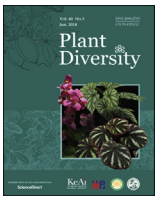
Plant Diversity 40 (2018) 127–137

Contents lists available at ScienceDirect

Plant Diversity

journal homepage: http://www.keaipublishing.com/en/journals/plant-diversity/
http://journal.kib.ac.cn



Plastome characteristics of Cannabaceae

Huanlei Zhang [a, b], Jianjun Jin [a, b], Michael J. Moore [c], Tingshuang Yi [a, *], Dezhu Li [a, **]

[a] Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China
[b] Kunming College of Life Sciences, University of Chinese Academy of Sciences, Kunming 650201, China
[c] Department of Biology, Oberlin College, Oberlin, OH 44074, USA

ARTICLE INFO

Article history:
Received 17 March 2018
Received in revised form
11 April 2018
Accepted 18 April 2018
Available online 23 April 2018

(Editor: Lianming Gao)

Keywords:
Plastome
IR expansion/contraction
Repeats
SSR
Sequence divergence
Phylogenomics

ABSTRACT

Cannabaceae is an economically important family that includes ten genera and ca. 117 accepted species. To explore the structure and size variation of their plastomes, we sequenced ten plastomes representing all ten genera of Cannabaceae. Each plastome possessed the typical angiosperm quadripartite structure and contained a total of 128 genes. The Inverted Repeat (IR) regions in five plastomes had experienced small expansions (330–983 bp) into the Large Single-Copy (LSC) region. The plastome of *Chaetachme aristata* has experienced a 942-bp IR contraction and lost *rpl22* and *rps19* in its IRs. The substitution rates of *rps19* and *rpl22* decreased after they shifted from the LSC to IR. A 270-bp inversion was detected in the *Parasponia rugosa* plastome, which might have been mediated by 18-bp inverted repeats. Repeat sequences, simple sequence repeats, and nucleotide substitution rates varied among these plastomes. Molecular markers with more than 13% variable sites and 5% parsimony-informative sites were identified, which may be useful for further phylogenetic analysis and species identification. Our results show strong support for a sister relationship between *Gironniera* and *Lozanell* (BS = 100). *Celtis*, *Cannabis-Humulus*, *Chaetachme-Pteroceltis*, and *Trema-Parasponia* formed a strongly supported clade, and their relationships were well resolved with strong support (BS = 100). The availability of these ten plastomes provides valuable genetic information for accurately identifying species, clarifying taxonomy and reconstructing the intergeneric phylogeny of Cannabaceae.

1. Introduction

Cannabaceae *sensu* APG IV (Byng et al., 2016) comprise ten genera (Lipton, 1997; Sytsma et al., 2002; Haston et al., 2007, 2009; Mabberley, 2008; Bell et al., 2010) and ca. 117 species (Jin et al., unpublished). Most Cannabaceae species are trees and shrubs, while some are herbs (*Cannabis* L.) or vines (*Humulus* L.). The family has a cosmopolitan distribution; *Aphananthe* (Thunb.) Planch., *Celtis* L. and *Trema* Lour. are widely distributed in tropical and temperate regions (Yang et al., 2013; Jin et al., unpublished); the remaining genera have restricted distributions. A few species of this family are of great economic importance. *Cannabis sativa* L. (hemp) is one of earliest and most important domesticated food and fiber crops, and an increasingly important drug used for its anesthetic and antipsychotic properties (Measham et al., 1994; Kostic et al., 2008; Marks et al., 2009). *Humulus lupulus* L. (hops) is a key ingredient for brewing beer (Wilson, 1975; Murakami et al., 2006), and the phloem fiber of *Pteroceltis tatarinowii* Maxim. is the sole raw material for manufacturing traditional Chinese Xuan paper (Cao, 1993).

There are long-standing controversies over the circumscription and phylogenetic position of Cannabaceae. Cannabaceae was first separated from Moraceae by Rendle (1925). The circumscription of this family has been expanded significantly to include most former members of Ulmaceae subfam. Celtidoideae *sensu* Engler and Prantl (1893) or Celtidaceae *sensu* Link (1829) (Yang et al., 2013). A series of molecular studies elucidated the phylogenetic position of this family, which was supported to be a member of Rosales and sister to Moraceae and Urticaceae (Sytsma et al., 2002; Van Velzen et al., 2006; Wang et al., 2009; Zhang et al., 2011a,b). Multiple molecular studies have also helped to clarify intergeneric relationships of the family (Yang et al., 2013; Jin et al., unpublished). However, a few

* Corresponding author.
** Corresponding author.
E-mail addresses: tingshuangyi@mail.kib.ac.cn (T. Yi), dzl@mail.kib.ac.cn (D. Li).
Peer review under responsibility of Editorial Office of Plant Diversity.

https://doi.org/10.1016/j.pld.2018.04.003

nodes among genera have remained unresolved with weak support (Yang et al., 2013).

The plastome of angiosperms is usually conserved in gene content and structure, typically featuring two ~25 kb Inverted Repeat (IR) regions separating the remainder of the genome into Large and Small Single-Copy regions (LSC, SSC). Size variation among plastomes is mostly due to the expansion or contraction of the IR and/or larger indels, as for example caused by the loss of genes (especially the *ndh* genes) (Downie and Jansen, 2015). Plastomes have proved highly valuable in resolving difficult phylogenetic relationships at both deeper taxonomic levels (e.g. Jansen et al., 2007; Moore et al., 2007, 2010), as well as at more shallow levels (e.g. Zhang et al., 2011a,b; Givnish et al., 2015; Wysocki et al., 2015; Duvall et al., 2016).

In this article, we report the complete plastome sequences of ten species representing all ten genera of Cannabaceae. We annotated the plastomes in detail, identified structure and size variation, and determined the distribution and location of microsatellites (SSRs) and repeats. We demonstrate that the resulting plastome information will be widely useful for understanding phylogenetic relationships, population genetics and breeding programs across the family.

2. Materials and methods

2.1. Chloroplast DNA extraction and sequencing

We used about 100 mg of fresh leaf material of each species (see Table S1 for voucher specimens). Total genomic DNA was extracted with a modified CTAB (Cetyl Trimethyl Ammonium Bromide) method (Doyle and Doyle, 1987), in which 4% CTAB with approximately 1% polyvinyl polypyrrolidone (PVP) and 0.2% DL-dithiothreitol (DTT) was included (Yang et al., 2014). Long-range polymerase chain reaction (PCR) was used for DNA amplification of the plastome using 15 universal primers pairs and methods described by Zhang et al. (2016). Illumina Nextera XT libraries (Illumina, San Diego, CA, USA) with 500 bp inserts were constructed following the manufacturer's protocol. Paired-end (PE) sequencing was performed on an Illumina Hiseq 2500 instrument at the Beijing Genomics Institute (BGI, Shenzhen, Guangdong, China) or on a Hiseq 2000 instrument at the Plant Germplasm and Genomics Center (Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China).

2.2. Plastome assembly and annotation

Raw reads were filtered using NGSQCToolkit (Patel and Jain, 2012; cut-off value for percentage of read length = 80, cut-off value for PHRED quality score = 30) to obtain high quality reads that were free of vector and adaptor sequences. Filtered reads were then assembled into contigs using the software CLC Genomics Workbench 8, via the *de novo* method using a k-mer of 63 and a minimum contig length of 1 kb. Using BLAST (Altschul et al., 1990) with default search parameters, all contigs were aligned to the *Morus mongolica* Schneid. plastome (NC025772.2) as a reference. We mapped the paired reads to the assembled plastomes using Bowtie 2 (Langmead and Salzberg, 2012), as implemented in Geneious v9.5 (Kearse et al., 2012), to verify the IR boundaries, correct some biased bases brought in by the CLC assembler, and detect the number of matched paired-end (PE) reads and the depth of coverage. Lastly, we filled the remaining gaps using long-range PCR and Sanger sequencing. We designed primers based on previous incomplete plastomes (Table S2). Each amplification was performed in 25 μL reaction volume containing 12.5 μL Taq DNA polymerase, 0.5 μL each of forward and reverse primers (dissolved in 10× ddH2O), and 1 μL (30 ng/μL) template DNA. The amplification was conducted using 94 °C for 3 min, 35 cycles of 94 °C for 50 s, 50 °C for 2 min, and 72 °C for 1 min, followed by a final extension step at 72 °C for 8 min. PCR products were sequenced at the Kunming Sequencing Department of Biosune Biotechnology Limited Company (Shanghai, China).

