

Comparative Analysis of Complete Chloroplast Genomes of Eight *Ficus* Species and Insights Into Phylogenetic Relationships of *Ficus*

Xi Xia

xiaxi888@outlook.com

Southwest Forestry University

Jingyu Peng

China Agricultural University

Lin Yang

Southwest Forestry University

Xueli Zhao

Southwest Forestry University

Anan Duan

Southwest Forestry University

Dawei Wang

Southwest Forestry University

Research Article

Keywords: Chloroplast genome, *Ficus*, hypervariable regions, Phylogenetic relationship

Posted Date: January 25th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1268022/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Life on June 7th, 2022. See the published version at <https://doi.org/10.3390/life12060848>.

Abstract

Background

The genus *Ficus* is an evergreen plant, the most numerous species in the family Moraceae, and often used as food and pharmacy source. The phylogenetic relationships of the genus *Ficus* have been debated for many years.

Results

In this study, the eight *Ficus* species (*Ficus altissima*, *F. auriculata*, *F. benjamina*, *F. curtipes*, *F. heteromorpha*, *F. lyrata*, *F. microcarpa*, and *F. virens*) complete chloroplast(cp) genomes were successfully sequenced and phylogenetic analyses with other *Ficus* species. The result showed that the eight *Ficus* cp genomes ranged from 160,333 bp (*F. heteromorpha*) to 160,772 bp (*F. curtipes*), with a typical quadripartite structure. It was found that the eight *Ficus* cp genomes had similar genome structures, containing 127 unique genes. The cp genomes of eight *Ficus* species contained 89–104 SSR loci, which were dominated by mono-nucleotides repeats. Moreover, we identified eight hypervariable regions eight hypervariable regions (*trnS-GCU_trnG-UCC*, *trnT-GGU_psbD*, *trnV-UAC_trnM-CAU*, *clpP_psbB*, *ndhF_trnL-UAG*, *trnL-UAG_ccsA*, *ndhD_psaC* and *ycf1*). Phylogenetic analysis showed that three subgen. (*Sycomorus*, *Ficus*, *Pharmacosycea*) of *Ficus* as monophyletic, but *Urostigma* is non-monophyletic.

Conclusions

In our study, we have successfully assembled and analysed the cp genomes of eight *Ficus* species. The results give useful data for further studies on molecular identification, phylogeny, species identification and population genetics of speciation in the *Ficus*.

Background

Genus *Ficus* is the most numerous species of evergreen plants in the family Moraceae and is mainly found in subtropical and tropical areas [1]. For a long time, numerous *Ficus* species are utilized as sources for food and pharmacy because of their rich nutritional content [2]. The genus *Ficus* is one of the largest woody genera of angiosperms, and the subdivision of the genus has been concerned by taxonomists [3]. The present relatively complete taxonomic system for the genus *Ficus* was proposed by Berg [4], which divided the genus *Ficus* into six subgenera (subgenus *Urostigma*, subgenus *Pharmacosycea*, subgenus *Sycomorus*, subgenus *Ficus*, subgenus *Sycidium*, and subgenus *Synoecia*) based on previous molecular phylogenetic analyses and morphological characters. However, the use of morphological and molecular markers for the identification of *Ficus* genus are controversial or limited due to overlapping phenotypic characteristics or morphological similarity among *Ficus* genus [3, 5–7].

Therefore, providing more genomic information is imperative for the understanding of *Ficus* genus and the safe and effective utilization of *Ficus*.

The chloroplast (cp) is a unique organelle in plant cell that not only play a vital part in photosynthesis, but also participate in the synthesis of other organisms [8–10]. The cp genomes in plants are uniparentally inherited and consists of a highly conserved genomic structure a small single-copy (SSC), a couple of inverse-repeat (IR) regions, and a large single-copy (LSC), which also has a large number of variable loci [11, 12]. Therefore, cp genomes were often used to research the evolution, interspecific divergence, adaptive population history and phylogeny of related species [12–14]. With the coming of the era of whole-genome sequencing, the cp whole genomes are extensively used in phylogenetic differentiation studies, species taxonomic identification and genetic breeding [15–19]. Although the cp genomes of some genus *Ficus* have been reported previously [20–27], the eight *Ficus* species studied here have not been reported. As such, we compared the cp genome structures of eight *Ficus* species to provide a useful genomic resource for *Ficus* species.

In our study, the complete cp genomes of eight *Ficus* species were sequenced, which possess economically significant values (*Ficus altissima*, *Ficus auriculata*, *Ficus benjamina*, *Ficus curtipes*, *Ficus heteromorpha*, *Ficus lyrata*, *Ficus microcarpa*, and *Ficus virens*). The structural features and sequence differences of the eight *Ficus* species were then compared and analyzed, while the phylogeny of the eight *Ficus* species was inferred from reported cp genomes of the *Ficus* genus. The results of this study can provide guidance for the subsequent conservation and utilization of *Ficus* species.

Results

Genomic Characteristics of Chloroplast

The eight newly sequenced chloroplast genomes of the genus *Ficus* (*F. altissima*, *F. auriculata*, *F. benjamina*, *F. curtipes*, *F. heteromorpha*, *F. lyrata*, *F. microcarpa*, and *F. virens*) were typical quadripartite structure with a length range of 160,333 bp (*F. heteromorpha*) ~ 160,772 bp (*F. curtipes*) (Fig. 1). The LSC region of the eight *Ficus* species cp genomes ranged from 88,426 (*F. auriculata*) to 89,149 (*F. curtipes*) bp, the SSC region of 19,183 (*F. virens*) to 20,158 (*F. microcarpa*) bp, and IRb region of 25,743 (*F. curtipes*) to 26,249 (*F. virens*) bp (IRa: 25,743 to 26,249 bp) (Table S1). The cp genomes of all eight *Ficus* were uploaded to NCBI and obtained GenBank accession numbers (Table S2).

