parsimony-informative). The MP analysis of the *trnL-trnF* dataset resulted in three equally parsimonious trees of 516 steps (17.5% of characters potentially parsimony-informative). The MP analysis for *rps16* resulted in eighteen equally parsimonious trees of 645 steps: this gene provided the highest percentage of informative sites of all four markers examined (20.2%). The ML and MP analyses of each locus yielded topologies similar to the BI phylogeny and differed only in weakly supported clades (Electr. Suppl.: Fig. S1).

Phylogenetic analyses of combined chloroplast genes. — The chloroplast genome typically behaves as a single, nonrecombining region, and there was no significant conflict in well-supported clades among the different gene trees. We thus combined all four chloroplast loci in our analysis. The combined dataset included 4538 unambiguously aligned positions with 731 parsimony-informative characters and gaps treated as missing. Parsimony analysis of the combined data yielded one most parsimonious tree of 2048 steps (consistency index, CI = 0.746; retention index, RI = 0.842). The coded indels of the combined data provided another 82 characters, of which 77 (93.9%) were parsimony-informative (PI indels). When the PI indels were excluded from the analyses, the same relationships were observed, but some clades received lower support.

ML and BI analyses of the combined dataset yielded topologies similar to the MP phylogeny (Fig. 1) and differed only in weakly supported clades ($P > 0.05$). The Bayesian combined chloroplast tree was selected to represent our results.

In the plastid combined analysis (Fig. 1), Cannabaceae was strongly supported as monophyletic (MPBS = 91%; MLBS = 97%; PP = 1.0). The monophyly of each genus of Cannabaceae was also strongly supported except for *Parasponia* nested within *Trema*. *Aphananthe* (MPBS = 100%; MLBS = 100%; PP = 1.0) was sister to the rest of the family. Three strongly supported clades were recovered: the *Lozanella* clade, the *Gironniera* clade and the clade formed by the remaining genera (MPBS = 98%; MLBS = 100%; PP = 1.0). Within the last clade, three clades were resolved, and *Chaetachme* and *Pteroceltis* were strongly supported as sister genera (MPBS = 100%; MLBS = 100%; PP = 1.0). Likewise, *Cannabis* and *Humulus* were also strongly supported as sister genera (MPBS = 100%; MLBS = 100%; PP = 1.0).

The phylogeny from the expanded data matrix using additonal GenBank sequences (Electr. Suppl.: Fig. S2) is largely

Table 3. Characteristics of individual and combined datasets.

			Variable sites		Parsimony-informative sites				
Dataset	Taxa	Aligned length	Within ingroup	Entire dataset	Within ingroup	Entire dataset	Consistency index	Retention index	Model selected by AIC
rbcL	42	1227	221	170	147	112	0.616	0.763	TIM+T+G
atpB-rbcL	34	1074	300	161	162	119	0.858	0.906	TVM+G
trnL-trnF	42	1117	329	174	196	158	0.789	0.887	GTR+G
rps16	40	1120	392	241	226	182	0.777	0.86	GTR+G
Combined	42	4538	1242	746	731	571	0.746	0.842	GTR+I+G

Version of Record (identical to print version).

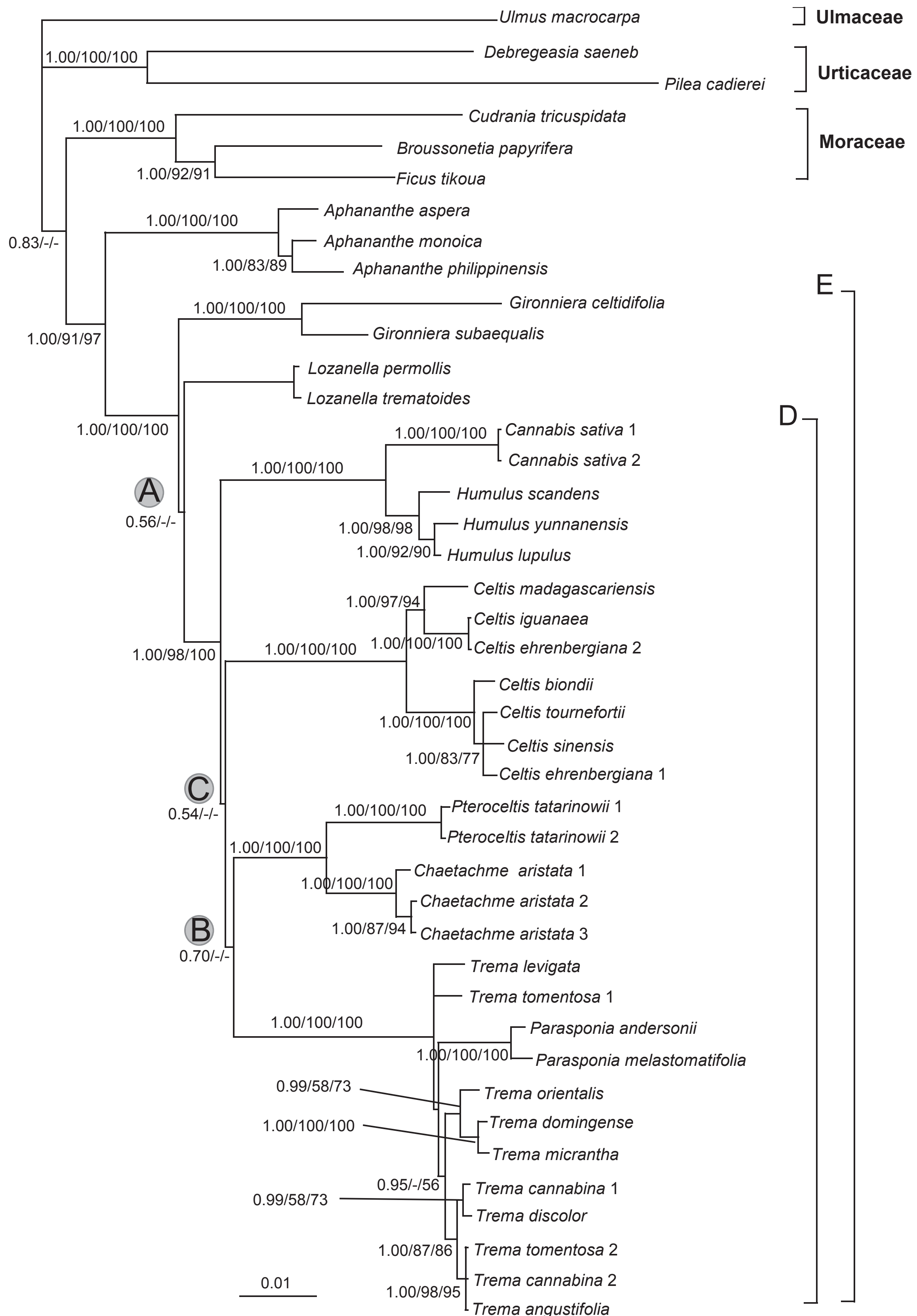


Fig. 1. Bayesian tree based on the combined data of four chloroplast genes. A, B and C represent three unresolved nodes. Clades D and E are discussed in the text. PP values from the BI analysis, and bootstrap values (%) of the MP and ML analyses are shown (PP/MPBS/MLBS). Dashes indicate that branches are supported by less than 50% PP, MPBS, or MLBS.

congruent with that based on the four loci sequenced by us. Exceptions were: (1) *Celtis* was not monophyletic in the BI and ML analyses, which suggested that *C. schippii* is sister to a clade formed by all genera except *Aphananthe*, *Gironniera* and *Lozanella*. However, this relationship only obtained weak support. *Celtis* was supported as monophyletic in the MP analysis without BS support; (2) the relationships among *Celtis*, the *Cannabis*-*Humulus* clade, the *Chaetachme*-*Pteroceltis* clade, and the *Parasponia*-*Trema* clade are slightly different between the two phylogenies, but relationships among these clades were not resolved in both analyses (Fig. 1; Electr. Suppl.: Fig. S2). The expanded data did not improve the resolution of the phylogeny of Cannabaceae, and some apparently spurious results may have been caused by missing data. We thus only included this phylogeny as a supporting figure and use it only to address relationships in the *Parasponia*-*Trema* clade.