Assembled genomes were annotated using DOGMA (Wyman et al., 2004) along with manual correction of start and stop codons and intron/exon boundaries in Geneious. Transfer RNA (tRNA) genes were further annotated using tRNAscan-SE (Schattner et al., 2005). Genome maps were created in OGDraw 1.2 (Lohse et al., 2013). All annotated plastomes were deposited in GenBank; accession numbers are MH118117–MH11812 that provided in Table S1.

2.3. Phylogenetic analysis

Phylogenetic analyses included all ten genera of Cannabaceae as ingroups, two species of *M. mongolica* (Moraceae) and *Ulmus macrocarpa* Hance (Ulmaceae) representing closely related families as outgroups (Table S1). A total of 237 loci (112 coding and 125 noncoding regions) were extracted from each plastome (exons were joined) for phylogenetic analysis. Loci shared by less than 6 taxa or with length <30 bp were excluded (Table S3). Sequences were aligned using MAFFT version 7 (Katoh and Standley, 2013) with default parameters. Maximum likelihood analysis was performed with RAxMLv8.2.10 (Stamatakis, 2006), by using the '-f a' option, GTRGAMMA model, and 1000 bootstrap replicates, with data partitioned by locus.

2.4. Analysis of sequence divergence

To characterize sequence divergence among all sequenced plastomes of Cannabaceae, we extracted 133 coding and 129 noncoding regions (including intergenic spacers and introns), each of them treated as a separate locus. These regions were aligned using MEGA v6.06 (Tamura et al., 2013). For each alignment, the number of invariant sites, variable but parsimony-uninformative sites, and parsimony-informative sites were calculated, as was pairwise sequence divergence (uncorrected "p" distance), all using PAUP* 4.0a147 (Swofford, 2002). Gaps were treated as missing data. Using the *Humulus scandens* plastome as a reference, sequence identity was also plotted using mVISTA (Frazer et al., 2004) in LAGAN mode.

2.5. Repeat analysis

REPuter (Kurtz et al., 2001) was used to locate sequence repeats including forward, reverse, and palindromic repeats. The minimal repeat size was set to 30 bp and repeat identity was set to ≥90% (hamming distance equal to 3). Before using REPuter to detect repeats, to avoid redundancy we removed the IR_A region from each plastome. However, IR repeats were treated twice (to represent both copies) when summarizing repeats across the genome. Tandem repeats were analyzed using TRF (Tandem Repeat Finder program) web interface (Benson, 1999) with the parameters setting as 2, 7 and 7 for match, mismatch and indel respectively. The minimum alignment score and maximum period size were set as 50 and 500. After analysis, tandem repeats <15 bp in length and the redundant results of REPuter were manually removed (Wang et al., 2017). We also tallied the total number of repeats, measured repeat lengths, and calculated the proportion of repeats in the LSC, SSC, and IR.

2.6. SSR analysis

Microsatellite detection was performed using MISA with minimum number of repeats of 8, 5, 4, 3, 3, and 3 respectively for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats. One copy of the IR was removed prior to microsatellite detection. All of the repeats were manually verified, and redundant results were removed.

3. Results and discussion

3.1. Conservation of Cannabaceae plastomes

Illumina sequencing produced from 289,464 (*Celtis blondii*) to 4,807,452 (*Trema orientalis*) paired-end reads, among which 257,965 (*Celtis blondii*) to 4,346,229 (*T. orientalis*) reads were mapped to their respective assembled genomes. De novo and reference-guided assembly produced full coverage for all plastomes, with mean coverages ranging from 120.3 × (*Celtis blondii*) to 2569.3 × (*T. orientalis*) (Table 1).

All sequenced plastomes displayed the typical quadripartite structure of most angiosperms (Wang et al., 2013; Li et al., 2014). The ten plastomes ranged in size from 153,776 bp (*H. scandens*) to 159,001 bp (*Celtis blondii*). The length of their LSC region varied from 83,885 bp (*H. scandens*) to 87,620 bp (*P. tatarinowii*), that of the SSC region from 17,751 bp (*H. scandens*) to 20,064 bp (*Chaetachme aristata*), and their IR region from 25,512 bp (*T. orientalis*) to 26,879 bp (*Celtis blondii*) (Table 1). The overall GC content was approximately 37.3% across all ten sampled plastomes. The gene content and structural organization of all ten sequenced plastomes were also highly conserved (Fig. 1, Fig. S1). Most plastomes harbored 112 unique genes, including 78 protein-coding genes, 30 transfer RNA (tRNA) genes, and four ribosomal RNA (rRNA) genes. The exceptions were the plastomes of *P. tatarinowii* and *C. aristata*; the former had a pseudogenic *rpl22* and the latter lost *rpl22* (Table 2). All plastomes lost *infA*, which was consistent with those of most eurosids (Millen et al., 2001).

The IR, LSC, and SSC gene content, as well as intron content, for most of the Cannabaceae plastomes matched the typical content for angiosperms, with some differences in IR gene content (Fig. 2, Table S4). The plastomes of *Aphananthe aspera*, *Lozanella enantiophylla*, *Parasponia rugosa* and *T. orientalis* possessed canonical IRs ranging from 25,512 bp in *T. orientalis* to 26,015 bp in *A. aspera*. Their IRs contained 17 complete genes (including six protein-coding genes, seven tRNAs, and all four rRNAs) as well as the 5′ ends of *ycf1* (1037–1076 bp) and *rps19* (0–100 bp). The plastomes of *C. sativa*, *H. scandens*, *P. tatarinowii*, *Celtis blondii* and *Gironniera subaequalis* had longer IRs, ranging from 26,011 bp (*C. sativa*) to 26,879 bp (*Celtis blondii*), caused by 330-bp (*C. sativa*) to 983-bp (*Celtis blondii*) IR expansions into the LSC; specifically, IRs expanded into all of *rps19* and all or part of *rpl22* (25–408 bp). In contrast, *C. aristata* had the shortest IR at 25,566 bp, due to a 942-bp IR contraction. Its IRs lost *rps19* and *rpl22*, but *rps19* was found before *trnH-GUG* in LSC near the IRa/LSC junction (J_{LA}). IRs of *C. aristata* may have experienced more than a 942-bp IR expansion into LSC firstly to include *rps19* and *rpl22*, followed by the loss of *rps19* (279 bp) and *rpl22* (408 bp) from IRb and *rpl22* from IRa. In contrast, the IR/SSC junctions showed little variation, including 0 (*A. aspera*) to 45 bp (*L. enantiophylla*) of the 3′ end of *ndhF*.

IR expansion and contraction are common, especially small contractions and expansions of <100 base pairs (bp), and the positions of four IR/single-copy junctions can vary even among closely related species (Goulding et al., 1996; Plunkett and Downie, 2000). Large IR expansions occur less frequently and sometimes accompany structural rearrangements elsewhere in the plastid genome (Guisinger et al., 2011; Wicke et al., 2011). Cannabaceae provide yet another example of moderate to small IR expansion and contraction. IR expansion has been suggested to start with double-strand breaks followed by strand invasion and recombination (Goulding et al., 1996; Wang et al., 2008). Regions with a high content of short repeats or “poly A tracts” were inferred to be associated with the dynamics of IR-LSC junctions and expansions of IR (Wang et al., 2008; Dugas et al., 2015). In Cannabaceae plastomes with expanded IRs, a region ca. 100 bp upstream of the IR-LSC junctions was found to be extremely AT-rich (>90%), including many poly A tracts and short repeats, which could explain the IR expansion of Cannabaceae plastomes. Large IR contractions have been rarely reported, and illegitimate recombination has been considered as the most plausible explanation (Goulding et al., 1996; Downie and Jansen, 2015; Blazier et al., 2016), which may also account for the IR contraction in *C. aristata*.

Nucleotide substitution rates of most plastome coding genes have been demonstrated to decrease after translocation from SC regions to the IR (Lin et al., 2012; Li et al., 2016; Zhu et al., 2016; but see exceptions in Lin et al., 2012; Wang et al., 2017). In this study, we also found a decrease of substitution rates for *rps19* (0.0154) and *rpl22* (0.0229) after their shifts from LSC into IR.