We annotated a total of 127 unique genes, containing 83 protein-coding genes, 8 rRNAs and 36 tRNAs, in eight *Ficus* species showing similar genomic structures. Nineteen of the genes contained one intron and three genes contained two introns (Table 1). In all eight *Ficus* species cp genomes were AT-rich, with 64.1% of the genomes made up of A/T nucleotides. The GC content of IR regions, LSC and SSC regions, protein-coding sequence, tRNA and rRNA among the eight *Ficus* cp genomes was similar, with 35.9%, 42.4–42.7%, 32.7%, 37.1–37.2%, 52.9% and 55.4% GC content respectively (Table S3).

Table 1

List of genes encoded by Eight species of *Ficus* chloroplast genome. (×2) indicates that the gene has two copies. * and ** indicate genes containing one / two introns.

Category	Gene groups	Gene names	number
Self-replication	Large subunit of ribosomal proteins	rpl14, rpl16*, rpl2 (×2)*, rpl20, rpl22, rpl23 (×2), rpl32, rpl33, rpl36	11
	Small subunit of ribosomal proteins	rps11, rps12**, rps14, rps15, rps16*, rps18, rps19, rps2, rps3, rps4, rps7 (×2), rps8	13
	DNA dependent RNA polymerase	rpoA, rpoB, rpoC1*, rpoC2	4
	ribosomal RNAs	rrn16 (×2), rrn23 (×2), rrn4.5 (×2), rrn5, (×2)	8
	transfer RNAs	trnA-UGC (×2)*, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC*, trnH-GUG, trnI-CAU (×2), trnI-GAU (×2)*, trnK-UUU*, trnL-CAA (×2), trnL-UAA*, 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*, trnW-CCA, trnY-GUA	36
Photosynthesis	Photosystem I	psaA, psaB, psaC, psal, psaJ	5
	Photosystem II	psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbM, psbN, psbT, psbZ	14
	NADP dehydrogenase	ndhA*, ndhB (×2)*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK	12
	Cytochrome b6/f complex	petA, petB*, petD*, petG, petL, petN	6
	ATP synthase	atpA, atpB, atpE, atpF*, atpH, atpI	6
	Rubisco	rbcL	1
Other genes	Maturase	matK	1
	Protease	clpP**	1
	Envelop membrane protein	cemA	1
	Subunit Acetyl-CoA-Carboxylate	accD	1
	c-type cytochrome synthesis gene	ccsA	1

Category	Gene groups	Gene names	number
Unknown	Conserved Open reading frames	<i>ycf1</i> (×2), <i>ycf2</i> , (×2), <i>ycf3**</i> , <i>ycf4</i>	6
			127

SSR and Repetitive Sequence Analysis

We detected 94, 91, 100, 89, 92, 101, 93 and 104 SSR loci in *Ficus altissima*, *Ficus auriculata*, *Ficus benjamina*, *Ficus curtipes*, *Ficus heteromorpha*, *Ficus lyrata*, *Ficus microcarpa* and *Ficus virens* cp genomes, respectively. Among them, the richest repeat type was mono-nucleotides (53-68), followed by di-nucleotides (17-20), tetra-nucleotides (8-10), tri-nucleotides (4-5), penta-nucleotides (2-4), and hexa-nucleotides (0-2), respectively. The A/T repeat motif accounted for a large proportion (52-67) among all the repeat types, followed by AT/AT (17-19), but the other repeat motifs were very rare. Mono-nucleotide repeat motifs (A/T), di-nucleotide repeat motifs (AT/TA), tri-nucleotide repeat motifs (AAG/CTT, AAT/ATT), tetra-nucleotide repeat motifs (AAAG/CTTT, AAAT/ATTT, AATT/AATT, AGAT/ATCT) and penta-nucleotides repeat motifs (AAAGG/CCTTT, AATTC/AATTG) were present in each cp genome (Fig. 2a, b; Table S4).

The analysis of the eight cp genomes using Reputer showed that a total of 512 repeats were identified. The number of repeats each species ranged from 57 (*Ficus altissima*) to 71 (*Ficus auriculata*), with palindromic repeats being the most abundant types (39-51), followed by forward repeats (13-19) and reverse (2-4) repeats, the complement repeats are rare in these 8 cp genomes (Fig. 2c; Table S5). The results of tandem repeats showed a total number of 341 tandem repeats for the eight cp genomes, ranging from 33 (*Ficus auriculata*) to 47 (*Ficus altissima* and *Ficus benjamina*) (Fig. 2d).

IR Constriction and Expansion

We performed a comparative analysis of the contraction and expansion of the IR/SC region boundary in eight *Ficus* species (Fig. 3). The genes at the junctions of the eight *Ficus* species included *rps19*, *ndhF*, *rpl22*, *rpl2*, *psbA*, *trnH*, and *ycf1*. The *ndhF* gene of *Ficus curtipes*, *Ficus heteromorpha*, and *Ficus lyrata* was located exclusively in the SSC region, ranging from 3-657 bp away from the boundary of JSB. The *ndhF* gene of *Ficus altissima*, *Ficus auriculata*, *Ficus benjamina*, *Ficus microcarpa* and *Ficus virens* was mainly located in the SSC region, but a small proportion of *ndhF* gene spanned the junction 15-521 bp into the IRb region. The *rps19* gene of *Ficus curtipes*, *Ficus lyrata*, and *Ficus virens* was located in the LSC region, which of *Ficus auriculata* was located in the IRb region, while the gene in *Ficus altissima*, *Ficus benjamina*, *Ficus heteromorpha* and *Ficus microcarpa* spans the LSC and IRb boundary. We also observed that the *ycf1* gene in all eight *Ficus* cp genomes spanned the SSC, IRa junction, but the majority were located in the SSC region, and the length of the *ycf1* gene in the IRa ranged from 1008-1493 bp. In the eight *Ficus* cp genomes, the *trnH*, *psbA*, and *rpl22* genes were located exclusively in the LSC region, and the distance between the *trnH* gene and JLA ranged from 46-169 bp.