Ancestral character state reconstructions. — The ancestral condition of the sexual system of the family is equivocal in the parsimony analysis, and the Bayesian analysis indicated monoecy being the ancestral state with a slightly higher probability (69.8%; Fig. 2A). Four shifts to monoecy and dioecy occurred in *Gironniera*, the *Cannabis*-*Humulus* clade, *Chaetachme* and *Trema angustifolia*, two shifts to andromonoecy in *Celtis*, and one shift to dioecy in *Lozanella*. Polygamy independently evolved at least five times in *Trema* and *Celtis*.

Alternate leaves represent the ancestral state for Cannabaceae in both the parsimony and Bayesian analyses with high probobility (99.6%; Fig. 2B). One shift to opposite leaves occurred in *Lozanella*, and one shift to opposite and alternate leaves in the *Cannabis* and *Humulus* clade.

Triporate pollen is inferred to be the ancestral state for Cannabaceae in both the parsimony and Bayesian analyses

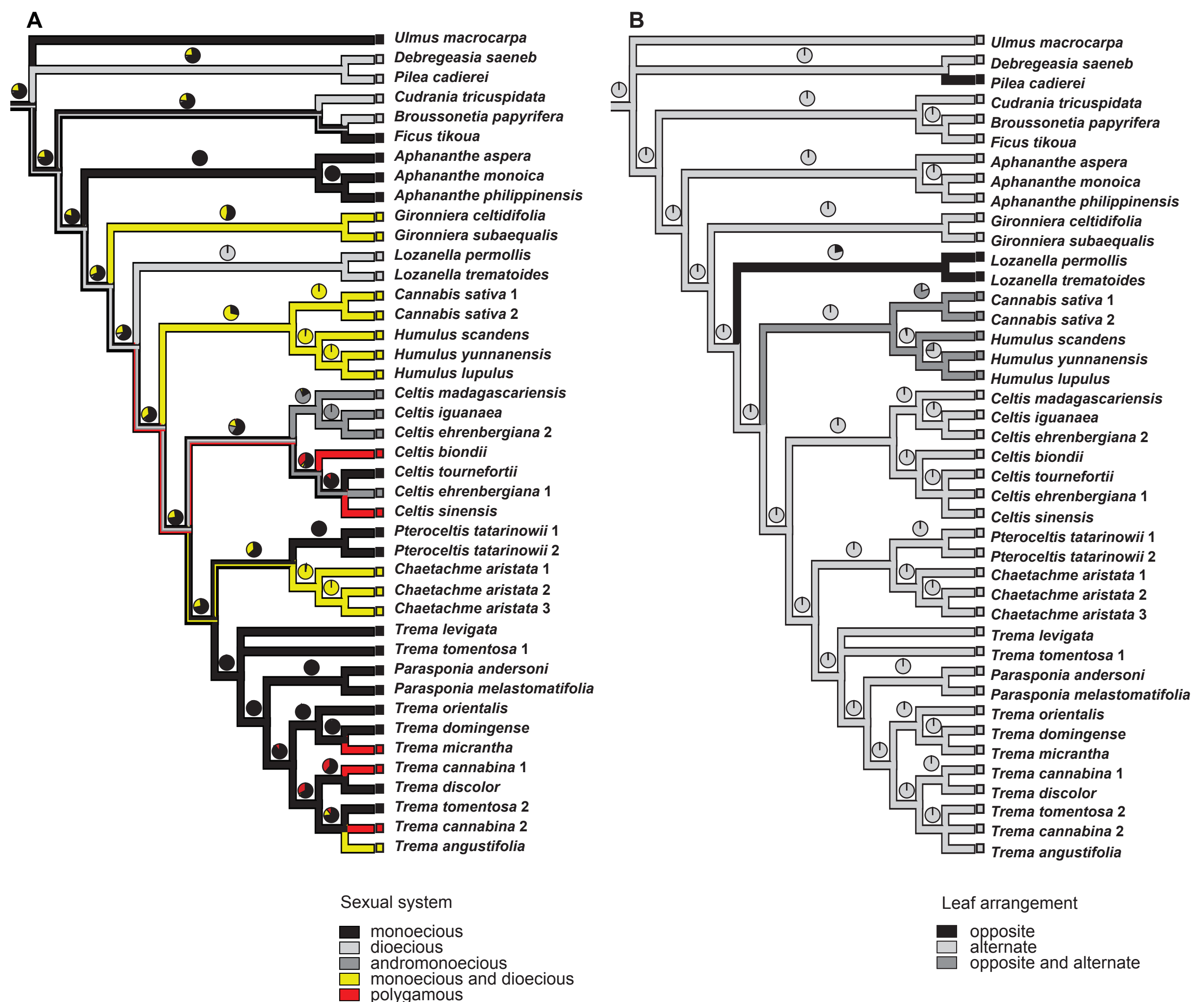


Fig. 2A–H. Reconstruction of ancestral states of morphological traits using parsimony and Bayesian approaches. Posterior probabilities for each character state are indicated as pie charts.

Version of Record (identical to print version).

with a relatively high probability (85.8%; Fig. 2C). One shift to diporate pollen occurred in the *Trema-Parasponia* clade.

The ancestral state of aestivation for Cannabaceae is equivocal in the parsimony analysis (Fig. 2D). The Bayesian analysis inferred imbricate aestivation as the ancestral condition for the family with a slightly higher probability (66.2%; Fig. 2D). Shifts to valvate aestivation occurred at least two times in *Aphananthe* and the *Chaetachme-Pteroceltis-Trema-Parasponia* clade. Within the latter clade, two reversals to imbricate aestivation took place in *Pteroceltis* and *Parasponia*.

Drupes are the ancestral state for Cannabaceae in both the parsimony and Bayesian analyses with high probability (96.2%; Fig. 2E). One shift to samaras occurred in *Pteroceltis*, and one shift to achenes in the *Cannabis-Humulus* clade.

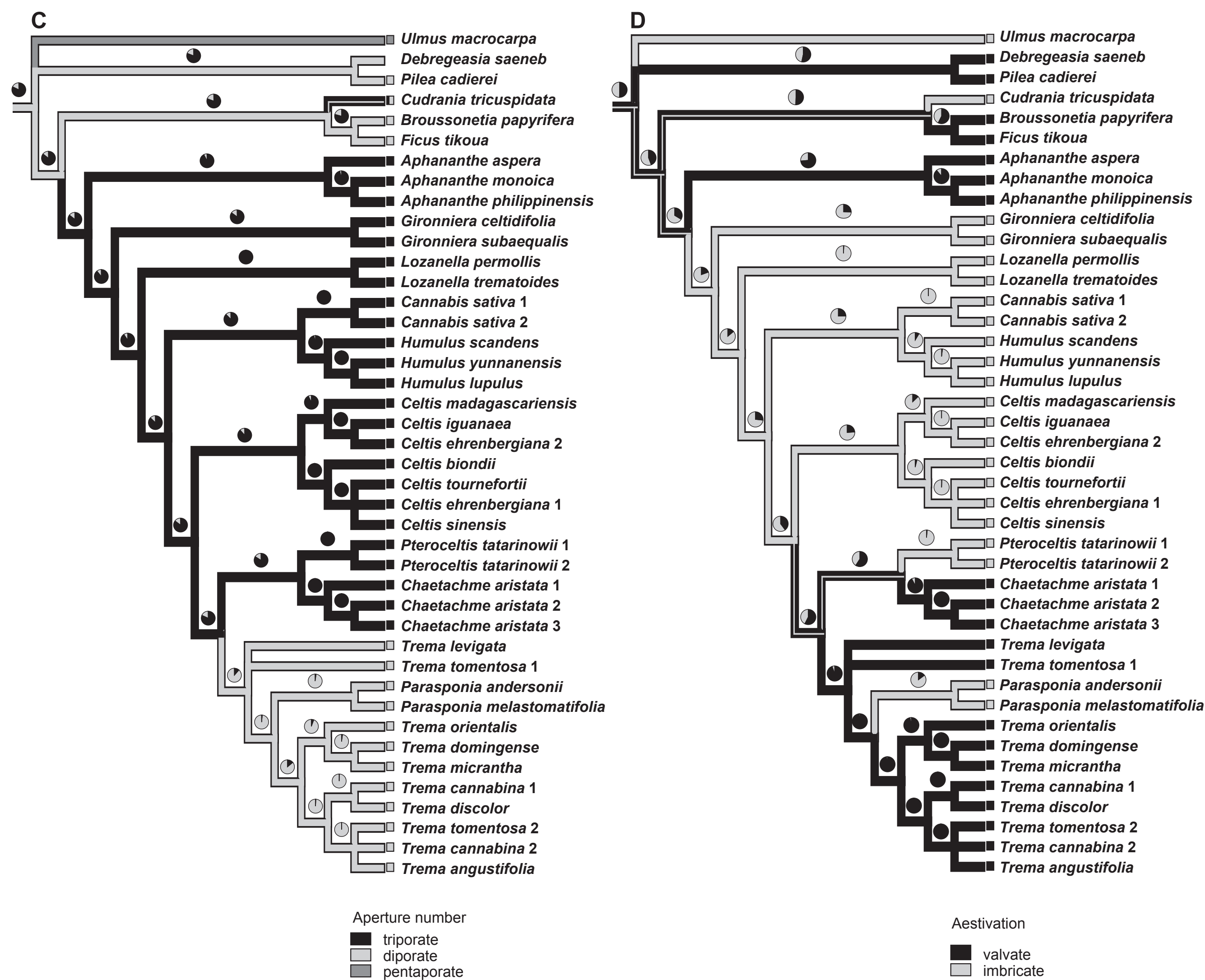
Seed coat without holes is reconstructed as the ancestral state by both the parsimony and Bayesian analyses (90.1%; Fig. 2F). One shift to seed coat with holes occurred in clade D (as defined in Fig. 1), and two reversals took place in *Humulus* and the *Trema-Parasponia* clade.