Finally, an interesting 270-bp inversion between *petN* and *psbM* was detected in the plastome of *P. rugosa*, representing the first known reasonably long inversion in Cannabaceae plastomes. A pair of 18-bp inverted repeats resided at the boundaries of this inversion, and it is likely that these repeats helped mediate this inversion, as seen for other smaller inversions (Kim et al., 2005; Qu et al., 2017a,b). Likewise, short repeats have also been inferred to associated with large inversions, such as the association of 29-kb repeats with a 36-kb inversion in legumes (Martin et al., 2014); the association ≥ 20-bp repeats with a 45-kb inversion of *Medicago truncatula* (Gurdon and Maliga, 2014); and the association of 11-bp repeats with a 36-kb inversion in *Calocedrus macrolepis* (Qu et al., 2017a,b).

Table 1
Assembly statistics and genome features for newly sequenced Cannabaceae plastomes.

Species	Total PE reads	Matched PE reads	Mean coverage (×)	Genome length (bp)	LSC length (bp)	SSC length (bp)	IR length (bp)	GC content (%)
Aphananthe aspera	1,695,716	374,611	583.7	157,687	86,135	19,442	26,015	36.4
Cannabis sativa	2,040,500	1,880,700	1351.8	153,910	84,059	17,829	26,011	36.7
Celtis blondii	289,464	257,965	120.3	159,001	86,072	19,171	26,879	36.3
Chaetachme aristata	1,142,608	1,045,891	1415.4	157,939	86,743	20,064	25,566	36.1
Gironniera subaequalis	396,352	374,583	583.6	157,807	86,215	18,942	26,325	36.3
Humulus scandens	1,010,646	839,251	1436.6	153,776	83,885	17,751	26,070	36.9
Lozanella enantiophylla	1,077,002	1,026,115	1573.4	156,711	85,928	19,133	25,825	36.6
Parasponia rugosa	586,024	498,328	627.5	157,434	86,961	19,313	25,580	36.3
Pteroceltis tatarinowii	1,051,832	992,380	1711.1	158,504	87,620	18,856	26,014	36.3
Trema orientalis	4,807,452	4,346,229	2569.3	157,192	86,859	19,309	25,512	36.3

PE = paired-end; LSC = Large Single-Copy region; SSC = Small Single-Copy region; IR = Inverted Repeat region.

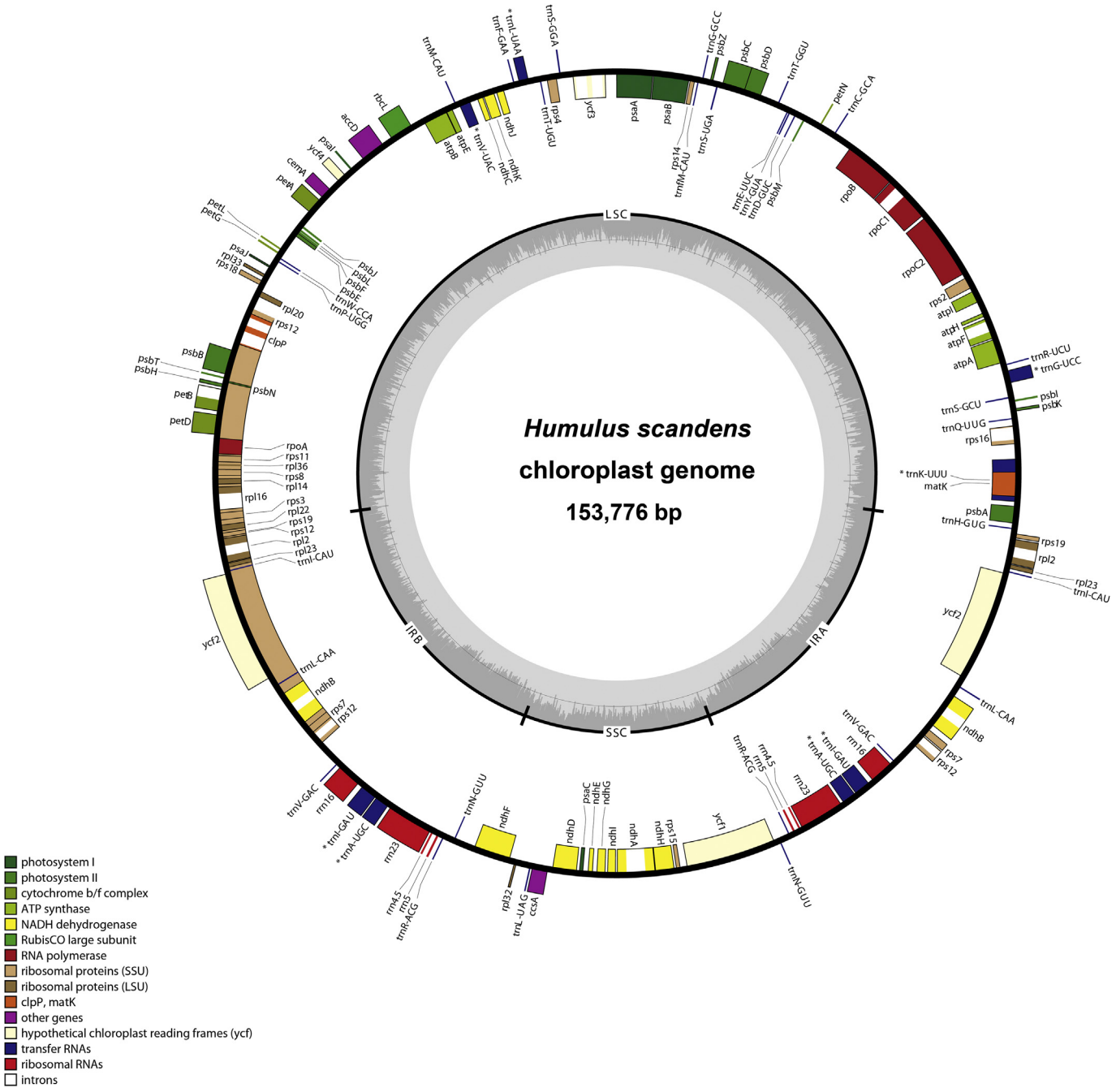


Fig. 1. Gene maps of the plastome of *Humulus scandens*. Genes are indicated by boxes on the inside (clockwise transcription) and outside (counterclockwise transcription) of the outermost circle. The inner circle identifies the major structural components of the plastome (LSC, IR, and SSC). Genes belonging to different functional groups are color-coded. Dashed area in the inner circle indicates the GC content of the plastome. * represents the tRNA with an intron.

3.2. Phylogenetic relationships

The monophyly of Cannabaceae was strongly supported (BS = 100). Relationships among the ten genera of Cannabaceae were also fully resolved with high bootstrap support (BS) (Fig. 3). Complete plastome sequences have also been used to successfully resolve intergeneric relationships in many other vascular plants (e.g. Givnish et al., 2015; Qu et al., 2017a,b; Zhang et al., 2017; Wang et al., 2018), and our study provides yet another example. Some previously resolved intrafamilial relationships were strongly supported in this study (Fig. 3): *Aphananthe* was sister to other genera of Cannabaceae (Song et al., 2001; Sytsma et al., 2002; Van Velzen et al., 2006; Yang et al., 2013); *Gironniera*, *Lozanella* and the clade B together formed a monophyletic group (Yang et al., 2013); *Chaetachme* and *Pteroceltis* were sisters (Van Velzen et al., 2006; Yang et al., 2013); *Cannabis* and *Humulus* were sisters (Song et al., 2001; Song and Li, 2002; Sytsma et al., 2002); *Parasponia* was nested within *Trema* (Zavada and Kim, 1996; Sytsma et al., 2002; Yesson et al., 2004; Van Velzen et al., 2006; Yang et al., 2013). However, our study supported some new relationships. Our results show strong support (BS = 100) for a sister relationship between *Gironniera* and *Lozanella*. *Celtis* was strongly supported to be sister

Table 2
Gene content in Cannabaceae plastomes.