Sequence Divergence Analysis

To determine sequence differences among the eight *Ficus* cp genomes, we used *Ficus religiosa* as a reference genome and compared them using mVISTA (Fig. 4). It is apparent from this Fig. that the eight *Ficus* cp genomes were highly conserved, but regions of divergence among intergenic were also found, such as *trnT-UGU_trnL-UAA*, *rps16_trnQ-UUG*, *trnT-GGU_psbD*, *petA_psbJ_rpoB_trnC-GCA*, *petN_psbM*, *atpB_rbcL*, and *rpl32_trnL-UAG*. Subsequently, we calculated the nucleotide variability (Pi), and the Pi value allows a better assessment of the degree of sequence divergence (Fig. 5). From the results, the degree of divergence in the LSC and SSC regions exceeded that of the IR region, and the average value of Pi in the eight *Ficus* species was about 0.0033. Then, $Pi > 0.0128$ was identified as the hotspot region, and a total of eight high variance regions were found: *trnS-GCU_trnG-UCC*, *trnT-GGU_psbD*, *trnV-UAC_trnM-CAU*, *clpP_psbB*, *ndhF_trnL-UAG*, *trnL-UAG_ccsA*, *ndhD_psaC* and *ycf1*, where the *ndhF_trnL-UAG* region has the highest Pi value (0.0202). The identification of these highly variable regions offers invaluable information for the evolution of markers in *Ficus*.

Non-synonymous (Ka) and synonymous (Ks) substitution rate analysis

We calculated Ka/Ks ratios of the eight *Ficus* chloroplast genomes using *Ficus religiosa* as the reference with a total of 53 protein genes reselected based on 300 bp length (Fig.6; Table S6). Ka/Ks ratio results indicate that gene-specificity is the majority, and the remaining small portion is region-specific (the IR, SSC, and LSC regions showed comparable values). The Ka range of eight *Ficus* chloroplast genomes was from 0 to 2.6149, and the Ks values ranged from 0 to 5.0893. The highest Ka and Ks was for the *ycf1* gene between *Ficus altissima* and *Ficus religiosa* ($ka=2.6149$, $ks=5.0893$), while the values of Ka and Ks of a large number of genes are equal to 0. The average Ka/Ks ratio for the 53 protein genes was 0.1211 and the average Ka/Ks value for *ycf1* was 1.0476, with only the *ycf1* gene having a Ka/Ks value greater than 1 among genes tested. Fig.6 The Ka/Ks ratio of 53 protein-coding genes of eight cp genomes for comparison with *Ficus religiosa*.

Phylogenomic Analysis

Phylogenetic analysis of 28 species revealed generally high support values for both phylogenetic methods (ML/BI), yielding nearly identical topologies, and strong support for monophyly in *Ficus* (Fig.7). *Ficus* was divided into five main branches, subgen. *Pharmacosycea*, subgen. *Ficus*, subgen. *Urostigma*, subgen. *Sycomorus*, and *F. lyrata* in subgen. *Urostigma*. Eight species of the subgen. *Urostigma* formed a monophyletic cluster with high support (Cluster I; BS = 100 for ML; PP = 1.00 for BI). Cluster II (BS = 100 for ML; PP = 1.00 for BI) includes four species of the subgen. *Sycomorus*. The Cluster III (BS = 76 for ML; PP = 1.00 for BI) is the subgen. *Ficus*, which is composed of eleven species. One species of the subgen. *Urostigma* (*F. lyrata*) formed Cluster III (BS = 63 for ML; PP = 1.00 for BI); The Cluster IV (BS = 100 for ML, PP = 1.00 for BI) contained two species of the subgen. *Pharmacosycea*. All of the 26 *Ficus* taxa formed a monophyletic clade with 100% bootstrap value or the Bayesian posterior probability.

Discussion

Chloroplast genomic DNA structures

In the present study, we sequenced the eight *Ficus* species cp genomes, which laid the foundation for a comprehensive comparison of the cp genome sequences of Moraceae. The eight *Ficus* species reported in this study differ somewhat in chloroplast size, but in general are roughly similar in chloroplast genome length to those reported so far for *Ficus* species, all around 160,000 bp in length [8, 27]. The structure of the eight *Ficus* cp genomes, like most angiosperms, also has a typical quadripartite structure [8, 28–30]. Moreover, we found that the content of eight *Ficus* cp genomes was dominated by A/T, up to 64.1%, and this result was similar to other plants [31–33]. The lower G/C content of the *Ficus* genomes may be related to its own spontaneous mutations [34]. We also found that the GC content of LSC and SSC were lower than IR regions, which were common in *Ficus* species [27, 35, 36], it may be because of the existence of rRNA and tRNA genes in the IR region [37].

Identification of SSRs and repeat sequences

Repeated sequences have been shown to be associated with rearrangement, evolution and divergence of cp genome sequences, as well as having an important function in the phylogeny [12, 38, 39]. The number and distribution of four repeats in the eight *Ficus* cp genomes, and the distribution of repeat were similar in the cp genomes of *Ficus*, which was strongly related to the high conserved in *Ficus* species. The cpSSR markers have been frequently used for intraspecific identification and genetic evolutionary analysis because of their high mutation rate [40–42]. The results of this study indicated that the mononucleotide repeats were the richer repeat type in eight *Ficus* species, with the A/T as the enriched repeat motif, similar to other higher plant studies [13, 43], indicating that the higher plants were dominated by low-level repeat motif [44]. The SSRs and repeats detected in this study were important for the subsequent phylogenetic studies and taxonomy of genus *Ficus* and even the family Moraceae.

IR Constriction and Expansion

Scientists have noticed that the expansion and contraction of the IR regions affects the size of chloroplasts [39, 45]. Moreover, it may cause border genes to enter the IR or SC [46]. In our study, most of the genes (*rpl22*, *rpl2*, *ycf1*, *trnH*, *psbA*) in the eight *Ficus* species were identical at the IR/SC boundary positions, and differed only in length, further suggested that *Ficus* species were closely related and have highly conserved cp genomes. The cp genomes of monocots and dicots have distinct differences in gene arrangement, for example, the *trnH* gene was located in the LSC region in dicots, whereas it was located in the IR region in monocots [47]. Interestingly, we found that the *rps19* gene of *F. auriculata* was located exclusively on the IRb region, which would be very rare in dicotyledons [48–50] and maybe because of expansion from the IR region to the LSC region.