Persistent perianth is the ancestral condition for the family in both the parsimony and Bayesian analyses with slightly higher probability (63.1%) (Fig. 2G). Four shifts to deciduous perianth occurred in *Celtis*, *Chaetachme*, *Lozanella* and *Trema levigata*.

Extrapetiolar stipules is the ancestral state for Cannabaceae supported by both the parsimony and Bayesian analyses with high probability (97.2%; Fig. 2H). Three shifts to intrapetiolar stipules occurred in *Chaetachme*, *Lozanella* and *Parasponia*, and one shift to interpetiolar stipules occurred in *Humulus*.

DISCUSSION

The monophyly of Cannabaceae. — Including all recognized genera and using four plastid loci, our molecular data strongly supported the monophyly of Cannabaceae (MPBS = 91%; MLBS = 97%; PP = 1.0). Members of this family share a 9-bp deletion at position 247 in the *trnL-trnF* intergenic spacer.

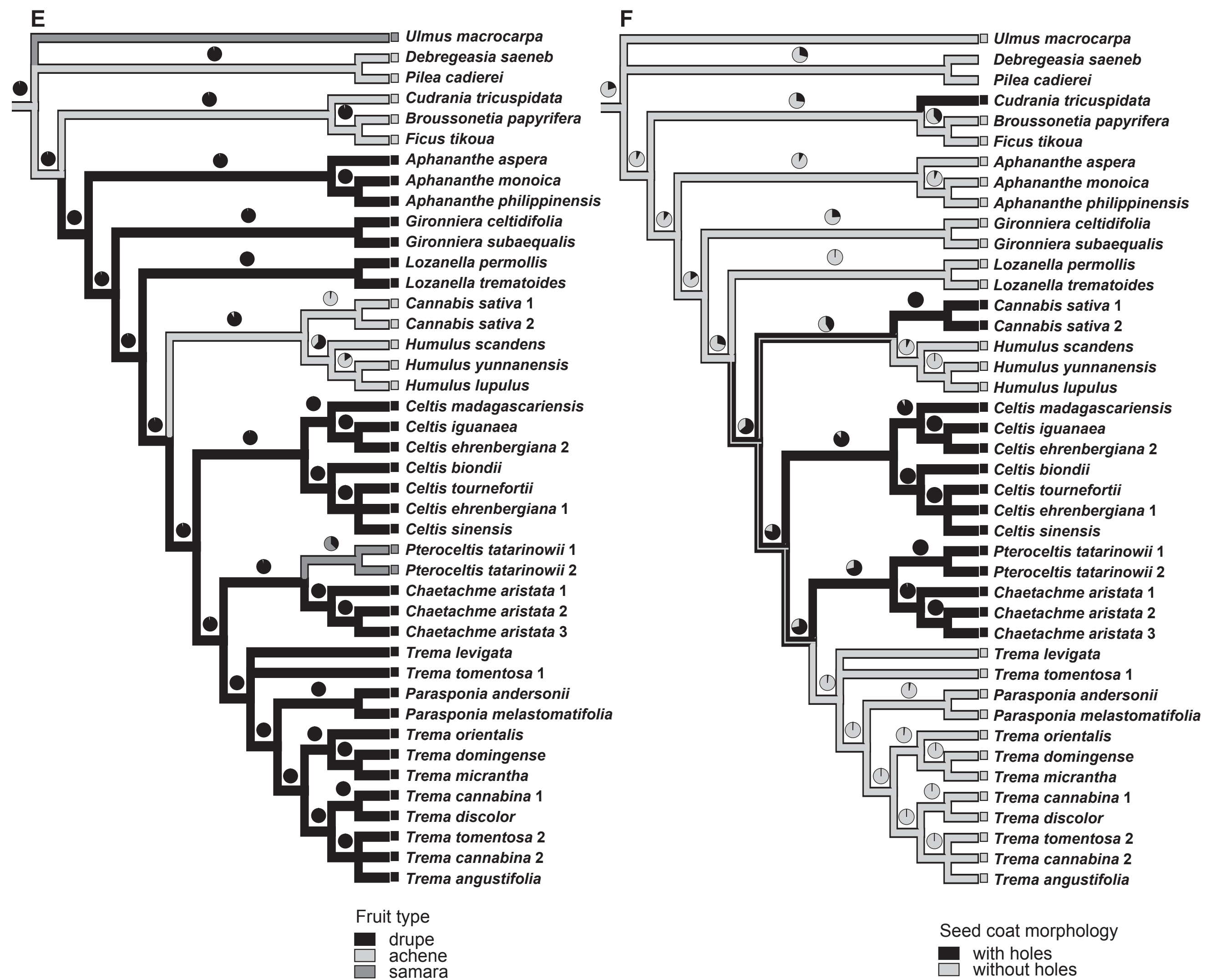


The recently expanded Cannabaceae can be identified by the following morphological characters: presence of cystoliths, absence of laticifers, antitepalous stamens, triporate or diporate pollen grains, curved embryo, two carpels, and superior ovary with apical placentation. However, most of these characters are shared with Moraceae and/or Urticaceae. More detailed studies should be carried out to identify synapomorphies of the family.

Intergeneric relationships among Cannabaceae. — Phylogenetic relationships within Cannabaceae were not completely resolved in previous studies (Song & al., 2001; Sytsma & al., 2002; Van Velzen & al., 2006). However, including all recognized genera, we were able to greatly improve phylogenetic resolution. Most nodes except nodes A, B and C obtained strong support (Fig. 1). Whether these nodes reflect rapid evolution remains to be tested. Below we discuss phylogenetic relationships in the family based on our results.

The *Aphananthe* clade. — Our results strongly supported *Aphananthe* (MPBS = 100%, MLBS = 100%, PP = 1.0) as sister to all other genera of Cannabaceae, in agreement with most previous molecular studies (Song & al., 2001; Sytsma & al., 2002; Van Velzen & al., 2006). *Aphananthe* differs from the other genera by its base chromosome number of $x = 13$ (Oginuma & al., 1990), asymmetrical ovules (Takaso, 1987), the presence of flavonols (Giannasi, 1978), and a unique seed coat morphology (Takaso & Tobe, 1990), supporting the finding that this genus has an isolated position in the family. Our molecular study strongly suggested the genus to be a member of Cannabaceae, with which it shares pollen structure (Kuprianova, 1962; Takahashi, 1989), leaf vernation pattern (Terabayashi, 1991), and gynoecial vasculature (Omori & Terabayashi, 1993).