Category	Gene groups	Name of genes
Self- replication	Large subunit of ribosomal proteins	*rpl2*[b] (×2), *rpl14*, *rpl16*[b], *rpl20*, *rpl22* (×2)[e,f], *rpl23* (×2), *rpl32*, *rpl33*, *rpl36*
	Small subunit of ribosomal proteins	*rps2*, *rps3*, *rps4*, *rps7* (×2), *rps8*, *rps11*, *rps12*[a–c] (×2), *rps14*, *rps15*, *rps16*[b], *rps18*, *rps19* (×2)[d]
	DNA-dependent RNA polymerase	*rpoA*, *rpoB*, *rpoC1*[b], *rpoC2*
	Ribosomal RNA genes	*rrn4.5* (×2), *rrn5* (×2), *rrn16* (×2), *rrn23* (×2)
	Transfer RNA genes	*trnA-UGC* (×2)[b], *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC*, *trnG-UCC*[b], *trnH-GUG*, *trnI-CAU* (×2), *trnI-GAU* (×2)[b], *trnK-UUU*[b], *trnL-CAA* (×2), *trnL-UAA*[b], *trnL-UAG*, *trnM-CAU*, *trnN-GUU* (×2), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG* (×2), *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC* (×2), *trnV-UAC*[b], *trnW-CCA*, *trnY-GUA*
Photosynthesis	Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
	Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
	NADH dehydrogenase	*ndhA*[b], *ndhB*[b] (×2), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
	Cytochrome b/f complex	*petA*, *petB*[b], *petD*[b], *petG*, *petL*, *petN*
	ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*[b], *atpH*, *atpI*
	RubisCo large subunit	*rbcL*
Other genes	Maturase K	*matK*
	Envelope membrane protein	*cemA*
	Subunit of acetyl- CoA carboxylase	*accD*
	c-type cytochrome synthesis gene	*ccsA*
	Protease	*clpP*[a]
	Proteins of unknown function	*ycf1*, *ycf2* (×2), *ycf3*[a], *ycf4*

(×2) = gene present twice due to position within the IR; [a] Contains two introns; [b] Contains one intron; [c] Exons separated and joined by trans-splicing; [d] gene present in the IRs in the IR-expanded species; [e] Gene present in the IR of *Celtis blondii*; [f] Gene present in the IR of *Chaetachme aristata*.

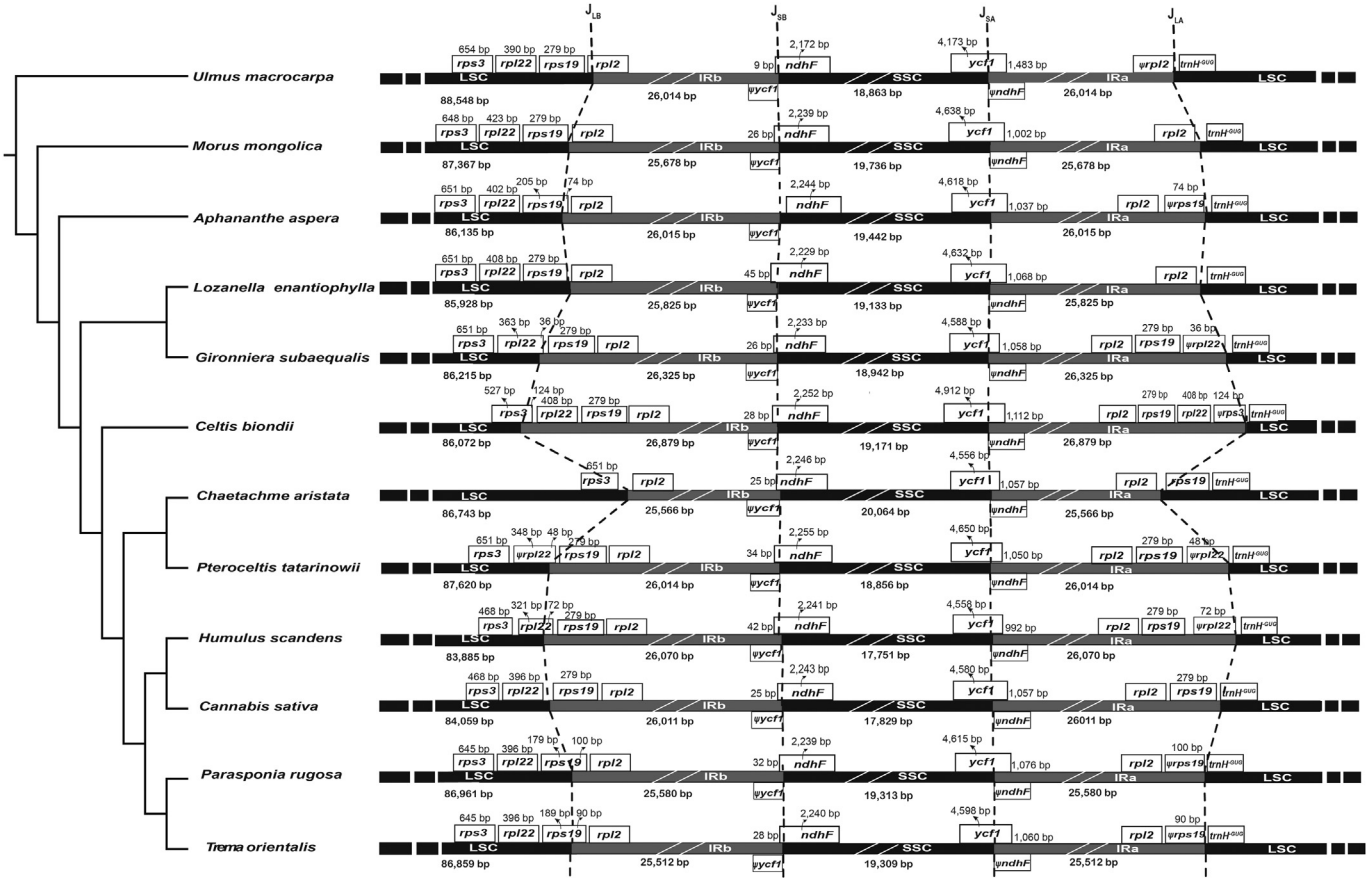


Fig. 2. Comparison of IR/SC boundaries among Cannabaceae plastomes. JSB, JSA and JLA refer to junctions of SSC/IRB, SSC/IRA, and LSC/IRA, respectively. Ψ indicates a pseudogene copy of a gene partially duplicated in the IR.

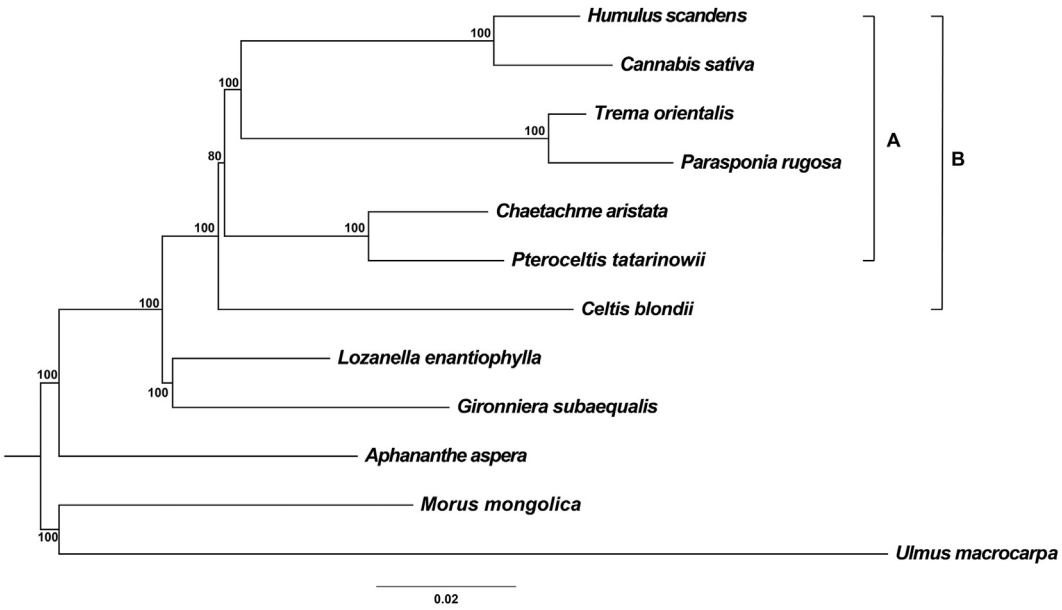


Fig. 3. The best maximum likelihood (ML) tree based on RAxML analysis. Bootstrap support values are provided next to each node.

of clade A (BS = 100). The *Humulus-Cannabis* clade and the *Trema-Parasponia* clade were sisters with strong support (BS = 100). Morphologically, they all have persistent tepals and stigmas. The *Chaetachme-Pteroceltis* clade was sister to the *Humulus-Cannabis-Trema-Parasponia* with relatively low support (BS = 80).