Sequence Divergence Analysis

Comparative analysis of hypervariable regions in plant cp genomes can better identify mutation hotspots and provide a basis for genetic diversity analysis [51, 52]. In this study, to identify variable regions, we

calculated the percentage of variable characters (coding and non-coding regions) in eight *Ficus* species cp genomes. The results of the study were similar to those of angiosperms, where the non-coding regions were much more variable than the coding regions [53–55]. At the same time, we also found that the SC region of the eight *Ficus* species in this study was more variable than the IR region, which was consistent with previous results [56, 57]. In previous studies, a large number of hypervariable regions were used as DNA barcodes for other plants [58–60]. We identified eight hypervariable regions in the *Ficus* cp genomes (*trnS-GCU_trnG-UCC*, *trnT-GU_psbD*, *trnV-UAC_trnM-CAU*, *clpP_psbB*, *ndhF_trnL-UAG*, *trnL-UAG_ccsA*, *ndhD_psaC*, and *ycf1*), and these hypervariable regions could be used as identification of the genus *Ficus*, as well as for phylogenetic analysis of the genus *Ficus* or as candidate markers for population genetics.

Non-synonymous (Ka) and synonymous (Ks) substitution rate analysis

The Ka/Ks reveal selection pressure on protein-coding genes and have a key role in evolutionary studies [61]. Non-synonymous nucleotide substitutions occurred with lower frequency than synonymous substitutions in most genes, subject to purifying selection [48]. Thus, the ratio of Ka/Ks > 1 indicates probable positive selection; Ka/Ks < 1 indicates purifying selection, while Ka/Ks values nearly 1 indicate neutral evolution [62]. The results of this study showed that a great many genes had Ka and Ks values equal to 0, indicating that these eight *Ficus* chloroplast genomes were relatively conservative. However, the highest Ka and Ks was for the gene between *Ficus altissima* and *Ficus religiosa* (ka=2.6149, ks=5.0893), suggested that the is more variable in *Ficus altissima*. The average Ka/Ks value of *ycf1* was 1.0476, and this gene probably has a vital role to play for the adaptation of *Ficus* to different environments [63]. The Ka/Ks values were less than 1 for all genes except *ycf1*, which indicated that these cp genomes underwent extensive purifying selection.

Phylogenomic Analysis

The cp genome was the most vital genetic resource for inferring plant evolutionary relationships and was well suited for analyzing comparative plant relatedness [64, 65]. In this study, in order to determined the phylogenetic relationships of eight *Ficus* genus, phylogenetic trees were structured using ML and BI methods. The findings indicated that all *Ficus* species clustered in the one clade and high BS values, a finding that is consistent with the phylogeny inferred by previous work using cpDNA fragments [8].

According to the results of the phylogenetic analysis, species of the three subgenera (Subgen. *Sycomorus* (Gasp.) Miq., Subgen. *Ficus* and Subgen. *Pharmacosycea* Miq.) were all monophyletic taxa and had strong support values, suggesting that the cp genome was suited to resolve the phylogeny of the *Ficus*. Rønsted's [66] results indicated that *Urostigma* is of non-monophyletic origin, and similar results were obtained in our ML and BI trees. Based on morphological traits [67] as well as molecular data [6], it was previously suggested that *Ficus tikoua* should be classified as Subgen. *Sycomorus* (Gasp.) Miq., and our phylogenetic analysis provides strong evidence for this classification. Subgen. *Urostigma* (Gasp.) Miq. and Subgen. *Sycomorus* (Gasp.) Miq. formed a clade, indicating that the two genera are closely related. Zhang [68] points out that *F. auriculata* and *F. beipeiensis* should be classified

as the same species, but our study results clearly do not support this classification. In our result, *F. auriculata* and *F. beipeiensis* were still somewhat different, despite being more closely related. It was noteworthy that both ML and BI trees exhibit similar topologies, confirm that cp genome data would be useful for studying intraspecific relationships in the genus *Ficus*.

Conclusion

In our study, we assembled and analyzed the cp genomes of eight *Ficus* species. By comparing these cp genomes, the results showed the structure of these cp genomes is similar to that of most angiosperm genomes, which exhibit a typical tetragonal structure. Also, we obtained valuable genetic resources, including SSRs, highly variable loci, and repetitive sequences. Phylogenetic analysis shows that all *Ficus* species were clustered in the same clade. The present study provides a clearer phylogenetic framework for the taxonomy of the genus *Ficus*, it will help in the species identification of these species, genetic diversity studies.

Methods

Plant materials and DNA extraction

Eight taxa, *Ficus altissima*, *Ficus auriculata*, *Ficus benjamina*, *Ficus curtipes*, *Ficus heteromorpha*, *Ficus lyrata*, *Ficus microcarpa*, and *Ficus virens*, representing the genus *Ficus* of the family Moraceae. We collected young, fresh, and non-diseased leaves from target species of adult plants, which were then put into liquid nitrogen frozen. Besides, the voucher specimens were preserved in the herbarium of Southwest Forestry University, Kunming, Yunnan, China (Table S2). DNA extraction Using a Modified CTAB Method [69].

Genome sequencing, assembly and annotation

Sequencing was performed by Annoroad Gene Technology (Beijing, China) to generate libraries with an average insert size of 400 bp from total DNA on an Illumina-based (Illumina Novaseq 6000). Then, the raw data obtained by sequencing were assembled using GetOrganelle v1.6.0 software with *Ficus religiosa* as the reference [70]. The assembled sequences were annotated with Geneious prime using *Ficus religiosa* as a reference and manually adjusted for start and stop codons. Genome mapping of the annotated *Ficus* cp genomes by OGdraw online [71]. The GenBank accession numbers of the eight *Ficus* species are shown in Table S2.