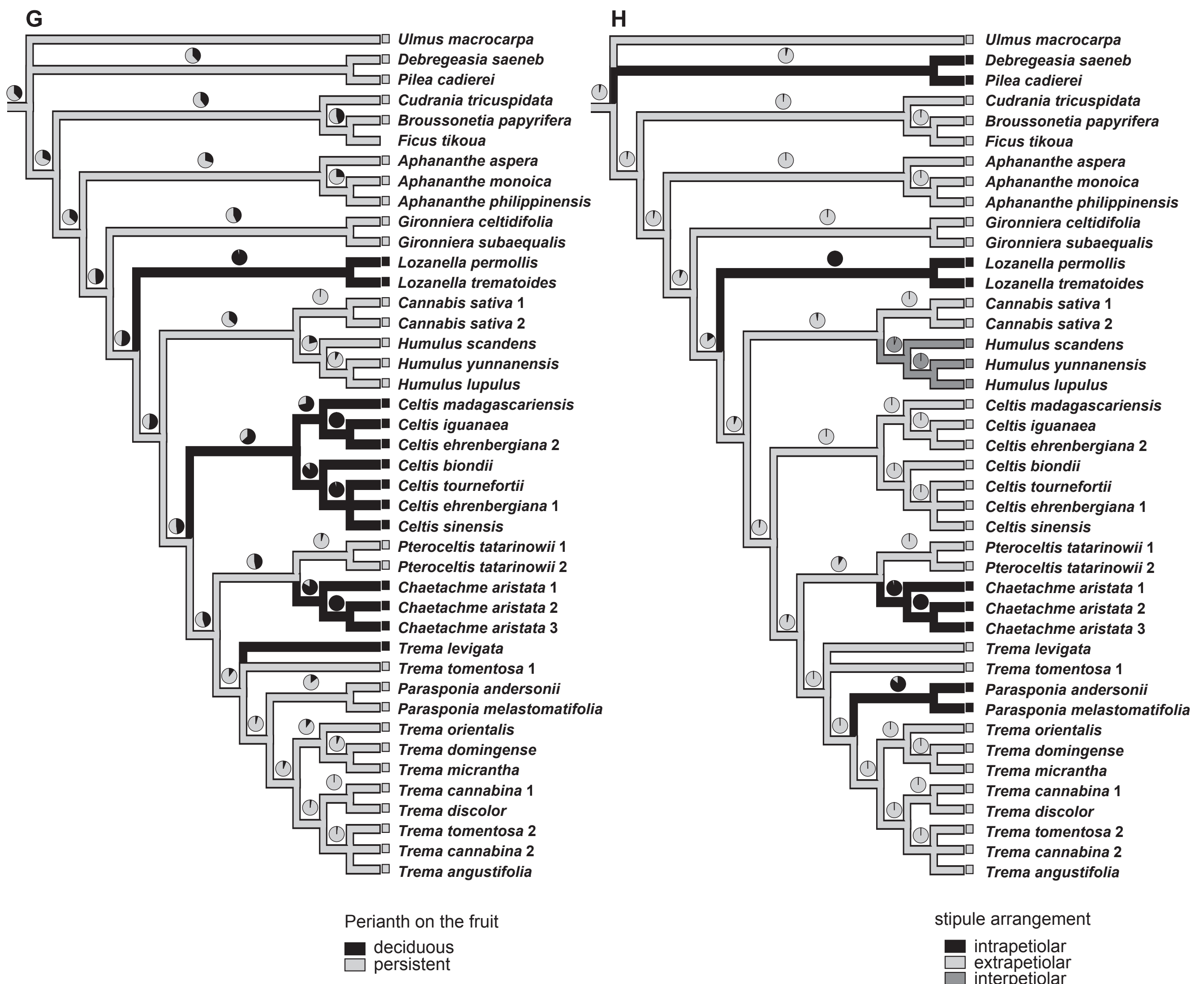
Phylogenetic relationships in clade E. — The *Gironniera* clade, the *Lozanella* clade and clade D together formed a strongly supported monophyletic group (MPBS = 100%, MLBS = 100%, PP = 1.0), marked as clade E in Fig. 1. Most members of this group contain glycoflavones (Giannasi, 1978). Exceptions to this are *Cannabis*, which produces flavonols as well as glycoflavones (Clark, 1978), and *Humulus* which contains flavonols (Alaniya & al., 2010). In addition, all members of clade E share two deletions in *trnL-trnF* (one 100-bp deletion at position 125–224; one 8-bp deletion at position 270–277). The phylogenetic relationships among *Gironniera*, *Lozanella* and clade D within this clade were not resolved. Our Bayesian



inference indicated that *Gironniera* is sister to a branch comprising *Lozanella* and clade D, but this conclusion is inconsistent with the result of the AU test, which suggested that *Lozanella* is sister to a branch comprising *Gironniera* and clade D.

Phylogenetic relationships in clade D. — The monophyly of Clade D received strong support (MPBS = 100%, MLBS = 98%, PP = 1.0). Most members of this clade have a base chromosome number of $x = 10$, but *Humulus* has $x = 8$ and *Chaetachme* $x = 15$ (Oginuma & al., 1990; Sytsma & al., 2002). Four strongly supported subclades (*Celtis*, *Cannabis*-*Humulus*, *Chaetachme*-*Pteroceltis*, *Trema*-*Parasponia*) were reconstructed in accordance with results of Van Velzen & al. (2006), but relationships among them were largely unresolved. The phylogenetic position of *Cannabis* and *Humulus* has been controversial for a long time with relationships suggested to *Celtis* (Chase & al., 1993), to *Pteroceltis*-*Celtis* (Wiegrefe & al., 1998), to *Pteroceltis*-*Celtis*-*Trema* (Song & al., 2001) or to *Pteroceltis* (Sytsma & al., 2002). The limitations of these studies were that they did not sample all genera and did not discuss morphological evidence to support these relationships. *Cannabis*, *Humulus* and *Pteroceltis* share a distinctive S-type sieve-element plastid (Behnke, 1989), but this character is homoplastic because these three genera did not form a clade. Including all recognized genera, our phylogenetic analysis indicated that the *Cannabis*-*Humulus* clade is sister to all other genera of clade D with low support. Additionally, the sister relationship between *Celtis* and the *Pteroceltis*-*Chaetachme*-*Trema*-*Parasponia* clade was implied with low statistical support in our analysis.

The *Chaetachme*-*Pteroceltis* subclade. — The monotypic *Chaetachme* is endemic to tropical Africa, southern Africa and Madagascar, and the monotypic *Pteroceltis* is one of the so-called Tertiary relict trees (Chai & al., 2010) restricted to temperate habitats in several parts of China (Li & al., 2012). These two genera have previously been placed in Celtidaceae (Link, 1829; Grudzinskaya, 1967), but their phylogenetic positions were not resolved in previous studies. A phylogenetic study employing *rbcL* indicated that *Chaetachme* and *Pteroceltis* formed a clade with low support (Ueda & al., 1997), but the two genera were not sister in the study of Sytsma & al. (2002). Our results show strong support for a sister relationship



between *Chaetachme* and *Pteroceltis* (MPBS = 100%, MLBS = 100%, PP = 1.0), which is consistent with Van Velzen & al. (2006). Synapomorphies for the sister-group relationship include a 5-bp insertion at position 646–650 of *rps16*. However, *Pteroceltis* and *Chaetachme* show some differences in both morphological characters and chromosome number. *Pteroceltis* has serrate leaf blades, extrapetiolar stipules, leaves 3-veined from base, unarmed twigs and samaras, whereas *Chaetachme* has entire leaf blades, interpetiolar stipules, pinnate leaf venation, spiny twigs and drupes. Furthermore, the chromosome number of *Pteroceltis* is $x = 10$, while that of *Chaetachme* is $x = 15$. Synapomorphies for this clade are unknown. The geographical distribution of the two genera is non-overlapping, but *Pteroceltis* had a wider range during the Tertiary as indicated by fossil occurrences in the Oligocence of Germany (Weyland, 1937) and in the Middle Eocene of Tennessee (Manchester & al., 2009). This may imply that these two genera may have been geographically closer to each other than they are today.