3.3. Sequence divergence and phylogenetic informativeness

Sequence alignments and the mVISTA plot (Fig. 4) revealed high sequence similarity among Cannabaceae plastomes. Aligned lengths of 133 coding and 129 noncoding regions ranged from 9 bp (*psbF-psbE* intergenic spacer) to 6828 bp (*ycf2*). The number of variable sites ranged from 0 (for 20 loci) to 943 (*ycf1*), and the number of parsimony-informative sites ranged from 0 (for 26 loci) to 392 (*ycf1*). Percentages of variable and parsimony-informative sites in coding and noncoding regions are provided in Fig. 5A and B and Table S5. Among coding regions, *matK*, *rps8*, *rpl22*, *ndhF* and *ycf1* had the highest percentages of variable and parsimony-informative sites, with *matK* having an especially high percentage of variable sites (14.05%) and *rpl22* having a high percentage of parsimony-informative sites (6.70%). The percentages of variable sites in noncoding regions ranged from 0 to 28.93% with a mean value of 9.43%, which was nearly twice that of coding regions (5.24% on average). The five noncoding regions with highest percentages of variable sites were *trnfM-CAU-rps14*, *psaI-ycf4*, *petD-2-rpoA*, *rpl36-rps8* and *rps15-ycf1*, with *rpl36-rps8* having the highest percentage of variable (28.93%) and parsimony-informative sites (10.85%). The five noncoding regions with highest percentage of parsimony-informative sites were *rpl33-rps18*, *clpP-3-clpP-2*, *rpoA-rps11*, *rpl36-rps8* and *rps15-ycf1*. The proportions of parsimony-informative sites in noncoding regions ranged from 0 to 10.85% with a mean value of 2.99%, which was higher than that of the coding regions (2.19% on average). In IRs, both of the percentages of variable sites and informative sites ranged from 0 to 2.78% with a mean value of 0.88% in coding regions. Among noncoding regions, the percentages of variable sites ranged from 0 to 6.93% with a mean value of 2.65%, which was similar low to the percentages of PIS (0–2.97% and mean of 1.00%). These findings all showed that fewer mutations were observed within IR regions, including coding and non-coding regions, than LSC and SSC regions. Those with no mutations were mostly tRNAs and *rrn5*, illustrating that tRNAs are more conserved than other genes.

Plastomes supply many valuable loci for reconstructing phylogenetic relationships at multiple taxonomic scales. A number of plastid coding and noncoding loci have been used in phylogenetic studies among genera in the same family, including for example *atpB*, *atpB-rbcL*, *matK*, *ndhF*, *rbcL*, *rpl16*, *rps4-trnS*, *rps16*, *trnH-psbA*, *trnL-F*, and *trnS-G* (Kim and Jansen, 1995; Gao et al., 2008; Hilu et al., 2008; Wilson, 2009; Peterson et al., 2010). Some plastome regions, such as *atpF-H*, *matK*, *psbK-I*, *rbcL*, *rpoB*, *rpoC1*, *trnH-psbA*, etc., have been relied upon heavily for development of candidate markers for plant DNA barcoding (Kress et al., 2005; Newmaster et al., 2006; Chase et al., 2007; Hollingsworth et al., 2011; Dong et al., 2012). The fast-evolving loci we identified, such as *rpl36-rps8*, *rpl22*, *rpl33-rps18*, *rps15-ycf1*, *matK* and *rps8* could be applied to resolve inter- or intraspecific relationships.

3.4. Repetitive sequences

Repeat regions are thought to play an important role in genome recombination and rearrangement (Smith, 2002). In this study, a total of 431 repeats were detected across all Cannabaceae plastomes, including 116 dispersed repeats and 314 tandem repeats (Table S6). Among all ten plastomes, *T. orientalis* had the most repeats (56) and *C. sativa* had the fewest (29). After excluding overlapped repeats detected by REPuter and accounting for both IR copies, 7 (*G. subaequalis*) –19 (*C. aristata*) pairs of dispersed repeats were identified. Plastomes of *C. aristata*, *P. rugosa*, and *T. orientalis* had three repeat types—direct, reverse and palindromic repeats (Fig. 6). Among these, 61% were direct, 33% were palindromic and 6% were reverse. The lengths of repeats ranged from 30 to 55 bp. The total length of dispersed repeats ranged from 541 (*G. subaequalis*) to 1229 bp (*C. aristata*), and their proportion of the whole plastome ranged from 0.34% (*G. subaequalis*) to 0.77% (*C. aristata*). We detected 20 (*C. sativa*)–42 (*T. orientalis*) tandem repeats with a size $\geq$ 15 bp, of which 184 were 15–20 bp in size, 112 were 21–30 bp, 13 were 31–40 bp, four were 41–50 bp, and one was 61 bp (in *A. aspera*). The total length of tandem repeats ranged from

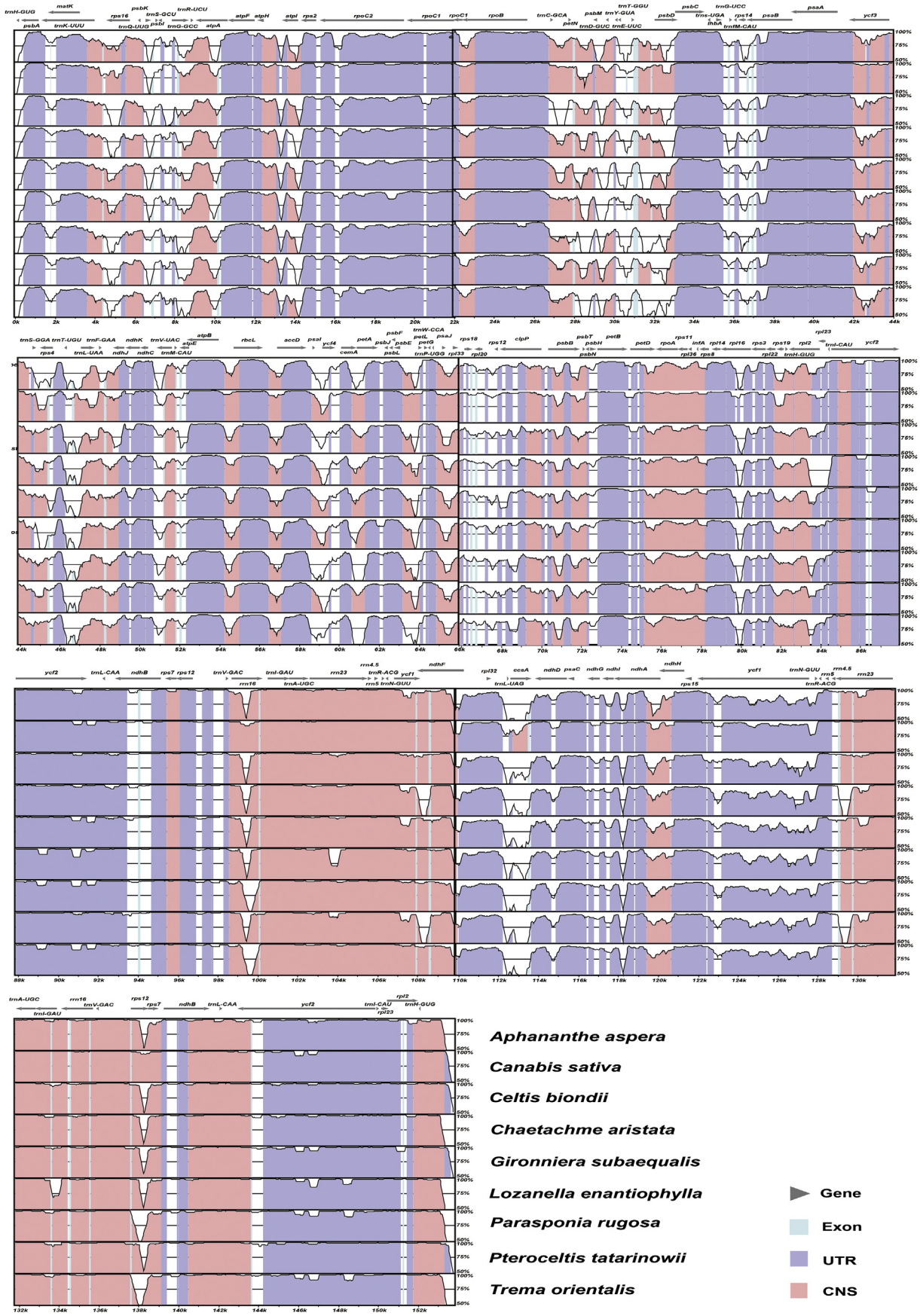


Fig. 4. mVISTA-based identity plot showing sequence identity among Cannabaceae plastomes. *Humulus scandens* is set as the reference. Coding and noncoding regions are colored in blue and red, respectively.