Simple sequence repeats (SSR) and repetitive sequence analysis

Identification of SSRs by using MISA [72], and the specific parameter settings were referred to Zhao *et al* (2020). The forward (F), reverse (R), complementary (C), and palindromic (P) oligonucleotide repeats were determined by the REPuter program [73], and then the tandem repeats were subdivided using a

network-based Tandem Repeat Finding [74], and the REPuter and Tandem Repeat specific parameters were set as in Yang *et al.* [75]

Genome comparison and divergent hotspots identification

The alignment of eight *Ficus* cp genome sequences of species was performed using MAFFT v7 [76], and the completed sequences were subsequently compared using the Shuffle-LAGAN mode in mVISTA [77].

Non-synonymous (Ka) and synonymous (Ks) substitution rate analysis

First, selected genes (length >300 bp) were paired in MAFFT. Then, the Ka/Ks value of the screened genes was calculated using the KaKs_calculator [78] with reference to Ivanova *et al.* [79].

Phylogenetic Analysis

To estimate phylogenetic relationships of eight *Ficus* species in the genus *Ficus*, 28 cp genomes taxa were compared (Table S1), containing 26 *Ficus* species, and 2 *Artocarpus* species (*Artocarpus camansi* and *Artocarpus altilis*) were set as outgroups. Sequence alignment of 28 cp genomic matrices using MAFFT v7 software. To obtain robust phylogenetic relationships, we performed maximum likelihood (ML) and Bayesian inference (BI) analyses. The ML phylogenetic tree was constructed using IQ-tree 1.5.5 [80] and the BI phylogenetic tree was constructed using MrBayes 3.1.2 [81]. Both ML trees and BI trees are visualized by using FigTree v1.4.0.

Declarations

Data availability

The chloroplast genomes sequences of eight *Ficus* species (*F. altissima*, *F. auriculata*, *F. benjamina*, *F. curtipes*, *F. heteromorpha*, *F. lyrata*, *F. microcarpa*, *F. virens*) have been uploaded to NCBI and the GenBank accession numbers are (MW043478, MW039147, MW039146, MW039143, MW039145, MW039148–MW039144, MW039142).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (No. 32060350) and the Forestry Science and Technology Development Project of National Forestry and Grassland Administration of China (No. KJZXSA2019036).

Authors' contributions

XX and JYP Wrote the manuscript and data curation, LY and AAD collected plant materials, XLZ interpreted the results, DWW conceived the study, funding acquisition. All authors reviewed the manuscript.

Ethical Requirement

This article does not contain any studies with human or animal subjects performed by any of the authors.

SUPPLEMENTARY MATERIALS

Table S1: Geographic locations of the eight *Ficus* species analyzed in this study, Table S2: List of samples voucher collection information and GenBank accession, Table S3: Base composition in the chloroplast genomes of *Ficus* species, Table S4: Types and numbers of SSRs motifs in eight *Ficus* cp genome, Table S5: Repeats in the chloroplast genome of eight *Ficus* species.

References

1. Badgujar SB, Patel VV, Bandivdekar AH, Mahajan RT: Traditional uses, phytochemistry and pharmacology of *Ficus carica*: a review. *Pharm Biol* 2014, 52(11):1487–1503.
2. Tkachenko G, Buyun L, Osadovskyy Z, Truhan M, Goncharenko V: *In vitro* screening of antimicrobial activity of ethanolic extract obtained from *Ficus lyrata* warb. (moraceae) leaves. *Agroecological Journal* 2016, 2:155–160.
3. Zhang Z, Wang XM, Liao S, Zhang JH, Li HQ: Phylogenetic reconstruction of *Ficus* subg. Synoecia and its allies (Moraceae), with implications on the origin of the climbing habit. *Taxon* 2020, 00(248):1–19.
4. Berg CC: Flora Malesiana precursor for the treatment of Moraceae 1: The main subdivision of *Ficus*: the subgenera. *Blumea* 2003, 48(1):166–177.
5. Xu L, Harrison RD, Yang P, Yang DR: New insight into the phylogenetic and biogeographic history of genus *Ficus*: Vicariance played a relatively minor role compared with ecological opportunity and dispersal. *J Syst Evol* 2011, 49(6):546–557.
6. Li HM, Wang S, Chen JY, Gui PJ: Molecular phylogeny of *Ficus* section *Ficus* in China based on four DNA regions. *Journal of Systematics Evolution* 2012, 50(5):422–432.
7. Fatihah HN, Mat N, Zaimah AR, Zuhailah MN, Norhaslinda H, Khairil M, Ghani AY, Ali AM: Morphological phylogenetic analysis of seven varieties of *Ficus deltoidea* Jack from the Malay Peninsula of Malaysia. *PLoS One* 2012, 7(12):e52441.