The *Parasponia-Trema* complex. — *Trema* is a pantropically distributed pioneer plant (Yesson & al., 2004), and *Parasponia* is distributed in Southeast Asia and the Pacific Islands and is the only genus of the family that is known to fix nitrogen (Becking, 1983). *Parasponia* differs from *Trema* in having connate intrapetiolar stipules and imbricate perianth lobes in male flowers (Soepadmo, 1977), and the ability to fix nitrogen (Becking, 1983; Sturms & al., 2010). However, our molecular results strongly suggested that *Parasponia* is nested within *Trema* (Fig. 1). These results are consistent with previous phylogenetic analyses (Zavada & Kim, 1996; Sytsma & al., 2002; Yesson & al., 2004; Van Velzen & al., 2006). *Parasponia* and *Trema* shared three *atpB-rbcL* insertions (7-bp insertion at position 386–392, 5-bp insertion at position 524–528, 7-bp insertion at position 942–948), one *trnL-trnF* insertion (6-bp insertion at position 30–35) and one *rps16* insertion (4-bp insertion at position 44–47). The expanded dataset also strongly supported that *Parasponia* is nested within *Trema* (MLBS = 100%, MPBS = 98%, PP = 1.0), and five *Parasponia* species formed a weakly supported clade (Electr. Suppl.: Fig S2). Furthermore, we found that diporate pollen is a good synapomorphy for this clade (Fig. 2C). Additionally, both genera have a basically lineate seed coat where the exotestal cells are characteristically longitudinally elongate (Takaso & Tobe, 1990), and pollen which lacks a granular layer in the exine (Takahashi, 1989). However, these two characters are shared with *Lozanella*. Based on this evidence, we suggested that *Parasponia* should be merged with *Trema*. The taxonomy of *Trema* is disputed because of the difficulty in finding good morphological features to delimit species (Yesson & al., 2004). Important diagnostic characters including leaf texture, venation, size, shape, colour and pubescence are variable and probably subject to ecotypic and ontogenetic variation. This may be the main reason why multiple individuals from one species did not form one clade in a previous molecular study (Yesson & al., 2004) and our study (see *Trema cannabina* and *Trema tomentosa* in Fig. 1).

The *Cannabis-Humulus* subclade. — *Cannabis* and *Humulus* were strongly supported as sister genera (MLBS = 100%, MPBS = 100%, PP = 1.0), in agreement with some previous molecular studies that sampled only one specimen from each genus (Song & al., 2001; Sytsma & al., 2002; Song & Li, 2002). Many morphological characters support this clade, including their herbaceous habit, palmately lobed or compound leaves, achenes, and female inflorescences which are bracteate spicate cymes.

Morphological character evolution in Cannabaceae. — Most morphological characters analyzed showed a complex evolution pattern and changed more than twice. For example, at least twelve shifts occurred among the five states of the sexual system. Both monoecy and dioecy occur frequently in the same family, but rarely in the same genus (Baker, 1959). Two or more sexual system states are present within several genera of Cannabaceae, including *Cannabis*, *Celtis*, *Chaetachme*, *Gironniera*, *Humulus* and *Trema*. Monoecy and dioecy even co-occur in the same species of *Cannabis*, *Chaetachme*, *Gironniera*, *Humulus* and *Trema*. Further studies should be carried out to reveal the underlying ecological and evolutionary mechanisms of frequent mating system transitions in Cannabaceae. Other characters such as seed coat with holes, valvate aestivation, deciduous perianth and intrapetiolar stipules also have changed several times in the family.

The expanded Cannabaceae include most genera of Ulmaceae or Celtidaceae, which makes the family morphologically highly diverse. Our character evolution analyses showed that some morphological characters traditionally used to delimit generic boundaries can not be used for this purpose. For example, unisexual or polygamous flowers, combined with other morphological characters, have been used to separate *Trema* and *Celtis* (Fu & al., 2003). In contrast to this, our analyses indicated that both genera have complex sexual system, and each state is homoplastic.

It is difficult to find morphological synapomorphies for most clades of Cannabaceae. For example, all morphological characters shared by the *Chaetachme-Pteroceltis* clade are homoplastic. However, some morphological characters such as diporate pollen, achenes or opposite and alternate leaves have evolved only once in the family and can be regarded as synapomorphies of certain clade: diporate pollen for the *Parasponia-Trema* clade, achenes and opposite or alternate leaves for the *Cannabis-Humulus* clade.

■ ACKNOWLEDGEMENTS

We thank Ulrike Bertram (the Ecological-Botanical Garden, Bayreuth) and Shudong Zhang (the Kunming Institute of Botany, Chinese Academy of Sciences, Kunming) for their help in obtaining material, the Botanic Garden Conservation International (BGCI) institution for offering samples, and Andrew Hipp (the Morton Arboretum, Chicago) for revising the manuscript. The DNA analysis was undertaken at the Molecular Biology Experiment Center, Kunming Institute of Botany. This study was supported by the National Natural Science Foundation of China (grant no. 40830209 and no. 31270274), the Knowledge Innovation Program of the Chinese Academy of Sciences (grant no. KSCX2-YW-R-136) and the Talent Project of Yunnan Province (project no. 2011CI042).

LITERATURE CITED

Alaniya, M., Kavtaradze, N.S., Skhirtladze, A., Pizza, C. & Piacente, S. 2010. Flavonol glycoside from *Humulus lupulus*. *Chem. Nat. Compounds* 46: 641–642. http://dx.doi.org/10.1007/s10600-010-9699-x

Alfaro, M.E., Zoller, S. & Lutzoni, F. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molec. Biol. Evol.* 20: 255–266. http://dx.doi.org/10.1093/molbev/msg028

Baker, H. 1959. Reproductive methods as factors in speciation in flowering plants. *Cold Spring Harbor Symp. Quant. Biol.* 24: 177–191. http://dx.doi.org/10.1101/SQB.1959.024.01.019

Becking, J. 1983. The *Parasponia parviflora*-Rhizobium symbiosis: Host specificity, growth and nitrogen fixation under various conditions. *Pl. & Soil* 75: 309–342. http://dx.doi.org/10.1007/BF02369969

Behnke, H.D. 1989. Sieve-element plastids, phloem proteins, and the evolution of flowering plants. IV. Hamamelidae. Pp. 105–128 in: Crane, P.R. & Blackmore, S. (eds.), *Evolution, systematics, and fossil history of the Hamamelidae*, vol. 2, *Lower Hamamelidae*. Oxford: Clarendon Press.

Bollback, J. 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *B. M. C. Bioinf.* 7: 88. http://dx.doi.org/10.1186/1471-2105-7-88

Brown, J.M. & Lemmon, A.R. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst. Biol.* 56: 643–655. http://dx.doi.org/10.1080/10635150701546249

Chai, X.Y., Chen, S.L. & Xu, W. 2010. Using inter-simple sequence repeat markers to analyze the genetic structure of natural *Pteroceltis tatarinowii* populations and implications for species conservation. *Pl. Syst. Evol.* 285: 65–73. http://dx.doi.org/10.1007/s00606-009-0256-7

Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., Duvall, M.R., Price, R.A., Hills, H.G., Qiu, Y.L., Kron, K.A., Rettig, J.H., Conti, E., Palmer, J.D., Manhart, J.R., Sytsma, K.J., Michaels, H.J., Kress, W.J., Karol, K.G., Clark, W.D., Hedren, M., Gaut, B.S., Jansen, R.K., Kim, K.J., Wimpee, C.F., Smith, J.F., Furnier, G.R., Strauss, S.H., Xiang, Q.Y., Plunkett, G.M., Soltis, P.S., Swensen, S.M., Williams, S.E., Gadek, P.A., Quinn, C.J., Eguiarte, L.E., Golenberg, E., Learn, G.H., Graham, S.W., Barrett, S.C.H., Dayanandan, S. & Albert, V.A. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Gard.* 80: 528–580. http://dx.doi.org/10.2307/2399846

Chiang, T.Y., Schaal, B.A. & Peng, C.I. 1998. Universal primers for amplification and sequencing a noncoding spacer between the *atpB* and *rbcL* genes of chloroplast DNA. *Bot. Bull. Acad. Sin.* 39: 245–250.