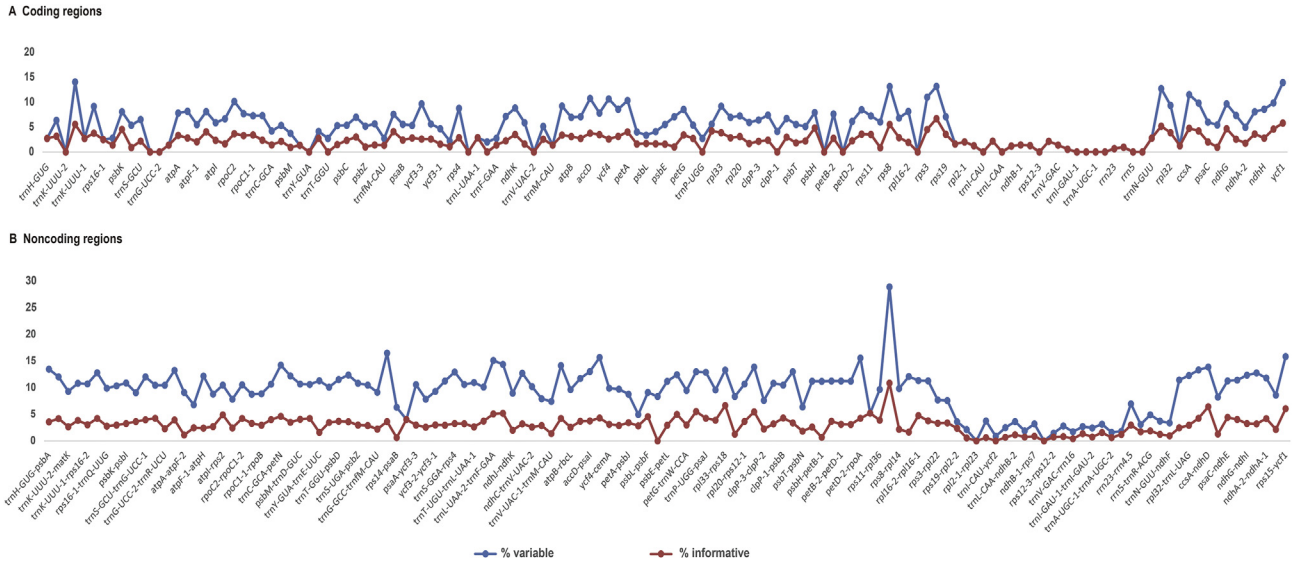


Fig. 5. Percentages of variable (blue, top line) and parsimony-informative (red, bottom line) sites across coding and non-coding loci. A coding regions; **B** noncoding regions. Regions are oriented according to their genome locations.

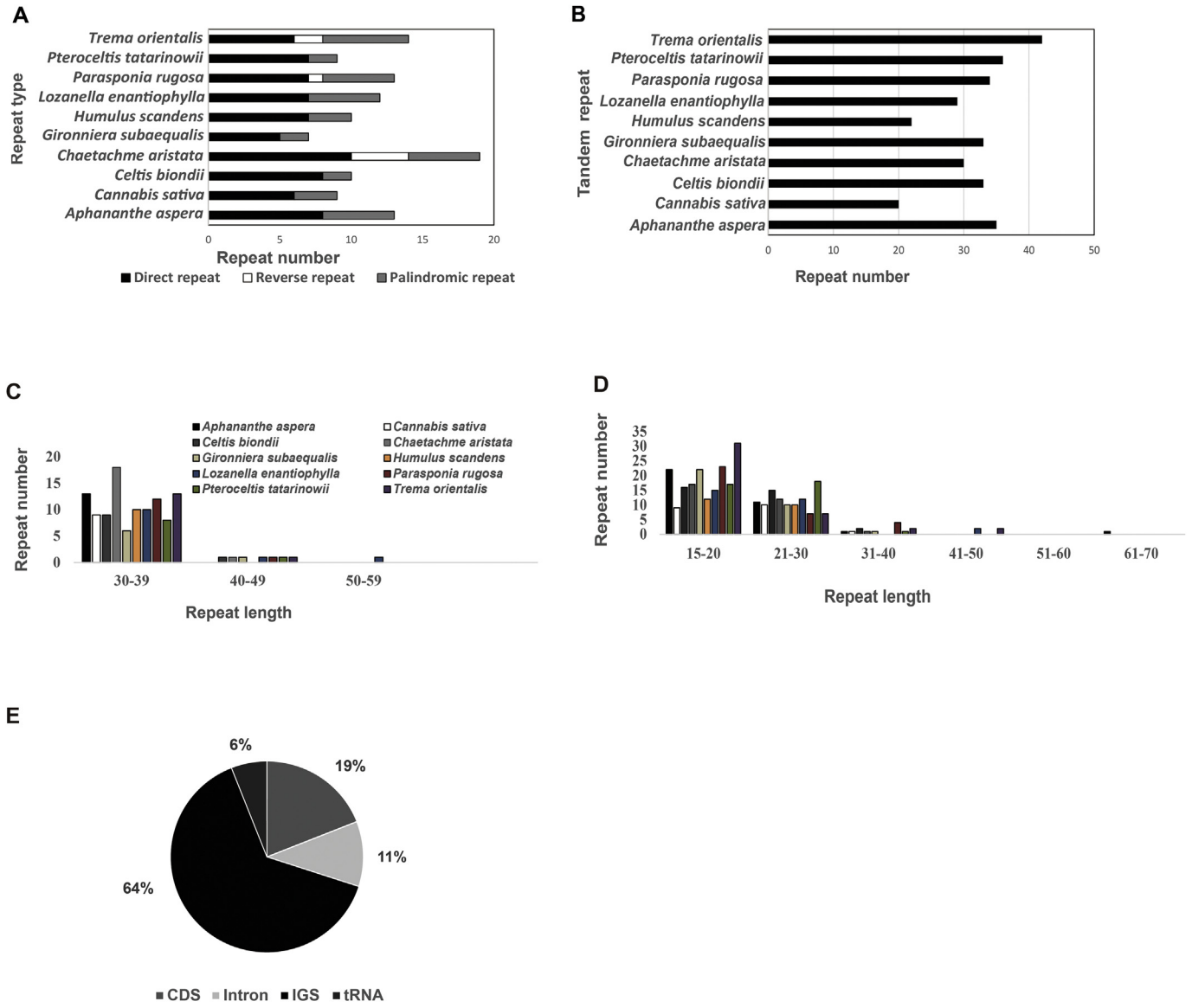


Fig. 6. Analyses of repeated sequences in Cannabaceae plastomes. A Numbers of the three dispersed repeat types; **B** Numbers of tandem repeats; **C** Frequency of dispersed repeats by length; **D** Frequency of tandem repeats by length; **E** The locations of repeats.