8. Bruun-Lund S, Clement WL, Kjellberg F, Rønsted N: First plastid phylogenomic study reveals potential cyto-nuclear discordance in the evolutionary history of *Ficus* L. (Moraceae). *Molecular phylogenetics and evolution* 2017, 109:93-104.
9. Neuhaus HE, Emes MJ: Nonphotosynthetic metabolism in plastids. *Annu Rev Plant Biol* 2000, 51:111–140.
10. Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, Yao H, Sun C, Li X, Li C *et al*: The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One* 2013, 8(2):e57607.
11. Hu S, Sablok G, Wang B, Qu D, Barbaro E, Viola R, Li M, Varotto C: Plastome organization and evolution of chloroplast genes in *Cardamine* species adapted to contrasting habitats. *BMC genomics* 2015, 16(1):306.
12. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D: The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant molecular biology* 2011, 76(3-5):273–297.
13. Hong Z, Wu Z, Zhao K, Yang Z, Zhang N, Guo J, Tembrock LR, Xu D: Comparative Analyses of Five Complete Chloroplast Genomes from the Genus *Pterocarpus* (Fabaceae). *Int J Mol Sci* 2020, 21(11):3758.
14. Zhong B, Xi Z, Goremykin VV, Fong R, McLenachan PA, Novis PM, Davis CC, Penny D: Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes. *Molecular biology and evolution* 2014, 31(1):177–183.
15. Alves-Pereira A, Clement CR, Picanço-Rodrigues D, Veasey EA, Dequigiovanni G, Ramos SLF, Pinheiro JB, Zucchi MI: Patterns of nuclear and chloroplast genetic diversity and structure of manioc along major Brazilian Amazonian rivers. *Annals of botany* 2018, 121(4):625–639.
16. Deng YW, Luo YY, He Y, Qin XS, Li CG, Deng XM: Complete Chloroplast Genome of *Michelia shiluensis* and a Comparative Analysis with Four Magnoliaceae Species. *Forests* 2020, 11(3).
17. Jiang L, Li M, Zhao F, Chu S, Zha L, Xu T, Peng H, Zhang W: Molecular Identification and Taxonomic Implication of Herbal Species in Genus *Corydalis* (Papaveraceae). *Molecules* 2018, 23(6).
18. Sun J, Wang Y, Liu Y, Xu C, Yuan Q, Guo L, Huang L: Evolutionary and phylogenetic aspects of the chloroplast genome of *Chaenomeles* species. *Scientific reports* 2020, 10(1):11466.
19. Yang C, Ma Y, Cheng B, Zhou L, Yu C, Luo L, Pan H, Zhang Q: Molecular Evidence for Hybrid Origin and Phenotypic Variation of *Rosa Section Chinenses*. *Genes* 2020, 11(9).
20. Chen H, Liu Q, Tang L: The plastid genome of winter cropping plants *Ficus tinctoria* (Moraceae). *Mitochondrial DNA B-Resources* 2020, 5(3):2703–2704.
21. Chi W, Chi LJ, Pan HW, Xia XL, Huang ZH: Characteristic Analysis of the Complete Chloroplast Genome of *Ficus concinna* Miq. var. *subsessilis* Corner -An Ancient Tree from Zhejiang Province, China. *Journal of Hainan Normal University (Natural Science)* 2020, 33(01):50–57.
22. Han F, Li J, Yuan J, Yu J: The complete chloroplast genome sequence of *Ficus beipeiensis* (Moraceae), an endemic and endangered plant in China. *Mitochondrial DNA B-Resources* 2021, 6(2):604–605.

23. Liu Y, Chen W, Li F, Li C, Xie X, Chao Z, Tian E: The complete chloroplast genome sequence of *Ficus hirta* (Moraceae). Mitochondrial DNA Part B, Resources 2019, 4(2):4041–4042.
24. Mao Q, Bi G: Complete chloroplast genome of *Ficus racemosa* (Moraceae). Mitochondrial DNA 2016, 27(6):4655–4656.
25. Wang J, Wang TY, Wang LY, Zhang JG, Zeng YF: Assembling and Analysis of the Whole Chloroplast Genome Sequence of *Elaeagnus angustifolia* and Its Codon Usage Bias. Acta Botanica Boreali-Occidentalia Sinica 2019, 39(09):1559–1572.
26. Xu S, Guo S, Fan D, Wang J: The complete chloroplast genome sequence of *Ficus formosana* Maxim (Moraceae) from Guangzhou, China. Mitochondrial DNA B-Resources 2021, 6(7):1895–1896.
27. Yu H, Nason JD: Nuclear and chloroplast DNA phylogeography of *Ficus hirta*: obligate pollination mutualism and constraints on range expansion in response to climate change. New phytologist 2013, 197(1):276–289.
28. Gu C, Tembrock LR, Zheng S, Wu Z: The Complete Chloroplast Genome of *Catha edulis*: A Comparative Analysis of Genome Features with Related Species. Int J Mol Sci 2018, 19(2):525.
29. Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK *et al*: Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proceedings of the National Academy of Sciences of the United States of America 2007, 104(49):19369–19374.
30. Ng WL, Cai Y, Wu W, Zhou R: The complete chloroplast genome sequence of *Melastoma candidum* (Melastomataceae). Mitochondrial DNA Part B, Resources 2017, 2(1):242–243.
31. Fu J, Liu H, Hu J, Liang Y, Liang J, Wuyun T, Tan X: Five Complete Chloroplast Genome Sequences from *Diospyros*: Genome Organization and Comparative Analysis. PLoS One 2016, 11(7):e0159566.
32. Mariotti R, Cultrera NG, Díez CM, Baldoni L, Rubini A: Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. BMC plant biology 2010, 10:211.
33. Yi DK, Kim KJ: Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. PLoS One 2012, 7(5):e35872.
34. Schmid P, Flegel WA: Codon usage in vertebrates is associated with a low risk of acquiring nonsense mutations. Journal of translational medicine 2011, 9:87.
35. Ghada, Baraket, and, Saddoud, Olfa, and, Chatti, Khaled, and, Mars: Chloroplast DNA analysis in Tunisian fig cultivars (*Ficus carica* L.): Sequence variations of the trnL-trnF intergenic spacer. Biochemical Systematics Ecology 2008, 36(11):828–835.
36. Wang J, Shi D, Bai Y, Liu Y: High-throughput sequencing uncover *Ficus tikoua* Bur. chloroplast genome. Journal of Plant Biochemistry Biotechnology 2019(9).
37. Doorduyn L, Gravendeel B, Lammers Y, Ariyurek Y, Chin AWT, Vrieling K: The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. DNA research: an international journal for rapid publication of reports on genes and genomes 2011, 18(2):93–105.

38. Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki K: Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. *DNA research* 2004, 11(2):93–99.
39. Weng ML, Blazier JC, Govindu M, Jansen RK: Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Molecular biology and evolution* 2014, 31(3):645–659.
40. Gómez A, González-Martínez SC, Collada C, Climent J, Gil L: Complex population genetic structure in the endemic Canary Island pine revealed using chloroplast microsatellite markers. *TAG Theoretical and applied genetics Theoretische und angewandte Genetik* 2003, 107(6):1123–1131.
41. Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, Mei W, Thomas PI, Gardner MF, Ennos RA *et al*: Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? *Molecular ecology resources* 2015, 15(5):1067–1078.
42. Urbaniak L, Wojnicka-Pótorak A, Celiński K, Lesiczka P, Pawlaczyk E, Aučina A: Genetic resources of relict populations of *Pinus sylvestris* (L.) in Western Carpathians assessed by chloroplast microsatellites. *Biologia* 2019, 74(9):1077–1086.
43. Zhou T, Zhu H, Wang J, Xu Y, Xu F, Wang X: Complete chloroplast genome sequence determination of *Rheum* species and comparative chloroplast genomics for the members of Rumiceae. *Plant cell reports* 2020, 39(6):811–824.
44. Zhang DF, Li SX, Xia T: Characterization of EST-SSR Among Six Genera of Rosaceae. *Bulletin of Botanical Research* 2014, 34(06):810–815.
45. He L, Qian J, Li X, Sun Z, Xu X, Chen S: Complete Chloroplast Genome of Medicinal Plant *Lonicera japonica*: Genome Rearrangement, Intron Gain and Loss, and Implications for Phylogenetic Studies. *Molecules* 2017, 22(2).
46. Wang W, Chen S, Zhang X: Whole-Genome Comparison Reveals Divergent IR Borders and Mutation Hotspots in Chloroplast Genomes of Herbaceous Bamboos (Bambusoideae: Olyreae). *Molecules* 2018, 23(7):1537.
47. Huotari T, Korpelainen H: Complete chloroplast genome sequence of *Elodea canadensis* and comparative analyses with other monocot plastid genome. *Gene* 2012, 508:96–105.
48. Du X, Zeng T, Feng Q, Hu L, Zhu B: The Complete Chloroplast Genome Sequence of Yellow Mustard (*Sinapis alba* L.) and Its Phylogenetic Relationship to Other Brassicaceae Species. *Gene* 2020, 731:144340.
49. Lin CP, Wu CS, Huang YY, Chaw SM: The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. *Genome biology and evolution* 2012, 4(3):374–381.
50. Zhang Y, Zhang JW, Yang Y, Li XN: Structural and comparative analysis of the complete chloroplast genome of a mangrove plant: *Scyphiphora hydrophyllacea* Gaertn. f. and related Rubiaceae species. *Forest* 2019, 10(11).
51. Abdullah, Mehmood F, Shahzadi I, Waseem S, Mirza B, Ahmed I, Waheed MT: Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): Comparative analyses and identification of mutational

- hotspots. *Genomics* 2020, 112(1):581–591.
52. Kuang DY, Wu H, Wang YL, Gao LM, Zhang SZ, Lu L: Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome* 2011, 54(8):663-673.
 53. Clegg MT, Gaut BS, Learn GH, Jr., Morton BR: Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America* 1994, 91(15):6795–6801.
 54. Huang H, Shi C, Liu Y, Mao SY, Gao LZ: Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC evolutionary biology* 2014, 14:151.
 55. Perry AS, Wolfe KH: Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *Journal of molecular evolution* 2002, 55(5):501–508.
 56. Dong W, Liu H, Xu C, Zuo Y, Chen Z, Zhou S: A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. *BMC genetics* 2014, 15:138.
 57. Shen X, Wu M, Liao B, Liu Z, Bai R, Xiao S, Li X, Zhang B, Xu J, Chen S: Complete Chloroplast Genome Sequence and Phylogenetic Analysis of the Medicinal Plant *Artemisia annua*. *Molecules* 2017, 22(8):1330.
 58. Hu CY, Zheng XY, Teng YW: Characterization and phylogenetic utility of non-coding chloroplast regions trnL-trnF and accD-psaI in *Pyrus*. *Acta Horticulturae Sinica* 2011, 38(12):2261–2272.
 59. Scarcelli N, Barnaud A, Eiserhardt W, Treier UA, Seveno M, D'Anfray A, Vigouroux Y, Pintaud JC: A Set of 100 Chloroplast DNA Primer Pairs to Study Population Genetics and Phylogeny in Monocotyledons. *Plos One* 2011, 6(5):e19954.
 60. Shen JS, Li XQ, Zhu XT, Huang XL, Jin SH: The Complete Plastid Genome of *Rhododendron pulchrum* and Comparative Genetic Analysis of Ericaceae Species. *Forests* 2020, 11(2):158.
 61. Nazareno AG, Carlsen M, Lohmann LG: Complete Chloroplast Genome of *Tanaecium tetragonolobum*: The First Bignoniaceae Plastome. *PLoS One* 2015, 10(6):e0129930.
 62. Yang Z, Nielsen R: Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular biology and evolution* 2000, 17(1):32–43.
 63. Zeb U, Dong WL, Zhang TT, Wang RN, Li ZH: Comparative plastid genomics of *Pinus* species: Insights into sequence variations and phylogenetic relationships. *Journal of Systematics and Evolution* 2019, 58(2):118–132.
 64. Zhang Y, Wang Z, Guo Y, Chen S, Xu X, Wang R: Complete chloroplast genomes of *Leptodermis scabrida* complex: Comparative genomic analyses and phylogenetic relationships. *Gene* 2021, 791:145715.
 65. Kim GB, Lim CE, Kim JS, Kim K, Lee JH, Yu HJ, Mun JH: Comparative chloroplast genome analysis of *Artemisia* (Asteraceae) in East Asia: insights into evolutionary divergence and phylogenomic implications. *BMC genomics* 2020, 21(1):415.