Clark, M.N. 1978. *A study of infraspecific flavonoid variation of* Cannabis sativa *L. (Cannabaceae)*. Dissertation, University of British Columbia, Vancouver, Canada.

Doyle, J. & Doyle, J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull. Bot. Soc. Amer.* 19: 11–15.

Engler, A. & Prantl, K. 1893. *Die natürlichen Pflanzenfamilien*, vol. 3. Berlin: Bornträger.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783–791. http://dx.doi.org/10.2307/2408678

Fu, L., Xin, Y. & Whittemore, A. 2003. Ulmaceae. Pp. 1–19 in: Wu, Z.Y., Raven, P.H. & Hong, D.Y. (eds.), *Flora of China*, vol. 5. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press.

Giannasi, D.E. 1978. Generic relationships in Ulmaceae based on flavonoid chemistry. *Taxon* 27: 331–344. http://dx.doi.org/10.2307/1220369

Grudzinskaya, I.A. 1967. Ulmaceae and reasons for distinguishing Celtidoideae as a separate family Celtidaceae Link. *Bot. Zhurn. S.S.S.R.* 52: 172–349.

Huelsenbeck, J.P., Nielsen, R. & Bollback, J.P. 2003. Stochastic mapping of morphological characters. *Syst. Biol.* 52: 131–158. http://dx.doi.org/10.1080/10635150390192780

Judd, W.S., Campbell, C.S., Kellogg, E.A., Stevens, P.F. & Donoghue, M.J. 2008. *Plant systematics: A phylogenetic approach*, 3rd ed. Sunderland, Massachusetts: Sinauer.

Killip, E.P. & Morton, C.V. 1931. The genus *Lozanella*. *J. Washington Acad. Sci.* 21: 336–339.

Kuprianova, A. 1962. Palynological data and the systematics of the Fagales and Urticales. Pp. 17–25 in: Sladkov, A.N. (ed.), *Reports of Soviet Palinologists for the First International Conference on Palinology (Tucson, USA).* Moskow: Publishing House of USSR Academy of Sciences..

Li, X.H., Shao, J.W., Lu, C., Zhang, X.P. & Qiu, Y.X. 2012. Chloroplast phylogeography of a temperate tree *Pteroceltis tatarinowii* (Ulmaceae) in China. *J. Syst. Evol.* 50: 325–333. http://dx.doi.org/10.1111/j.1759-6831.2012.00203.x

Link, H.F. 1829. *Handbuch zur Erkennung der nutzbarsten und am häufigsten vorkommenden Gewächse*, vol. 2, Berlin: in der Haude und Spenerschen Buchhandlung.

Mabberley, D.J. 2008. *Mabberley's plant-book: A portable dictionary of plants, their classification and uses*. New York: Cambridge University Press.

Maddison, W.P. & Maddison, D.R. 2010. Mesquite: A modular system for evolutionary analysis, version 2.74. University of British Columbia & Oregon State University. http://mesquiteproject.org

Manchester, S.R., Chen, Z.D., Lu, A.M. & Uemura, K. 2009. Eastern Asian endemic seed plant genera and their paleogeographic history throughout the Northern Hemisphere. *J. Syst. Evol.* 47: 1–42. http://dx.doi.org/10.1111/j.1759-6831.2009.00001.x

Oginuma, K., Raven, P.H. & Tobe, H. 1990. Karyomorphology and relationships of Celtidaceae and Ulmaceae (Urticales). *Bot. Mag. (Tokyo)* 103: 113–131. http://dx.doi.org/10.1007/BF02489620

Omori, Y. & Terabayashi, S. 1993. Gynoecial vascular anatomy and its systematic implications in Celtidaceae and Ulmaceae (Urticales). *J. Pl. Res.* 106: 249–258. http://dx.doi.org/10.1007/BF02344593

Oxelman, B., Liden, M. & Berglund, D. 1997. Chloroplast *rps16* intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Pl. Syst. Evol.* 206: 393–410. http://dx.doi.org/10.1007/BF00987959

Posada, D. & Crandall, K.A. 1998. MODELTEST: Testing the model of DNA substitution. *B. M. C. Bioinf.* 14: 817–818. http://dx.doi.org/10.1093/bioinformatics/14.9.817

R Development Core Team 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

Rambaut, A. 2002. Sequence alignment editor, version 2.0a11. Oxford: University of Oxford. http://evolve.zoo.ac.uk

Rambaut, A. & Drummond, A. 2004. Tracer, version 1.3. Edinburgh: University of Edinburgh.

Rendle, A.B. 1925. *The classification of flowering plants*, vol. 2, London: Cambridge University Press.

Ronquist, F. & Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *B. M. C. Bioinf.* 19: 1572–1574. http://dx.doi.org/10.1093/bioinformatics/btg180

Sattarian, A. & Van der Maesen, L.J.G. 2006. Endocarp morphology of African *Celtis* (Celtidaceae/Ulmaceae). *Blumea* 51: 389–397. http://dx.doi.org/10.3767/000651906X622337

Sattarian, A., Van den Berg, R.G. & Van der Maesen, L.J.G. 2006. Pollen morphology of African *Celtis* (Celtidaceae). *Feddes Repert.* 117: 34–40. http://dx.doi.org/10.1002/fedr.200611087

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* 51: 492–508. http://dx.doi.org/10.1080/10635150290069913

Shimodaira, H. & Hasegawa, M. 2001. CONSEL: For assessing the

confidence of phylogenetic tree selection. *B. M. C. Bioinf.* 17: 1246–1247. http://dx.doi.org/10.1093/bioinformatics/17.12.1246

Simmons, M.P. & Ochoterena, H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49: 369–381. http://dx.doi.org/10.1093/sysbio/49.2.369

Soepadmo, E. 1977. Ulmaceae. Pp. 31–76 in: Van Steenis, C.G.G.J. (ed.), *Flora Malesiana*, ser. 1, vol. 8. Alphen aan den Rijn: Sijthoff & Noordhoff.

Song, B. & Li, F.Z. 2002. The utility of *trnK* intron 5′ region in phylogenetic analysis of Ulmaceae s. l. *Acta Phytotax. Sin.* 40: 125–132.

Song, B.H., Wang, X.Q., Li, F.Z. & Hong, D.Y. 2001. Further evidence for paraphyly of the Celtidaceae from the chloroplast gene *matK*. *Pl. Syst. Evol.* 228: 107–115. http://dx.doi.org/10.1007/s006060170041

Sturms, R., Kakar, S., Trent, J., III & Hargrove, M.S. 2010. *Trema* and *Parasponia* hemoglobins reveal convergent evolution of oxygen transport in plants. *Biochemistry* 49: 4085–4093. http://dx.doi.org/10.1021/bi1002844

Swofford, D.L. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sunderland, Massachusetts: Sinauer.

Sytsma, K.J., Morawetz, J., Pires, J.C., Nepokroeff, M., Conti, E., Zjhra, M., Hall, J.C. & Chase, M.W. 2002. Urticalean rosids: Circumscription, rosid ancestry, and phylogenetics based on *rbcL*, *trnL-trnF*, and *ndhF* sequences. *Amer. J. Bot.* 89: 1531–1546. http://dx.doi.org/10.3732/ajb.89.9.1531

Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Pl. Molec. Biol.* 17: 1105–1109. http://dx.doi.org/10.1007/BF00037152

Takahashi, M. 1989. Pollen morphology of Celtidaceae and Ulmaceae: A reinvestigation. Pp. 253–265 in: Crane, P.R. & Blackmore, S. (eds.), *Evolution, systematics, and fossil history of the Hamamelidae*, vol. 2, *Higher Hamamelidae*. Oxford: Clarendon Press.

Takaso, T. 1987. Ovule ontogeny and morphology in Ulmaceae. Pp. 220 in: Greuter, W., Zimmer, B. & Behnke, H.D. (eds.), *Abstracts, XIV International Botanical Congress*. Königstein: Koeltz Scientific Books.