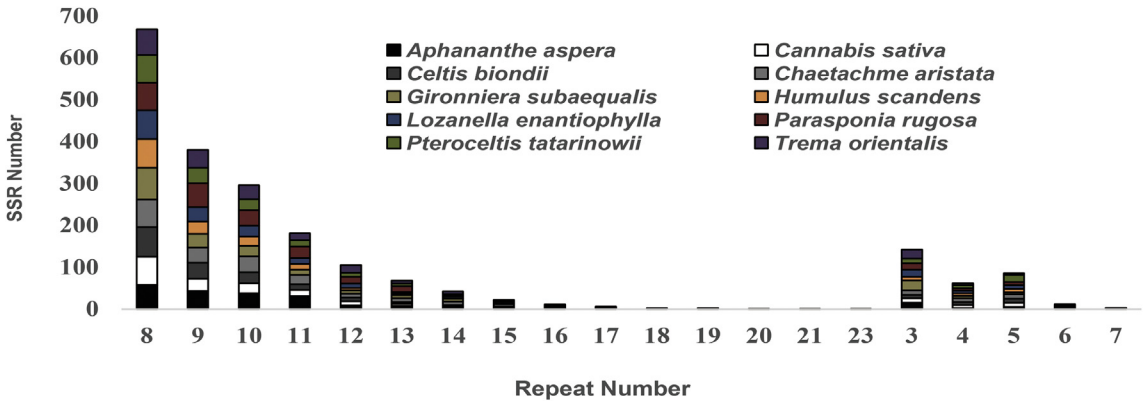


Fig. 7. The distribution of the simple sequence repeats (SSRs) in Cannabaceae plastomes.

950 (*H. scandens*) to 1727 bp (*T. orientalis*), and their proportion of the whole plastome ranged from 0.62% (*H. scandens*) to 1.59% (*C. aristata*). Across all repeats, most were located in intergenic spacer regions (64%), followed by coding sequences (19%), introns (11%), and tRNAs (6%).

3.5. Simple sequence repeat (SSR) polymorphisms

SSRs, including mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, were detected in all plastomes, although hexanucleotide repeats were absent from the plastomes of *Celtis blondii, H. scandens,* and *P. rugosa*. (see Table S7 for a comprehensive list of SSRs, including their positions within the plastome). In total, 221, 186, 193, 229, 210, 172, 195, 250, 209 and 228 SSRs were found in the plastomes of *Aphananthe spera*, *C. sativa, Celtis blondii, C. aristata, G. subaequalis, H. scandens, L. enantiophylla, P. rugosa, P. tatarinowii* and *T. orientalis*, respectively. The majority of mononucleotide repeat units were A/T, ranging from 8 to 23 bp in length (Fig. 7; the longest was present in *T. orientalis*). This finding is consistent with previous observations that cpSSRs are dominated by A/T mononucleotide repeats (Kuang et al., 2011). SSR loci were mainly located within intergenic spacers, followed by coding sequences and introns. Most SSRs were located in the LSC region, followed by the IR and SSC regions. SSRs have been used to understand evolutionary relationships among some closely related plant taxa, and are also effective genetic markers for studying plant breeding, population genetics, biological conservation, mating systems, and uniparental lineages (Terrab et al., 2006; Cardle et al., 2000; Peakall et al., 1998). The SSRs characterized in this study may prove useful for understanding phylogeography and genetic structure of populations.

4. Conclusion

We reported ten complete plastomes in Cannabaceae using Illumina sequencing technology via a combination of de novo and reference-guided assembly. These plastomes were relatively conserved, but the IR regions in some plastomes experienced small expansions and contractions. Substitution rates were calculated after the genes shifted from the LSC to IR. We investigated the variation of repeat sequences, SSRs, and sequence divergence among the ten complete plastomes. Molecular markers with rapid evolution rates were identified, which may be useful for further phylogenetic analysis and species identification. Phylogenies were constructed using the entire genomes. The availability of these ten plastomes provided valuable genetic information for accurately identifying species, clarifying taxonomy and reconstructing the intergeneric phylogeny of Cannabaceae.

Acknowledgments

This study was supported by grants from the National Natural Science Foundation of China, key international (regional) cooperative research project (31720103903), The Strategic Priority Research Program of the Chinese Academy of Sciences (XDPB0201). We would like to thank the Beijing Botanical Garden, Shanghai Chen Shan Botanical Garden, Wuhan Botanical Garde, Missouri Botanical Garden, and San Francisco Botanical Garden for permission to sample fresh leaves, Shudong Zhang, Jie Cai for providing samples, Yinhuan Wang, Rong Zhang for experimental assistance, Xiaojian Qu, Siyun Chen, Yingying Yang for data analysis and their valuable comments. This study was conducted in the Key Laboratory of the Southwest China Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.pld.2018.04.003.

References

Altschul, S.F., Gish, W., Miller, W., et al., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Bell, C.D., Soltis, D.E., Soltis, P.S., 2010. The age and diversification of the angiosperms re-revisited. Am. J. Bot. 97, 1296–1303.
Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580.
Blazier, J.C., Jansen, R.K., Mower, J.P., et al., 2016. Variable presence of the inverted repeat and plastome stability in *Erodium*. Ann. Bot. 117, 1209–1220.
Byng, J.W., Chase, M.W., Christenhusz, M.J.M., et al., 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. 181, 1–20.
Cao, T.S., 1993. Xuan Paper of China. China Light Industry, Beijing, pp. 20–34.
Cardle, L., Ramsay, L., Milbourne, D., et al., 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156, 847–854.
Chase, M.W., Cowan, R.S., Hollingsworth, P.M., et al., 2007. A proposal for a standardised protocol to barcode all land plants. Taxon 56, 295–299.
Dong, W.P., Liu, J., Yu, J., et al., 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS One 7, e35071.
Downie, S.R., Jansen, R.K., 2015. A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. Syst. Bot. 40, 336–351.
Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.

Dugas, D.V., Hernandez, D., Koenen, E.J.M., et al., 2015. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. Sci. Rep. 5, 16958.
Duvall, M.R., Fisher, A.E., Columbus, J.T., et al., 2016. Phylogenomics and plastome evolution of the chloridoid grasses (Chloridoideae: Poaceae). Int. J. Plant Sci. 177, 235–246.
Engler, A., Prantl, K., 1893. Die natürlichen Pflanzenfamilien. III, 4. Engelmann, Leipzig, pp. 202–230.
Frazer, K.A., Pachter, L., Poliakov, A., et al., 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 32, W273–W279.
Gao, X., Zhu, Y.P., Wu, B.C., et al., 2008. Phylogeny of *Dioscorea* sect. *Stenophora* based on chloroplast *matK, rbcL* and *trnL-F* sequences. J. Syst. Evol. 46, 315–321.
Givnish, T.J., Spalink, D., Ames, M., et al., 2015. Orchid phylogenomics and multiple drivers of their extraordinary diversification. Proc. R. Soc. B. Biol. Sci. 282, 171–180.
Goulding, S.E., Olmstead, R.G., Morden, C.W., et al., 1996. Ebb and flow of the chloroplast inverted repeat. Mol. Gen. Genet. 252, 195–206.
Guisinger, M.M., Kuehl, J.V., Boore, J.L., et al., 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol. Biol. Evol. 28, 583–600.
Gurdon, C., Maliga, P., 2014. Two distinct plastid genome configurations and unprecedented intraspecies length variation in the *accD* coding region in *Medicago truncatula*. DNA Res. 21, 417–427.
Haston, E., Richardson, J.E., Stevens, P.F., et al., 2007. A linear sequence of Angiosperm Phylogeny Group II families. Taxon 56, 7–12.
Haston, E., Richardson, J.E., Stevens, P.F., et al., 2009. The Linear Angiosperm Phylogeny Group (LAPG) III: a linear sequence of the families in APG III. Bot. J. Linn. Soc. 56, 128–131.
Hilu, K.W., Black, C., Diouf, D., et al., 2008. Phylogenetic signal in *matK* vs. *trnK*: a case study in early diverging eudicots (angiosperms). Mol. Phylogen. Evol. 48, 1120–1130.
Hollingsworth, P.M., Graham, S.W., Little, D.P., 2011. Choosing and using a plant DNA barcode. PLoS One 6, e19254.
Jansen, R.K., Cai, Z., Raubeson, L.A., et al., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl. Acad. Sci. U. S. A 104, 19369–19374.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kearse, M., Moir, R., Wilson, A., et al., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Kim, K.J., Choi, K.S., Jansen, R.K., 2005. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). Mol. Biol. Evol. 22, 1783–1792.
Kim, K.J., Jansen, R.K., 1995. *NdhF* sequence evolution and the major clades in the sunflower family. Proc. Natl. Acad. Sci. U. S. A 92, 10379–10383.
Kostic, M., Pejic, B., Skundric, P., 2008. Quality of chemically modified hemp fibres. Bioresour. Technol. 99, 94–99.
Kress, W.J., Wurdack, K.J., Zimmer, E.A., et al., 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. U. S. A 102, 8369–8374.
Kuang, D.Y., Wu, H., Wang, Y.L., et al., 2011. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. Genome 54, 663–673.
Kurtz, S., Choudhuri, J.V., Ohlebusch, E., et al., 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29, 4633–4642.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.
Li, F.W., Kuo, L.Y., Pryer, K.M., et al., 2016. Genes translocated into the plastid inverted repeat show decelerated substitution rates and elevated GC content. Genome Biol. Evol. 8, 2452–2458.
Li, H., Cao, H., Cai, Y.F., et al., 2014. The complete chloroplast genome sequence of sugar beet (*Beta vulgaris* ssp. *vulgaris*). Mitochondr. DNA 25, 209–211.
Lin, C.P., Wu, C.S., Huang, Y.Y., et al., 2012. The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. Genome Biol. Evol. 4, 374–381.
Lipton, L.E., 1997. Flora of north America north of Mexico. Lib. J. 3, 122–150.
Lohse, M., Drechsel, O., Kahlau, S., et al., 2013. Organellar Genome DRAW-asuite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression datasets. Nucleic Acids Res. 41, 575–581.
Mabberley, D.J., 2008. Mabberley's Plant-book: a Portable Dictionary of Plants, Their Classification and Uses. Cambridge University, New York, p. 147.
Marks, M.D., Tian, L., Wenger, J.P., et al., 2009. Identification of candidate genes affecting D9-tetrahydrocannabinol biosynthesis in *Cannabis sativa*. J. Exp. Bot. 60, 3715–3726.
Martin, G.E., Rousseau-Gueutin, M., Cordonnier, S., et al., 2014. The first complete chloroplast genome of the Genistoid legume *Lupinus luteus*: evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. Ann. Bot. 113, 1197–1210.
Measham, F., Newcombe, R., Parker, H., 1994. The normalization of recreational drug use amongst young people in north-west England. Br. J. Sociol. 45, 287–312.
Moore, M.J., Bell, C.D., Soltis, P.S., et al., 2007. Using plastid genomescale data to resolve enigmatic relationships among basal angiosperms. Proc. Natl. Acad. Sci. U. S. A 104, 19363–19368.
Moore, M.J., Soltis, P.S., Bell, C.D., et al., 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc. Natl. Acad. Sci. U. S. A 107, 4623–4628.
Millen, R.S., Olmstead, R.G., Adams, K.L., et al., 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13, 645–658.
Murakami, A., Darby, P., Javornik, B., et al., 2006. Molecular phylogeny of wild hops, *Humulus lupulus* L. Heredity 97, 66–74.
Newmaster, S.G., Fazekas, A.J., Ragupathy, S., 2006. DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach. Can. J. Bot. 84, 335–341.
Patel, R.K., Jain, M., 2012. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7, e30619.
Peakall, R., Gilmore, S., Keys, W., et al., 1998. Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. Mol. Biol. Evol. 15, 1275–1287.
Peterson, P.M., Romaschenko, K., Johnson, G., 2010. A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees. Mol. Phylogen. Evol. 55, 580–598.
Plunkett, G.M., Downie, S.R., 2000. Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. Syst. Bot. 25, 648–667.
Qu, X.J., Jin, J.J., Chaw, S.M., et al., 2017a. Multiple measures could alleviate long-branch attraction in phylogenomic reconstruction of Cupressoideae (Cupressaceae). Sci. Rep. 7, 41005.
Qu, X.J., Wu, C.S., Chaw, S.M., et al., 2017b. Insights into the existence of isomeric plastomes in Cupressoideae (Cupressaceae). Genome Biol. Evol. 9, 1110–1119.
Rendle, A.B., 1925. The Classification of Flowering Plants, vol. 2. Cambridge University Press, London.
Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33, W686–W689.
Smith, T.C., 2002. Chloroplast evolution: secondary symbiogenesis and multiple losses. Curr. Biol. 12, R62–R64.
Song, B., Li, F.Z., 2002. The utility of *trnK* intron 5′region in phylogenetic analysis of Ulmaceae sl. Acta Phytotax. Sin. 40, 125–132.
Song, B.H., Wang, X.Q., Li, F.Z., et al., 2001. Further evidence for paraphyly of the Celtidaceae from the chloroplast gene *matK*. Plant Syst. Evol. 228, 107–115.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analysis with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Swofford, D., 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA, version 4.
Sytsma, K.J., Morawetz, J., Pires, J.C., et al., 2002. Urticalean rosids: circumscription, rosid ancestry, and phylogenetics based on *rbcL, trnL-trnF*, and *ndhF* sequences. Am. J. Bot. 89, 1531–1546.
Tamura, K., Stecher, G., Peterson, D., et al., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
Terrab, A., Paun, O., Talavera, S., et al., 2006. Genetic diversity and population structure in natural populations of Moroccan Atlas cedar (*Cedrus atlantica*; Pinaceae) determined with cpSSR markers. Am. J. Bot. 93, 1274–1280.
Van Velzen, R., Bakker, F.T., Sattarian, A., et al., 2006. Evolutionary relationships of Celtidaceae (Dissertation). In: Sattarian, A. (Ed.), Contribution to the Biosystematics of *Celtis* L. (Celtidaceae) with Special Emphasis on the African Species. Wageningen University, Wageningen, The Netherlands, pp. 7–30.
Wang, H.C., Moore, M.J., Soltis, P.S., et al., 2009. Rosid radiation and the rapid rise of angiosperm dominated forests. Proc. Natl. Acad. Sci. U. S. A 106, 3853–3858.
Wang, R.J., Cheng, C.L., Chang, C.C., et al., 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol. Biol. 8, 36.
Wang, S., Shi, C., Gao, L.Z., 2013. Plastid genome sequence of a wild woody oil species, *Prinsepia utilis*, provides insights into evolutionary and mutational patterns of Rosaceae chloroplast genomes. PLoS One 8, e73946.
Wang, Y.H., Qu, X.J., Chen, S.Y., et al., 2017. Plastomes of Mimosoideae: structural and size variation, sequence divergence, and phylogenetic implication. Tree Genet. Genomes 13, 41.
Wang, Y.H., Wicke, S., Wang, H., et al., 2018. Plastid genome evolution in the early-diverging legume subfamily Cercidoideae (Fabaceae). Front. Plant Sci. 9, 138.
Wicke, S., Schneeweiss, G.M., Muller, K.F., et al., 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol. Biol. 76, 273–297.
Wilson, C.A., 2009. Phylogenetic relationships among the recognized series in *Iris* section *Limniris*. Syst. Bot. 34, 277–284.
Wilson, D., 1975. Plant remains from the Graveney boat and the early history of *Humulus lupulus* L. in W. Europe. New Phytol. 75, 627–648.
Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20, 3252–3255.
Wysocki, W.P., Clark, L.G., Attigala, L., et al., 2015. Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. BMC Evol. Biol. 15, 50.

Yang, J.B., Li, D.Z., Li, H.T., 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. Mol. Ecol. Resour. 14, 1024–1031.
Yang, M.Q., van Velzen, R., Bakker, T.F., et al., 2013. Molecular phylogenetics and character evolution of Cannabaceae. Taxon 62, 473–485.
Yesson, C., Russell, S.J., Parrish, T., et al., 2004. Phylogenetic framework for *Trema* (Celtidaceae). Plant Syst. Evol. 248, 85–109.
Zavada, M.S., Kim, M., 1996. Phylogenetic analysis of Ulmaceae. Plant Syst. Evol. 200, 13–20.
Zhang, S.D., Jin, J.J., Chen, S.Y., et al., 2017. Diversification of Rosaceae since the late Cretaceous based on plastid phylogenomics. New Phytol. 214, 1355–1367.
Zhang, S.D., Soltis, D.E., Yang, Y., et al., 2011a. Multi-gene analysis provides a well-supported phylogeny of Rosales. Mol. Phylogen. Evol. 60, 21–28.
Zhang, Y.J., Ma, P.F., Li, D.Z., 2011b. High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). PLoS One 6, e20596.
Zhang, T., Zeng, C.X., Yang, J.B., et al., 2016. Fifteen novel universal primer pairs for sequencing whole chloroplast genomes and a primer pair for nuclear ribosomal DNAs. J. Syst. Evol. 54, 219–227.
Zhu, A., Guo, W., Gupta, S., et al., 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 209, 1747–1756.