Takaso, T. & Tobe, H. 1990. Seed coat morphology and evolution in Celtidaceae and Ulmaceae (Urticales). *Bot. Mag. (Tokyo)* 103: 25–41. http://dx.doi.org/10.1007/BF02488408

Terabayashi, S. 1991. Vernation patterns in Celtidaceae and Ulmaceae (Urticales), and their evolutionary and systematic implications. *Bot. Mag. (Tokyo)* 104: 1–13. http://dx.doi.org/10.1007/BF02493399

The Plant List 2010. Version 1. Published on the Internet; http://www.theplantlist.org/ (accessed 20 Mar. 2013).

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* 25: 4876–4882. http://dx.doi.org/10.1093/nar/25.24.4876

Todzia, C.A. 1993. Ulmaceae. Pp. 603–611 in: Kubitzki, K., Rohwer, J.G. & Bittrich, V. (eds.), *The families and genera of vascular plants*, vol. 2. Berlin: Springer.

Ueda, K., Kosuge, K. & Tobe, H. 1997. A molecular phylogeny of Celtidaceae and Ulmaceae (Urticales) based on *rbcL* nucleotide sequences. *J. Pl. Res.* 110: 171–178. http://dx.doi.org/10.1007/BF02509305

Van Velzen, R., Bakker, F.T., Sattarian, A. & Van der Maesen, L.J.G. 2006. Evolutionary relationships of Celtidaceae. Pp. 7–30 in: Sattarian, A. (ed.), *Contribution to the biosystematics of Celtis L. (Celtidaceae) with special emphasis on the African species*. Dissertation, Wageningen University, Wageningen, the Netherlands.

Wang, H.C., Moore, M.J., Soltis, P.S., Bell, C.D., Brockington, S.F., Alexandre, R., Davis, C.C., Latvis, M., Manchester, S.R. & Soltis, D.E. 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl. Acad. Sci. U.S.A.* 106: 3853–3858. http://dx.doi.org/10.1073/pnas.0813376106

Weyland, H. 1937. Beiträge zur Kenntnis der rheinischen Tertiärflora. II. Erste Ergänzungen und Berichtigungen zur Flora der Blätterkohle und des Polierschiefers von Rott im Siebengebirge. *Palaeontographica, Abt. B, Paläophytol.* 83: 67–122.

Wiegrefe, S.J., Sytsma, K.J. & Guries, R.P. 1998. The Ulmaceae, one family or two? Evidence from chloroplast DNA restriction site mapping. *Pl. Syst. Evol.* 210: 249–270. http://dx.doi.org/10.1007/BF00985671

Yesson, C., Russell, S.J., Parrish, T., Dalling, J.W. & Garwood, N.C. 2004. Phylogenetic framework for *Trema* (Celtidaceae). *Pl. Syst. Evol.* 248: 85–109. http://dx.doi.org/10.1007/s00606-004-0186-3

Zavada, M.S. & Dilcher, D.L. 1986. Comparative pollen morphology and its relationship to phylogeny of pollen in the Hamamelidae. *Ann. Missouri Bot. Gard.* 73: 348–381. http://dx.doi.org/10.2307/2399117

Zavada, M.S. & Kim, M. 1996. Phylogenetic analysis of Ulmaceae. *Pl. Syst. Evol.* 200: 13–20. http://dx.doi.org/10.1007/BF00984745

Zhang, S., Soltis, D.E., Yang, Y., Li, D. & Yi, T. 2011. Multi-gene analysis provides a well-supported phylogeny of Rosales. *Molec. Phylogen. Evol.* 60: 21–28. http://dx.doi.org/10.1016/j.ympev.2011.04.008

Zhou, Z. & Bartholomew, B. 2003. Cannabaceae. Pp. 74–75 in: Wu, Z.Y., Raven & P.H., Hong, D.Y. (eds.), *Flora of China*, vol. 5. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press.

Zwickl, D.J. 2006. *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. Dissertation, The University of Texas at Austin, Texas, U.S.A.

Appendix 1. List of taxa studied: Taxon, locality, voucher specimen and GenBank accession numbers for *rbcL*, *atpB-rbcL*, *trnL-trnF* and *rps16*. A dash (–) indicates an unavailable sequence.

OUTGROUP TAXA: ***Broussonetia papyrifera*** Vent. (Moraceae), China, Shaanxi; *Yi & Zhang 080645* (KUN); JF317438; JN040322; JN040358; JN040284. ***Cudrania tricuspidata*** Bureau ex Lavallée (Moraceae), China, Yunnan; *Zhang 090045* (KUN); JF317440; JN040332; JN040371; JN040296. ***Debregeasia saeneb*** (Forssk.) Hepper & J.R.I. Wood (Urticaceae), China, Yunnan; *Yi 09052* (KUN); JF317441; JN040333; JN040372; JN040297. ***Ficus tikoua*** Bureau (Moraceae), China, Hunan; *Yi & Zhang 080142* (KUN); JF317445; JN040334; JN040373; JN040298. ***Pilea cadierei*** Gagnep. & Guillaumin (Urticaceae), China, Yunnan; *Yi 09051* (KUN); JF317451; JN040342; JN040383; JN040307. ***Ulmus macrocarpa*** Hance (Ulmaceae), China, Shandong; *Yi & Zhang 080312* (KUN); JF317455; JN040354; JN040396; JN040320. **INGROUP TAXA:** ***Aphananthe aspera*** (Thunb.) Planch., China, Hubei; *Sun 10001* (KUN); JN040397; JN040321; JN040355; JN040281. ***Aphananthe monoica*** (Hemsl.) J.-F. Leroy, Mexico; *Fernández 3462* (U); JN040398; – ; JN040356; JN040282. ***Aphananthe philippinensis*** Planch., Australia, Queensland; *Forster 6657* (L); JN040399; –; JN040357; JN040283. ***Cannabis sativa*** L. **1,** China, Yunnan; *Yang 003* (KUN); JN040400; JN040323; JN040359; JN040285. ***Cannabis sativa*** L., China, Yunnan; *Yi 10108* (KUN); JN040401; JN040324; JN040360; JN040286. ***Celtis biondii*** Pamp. **2,** China, Yunnan; *Yi 10330* (KUN); JN040404; JN040327; JN040363; JN040289. ***Celtis ehrenbergiana*** (Klotzsch) Liebm. **1,** Netherlands, Gelderland; *Leeuw 481* (WAG); JN040406; JN040329; JN040365; JN040291. ***Celtis ehrenbergiana*** (Klotzsch) Liebm. **2,** Belgium, Vlaams-Brabant; coll. unknown *s.n. 19371993* (Meise Botanical Garden); JN040408; JN040331; JN040367; JN040293. ***Celtis iguanaea*** (Jacq.) Sarg., Germany, Bayern; coll. unknown *s.n.* (KUN); JN040402; JN040325; JN040361; JN040287. ***Celtis madagascariensis*** Sattarian, Madagascar, Toliara; *Phillipson 2938* (MO); JN040405; JN040328; JN040364; JN040290. ***Celtis sinensis*** Pers., Finland, Helsinki; coll. unknown *s.n.* (Helsinki Botanical Garden); JN040407; JN040330; JN040366; JN040292. ***Celtis tournefortii*** Lam., Italy, Catania; coll. unknown *s.n.* (Catania Botanical Garden); JN040403; JN040326; JN040362; JN040288. ***Chaetachme aristata*** Planch. **1,** Madagascar,

Appendix 1. Continued.

Toliara; *McPherson 14423* (WAG); JN040410; –; JN040369; JN040295. ***Chaetachme aristata*** Planch. **2,** South Africa, Transvaal; *Schijff 3444* (L); JN040409; –; JN040368; JN040294. ***Chaetachme aristata*** Planch. **3,** Congo, Orientale; *Bamps 4325* (WAG); JN040411; –; JN040370; –. ***Gironniera celtidifolia*** Gaudich., Papua New Guinea, Hans Meyer Range; *Sands 855* (L); JN040412; JN040335; JN040374; JN040299. ***Gironniera subaequalis*** Planch., China, Yunnan; *DNA Barcoding Group B GBOWS 1411* (KUN); JN040413; JN040336; JN040375; –. ***Humulus lupulus*** L., China, Beijing; *Xie 002* (KUN); JN040416; JN040338; JN040378; JN040302. ***Humulus scandens*** (Lour.) Merr., U.S.A., Maryland; *Windler 4046* (U); JN040415; –; JN040377; JN040301. ***Humulus yunnanensis*** Hu, China, Yunnan; *Yang 005* (KUN); JN040414; JN040337; JN040376; JN040300. ***Lozanella permollis*** Killip & C.V. Morton, Bolivia, Cochabamba; *Solomon 18073* (U); JN040417; JN040339; JN040379; JN040303. ***Lozanella trematoides*** Greenm., Mexico, Hidalgo; *Pringle 8983* (L); JN040418; JN040340; JN040380; JN040304. ***Parasponia andersonii*** Planch., Polynesia; *Meyer 2556* (L); JN040419; JN040341; JN040381; JN040305. ***Parasponia melastomatifolia*** J.J. Sm., Papua New Guinea, Milne Bay; *Pullen 7963* (L); JN040420; –; JN040382; JN040306. ***Pteroceltis tatarinowii*** Maxim. **1,** China, Beijing; *Xie 003* (KUN); JN040421; JN040343; JN040384; JN040308. ***Pteroceltis tatarinowii*** Maxim. **2,** China, Guizhou; *Yi 10081* (KUN); JN040422; JN040344; JN040385; JN040309. ***Trema angustifolia*** (Planch.) Blume, China, Guangxi; *Gong zw2009082601* (KUN); JN040425; JN040347; JN040388; JN040312. ***Trema cannabina*** Lour. **1,** Hawaii, Papaikou; *Lorence 9381* (Hawaii Botanical Garden); JN040426; JN040348; JN040389; JN040313. ***Trema cannabina*** Lour. **2,** China, Guangxi; *Jiang & Yang 09525* (KUN); JN040424; JN040346; JN040387; JN040311. ***Trema discolor*** Blume, Hawaii, Papaikou; *Lorence 9329* (Hawaii Botanical Garden); JN040427; JN040349; JN040390; JN040314. ***Trema domingense*** Urb., Dominican Republic, Duarte; *Ekman 12293* (U); JN040428; JN040350; JN040391; JN040315. ***Trema levigata*** Hand.-Mazz., China, Yunnan; *Tian 0020* (KUN); JN040429; JN040351; JN040392; JN040316. ***Trema micrantha*** (L.) Blume, Bolivia, Beni; *Chatrou 413* (U); JN040430; JN040352; JN040393; JN040317. ***Trema orientalis*** (L.) Blume, Gabon; coll. unknown *s.n.*; JN040431; JN040353; JN040394; JN040318. ***Trema tomentosa*** (Roxb.) H. Hara **1,** China, Guangxi; *Liu 0096* (KUN); JN040432; –; JN040395; JN040319. ***Trema tomentosa*** (Roxb.) H. Hara **2,** China, Guangxi; *Jiang & Yang 09524* (KUN); JN040423; JN040345; JN040386; JN040310.

Appendix 2. Accession numbers of additional sequences from GenBank in the order *trnL–trnF*, *rbcL*, *ndhF*, *matK*. A dash (–) indicates an unavailable sequence.

Aphananthe aspera –; –; AF500366; AF345320; ***Broussonetia papyrifera*** –; –; AY289269; AF345326; ***Cannabis sativa*** **1** –; –; AY289250; AF345317; ***Cannabis sativa*** **2** –; –; AY289250; –; ***Celtis iguanaea*** –; –; –; JQ589979; ***Celtis africana*** –; –; –; JF270686; ***Celtis australis*** –; HE963395; –; HE967374; ***Celtis latifolia*** –; JF738634; –; –; ***Celtis occidentalis*** –; –; –; AY257535; ***Celtis schippii*** –; JX987578; –; GQ981961; ***Celtis sinensis*** –; –; –; AF345316; ***Celtis tetrandra*** –; JF317439; JF317479; JF317420; ***Celtis tournefortii*** **2** AJ575060; –; –; –; ***Chaetachme aristata*** –; –; –; JF270688; ***Cudrania tricuspidata*** –; –; AY289272; JF317421; ***Debregeasia saeneb*** –; –; –; JF317422; ***Ficus tikoua*** –; –; –; JF317426; ***Gironniera subaequalis*** –; –; –; AF345319; ***Humulus scandens*** –; –; –; JQ773628; ***Humulus lupulus*** –; –; AY289251; AF345318; ***Lozanella enantiophylla*** AF501595; AF500341; AF500367; –; ***Parasponia parviflora*** AF501596; AF500342; AF500368; –; ***Parasponia rigida*** AY488675; U59820; –; –; ***Parasponia simulans*** AY488674; –; –; –; ***Pilea cadierei*** –; –; –; JF317431; ***Pteroceltis tatarinowii*** –; –; AF500369; AF345324; ***Trema micrantha*** –; –; –; JQ589372; ***Trema orientalis*** –; –; –; JF270972; ***Trema angustifolia*** –; –; –; JF317434; ***Trema aspera*** AY488681; –; –; –; ***Trema integerrima*** AY488718; –; –; –; ***Trema lamarckiana*** AY488698; –; –; –; ***Trema politoria*** AY488676 ; –; –; –; ***Trema tomentosa*** –; –; –; AF345325; ***Ulmus americana*** –; –; AF500365; –; ***Ulmus parvifolia*** –; –; –; AF345321.

Appendix 3. Morphological data matrix.

Taxon/Character	Sexual system	Leaf arrangement	Pollen aperture number	Aestivation	Fruit type	Seed coat morphology	Perianth at fruiting time	Stipule arrangement
Aphananthe aspera	0	1	0	0	0	1	1	1
Aphananthe monoica	0	1	0	0	0	1	1	1
Aphananthe philippinensis	0	1	0	0	0	1	1	1
Broussonetia papyrifera	1	1	1	0	1	1	1	1
Cannabis sativa 1	3	2	0	1	1	0	1	1
Cannabis sativa 2	3	2	0	1	1	0	1	1
Celtis biondii	4	1	0	1	0	0	0	1
Celtis ehrenbergiana 1	2	1	0	1	0	0	0	1
Celtis ehrenbergiana 2	2	1	0	1	0	0	0	1
Celtis iguanaea	2	1	0	1	0	0	0	1
Celtis madagascariensis	2	1	0	1	0	0	0	1
Celtis sinensis	4	1	0	1	0	0	0	1
Celtis tournefortii	0	1	0	1	0	0	0	1
Chaetachme aristata 1	3	1	0	0	0	0	0	0
Chaetachme aristata 2	3	1	0	0	0	0	0	0
Chaetachme aristata 3	3	1	0	0	0	0	0	0
Cudrania tricuspidata	1	1	0&1	1	1	0	1	1
Debregeasia saeneb	1	1	?	0	1	?	1	0
Ficus tikoua	0	1	1	0	1	1	?	1
Gironniera celtidifolia	3	1	0	1	0	1	1	1
Gironniera subaequalis	3	1	0	1	0	1	1	1
Humulus lupulus	3	2	0	1	1	1	1	2
Humulus scandens	3	2	0	1	1	1	1	2
Humulus yunnanensis	3	2	0	1	1	1	1	2
Lozanella permollis	1	0	0	1	0	1	0	0
Lozanella trematoides	1	0	0	1	0	1	0	0
Parasponia andersonii	0	1	1	1	0	1	1	0
Parasponia melastomatifolia	0	1	1	1	0	1	1	0
Pilea cadierei	1	0	1	0	1	?	1	0
Pteroceltis tatarinowii 1	0	1	0	1	2	0	1	1
Pteroceltis tatarinowii 2	0	1	0	1	2	0	1	1
Trema angustifolia	3	1	1	0	0	1	1	1
Trema cannabina 1	4	1	1	0	0	1	1	1
Trema cannabina 2	4	1	1	0	0	1	1	1
Trema discolor	0	1	1	0	0	1	1	1
Trema domingense	0	1	1	0	0	1	1	1
Trema levigata	0	1	1	0	0	1	0	1
Trema micrantha	4	1	1	0	0	1	1	1
Trema orientalis	0	1	1	0	0	1	1	1
Trema tomentosa 1	0	1	1	0	0	1	1	1
Trema tomentosa 2	0	1	1	0	0	1	1	1
Ulmus macrocarpa	0	1	2	1	2	1	1	1