66. Rønsted N, Weiblen GD, Cook JM, Salamin N, Machado CA, Savolainen V: 60 million years of co-divergence in the fig-wasp symbiosis. *Proceedings Biological sciences* 2005, 272(1581):2593-2599.
67. Gui P: Morphological and Anatomical Studies on *Ficus* subg. *Ficus* (Moraceae) in China. East China Normal University 2013:66.
68. Zhang LF: Phylogeny and taxonomy of the *Ficus auriculata* Complex. East China Normal University 2016:40–42.
69. Yang JB, Li DZ, Li HT: Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Molecular ecology resources* 2014, 14(5):1024–1031.
70. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ: GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome biology* 2020, 21(1):241.
71. Lohse M, Drechsel O, Bock R: OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current genetics* 2007, 52(5-6):267–274.
72. Beier S, Thiel T, Münch T, Scholz U, Mascher M: MISA-web: a web server for microsatellite prediction. *Bioinformatics (Oxford, England)* 2017, 33(16):2583–2585.
73. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R: REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic acids research* 2001, 29(22):4633–4642.
74. Benson G: Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* 1999, 27(2):573–580.
75. Yang Z, Wang G, Ma Q, Ma W, Liang L, Zhao T: The complete chloroplast genomes of three Betulaceae species: implications for molecular phylogeny and historical biogeography. *PeerJ* 2019, 7:e6320.
76. Katoh K, Standley DM: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 2013, 30(4):772–780.
77. Dubchak I, Ryaboy DV: VISTA family of computational tools for comparative analysis of DNA sequences and whole genomes. *Methods in molecular biology (Clifton, NJ)* 2006, 338:69–89.
78. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J: KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, proteomics & bioinformatics* 2010, 8(1):77–80.
79. Ivanova Z, Sablok G, Daskalova E, Zahmanova G, Apostolova E, Yahubyan G, Baev V: Chloroplast Genome Analysis of Resurrection Tertiary Relict *Haberlea rhodopensis* Highlights Genes Important for Desiccation Stress Response. *Frontiers in plant science* 2017, 8:204.
80. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ: IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* 2015, 32(1):268–274.
81. Ronquist F, Huelsenbeck JP: MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)* 2003, 19(12):1572–1574.

Figures

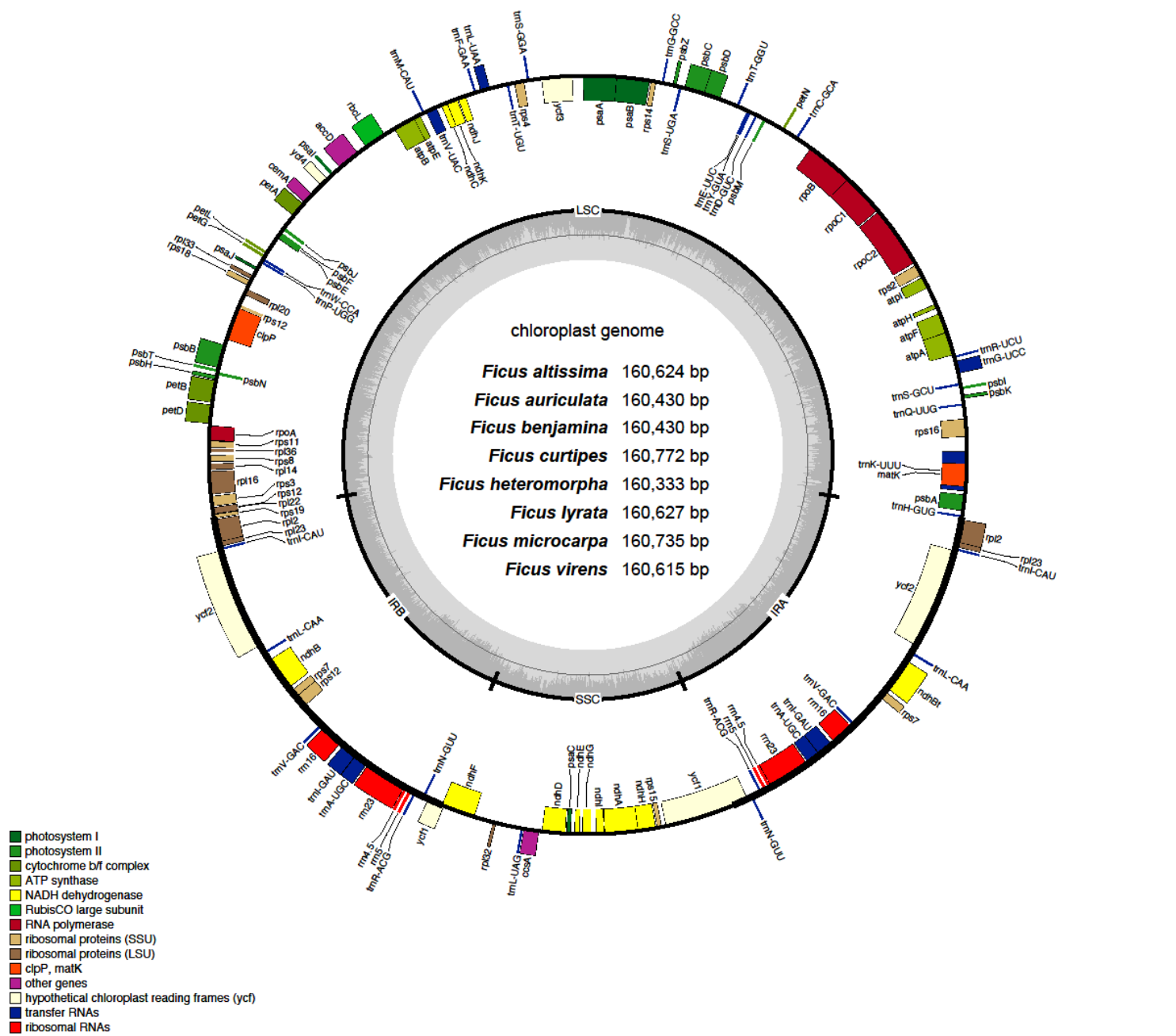


Figure 1

The complete chloroplast genome of eight *Ficus* species. Genes shown outside the circle are transcribed clockwise and those inside counterclockwise. Different colors represent different kinds of functional genes. The dark gray area and light gray area of inner circle represent the GC content to AT content of the genome respectively.

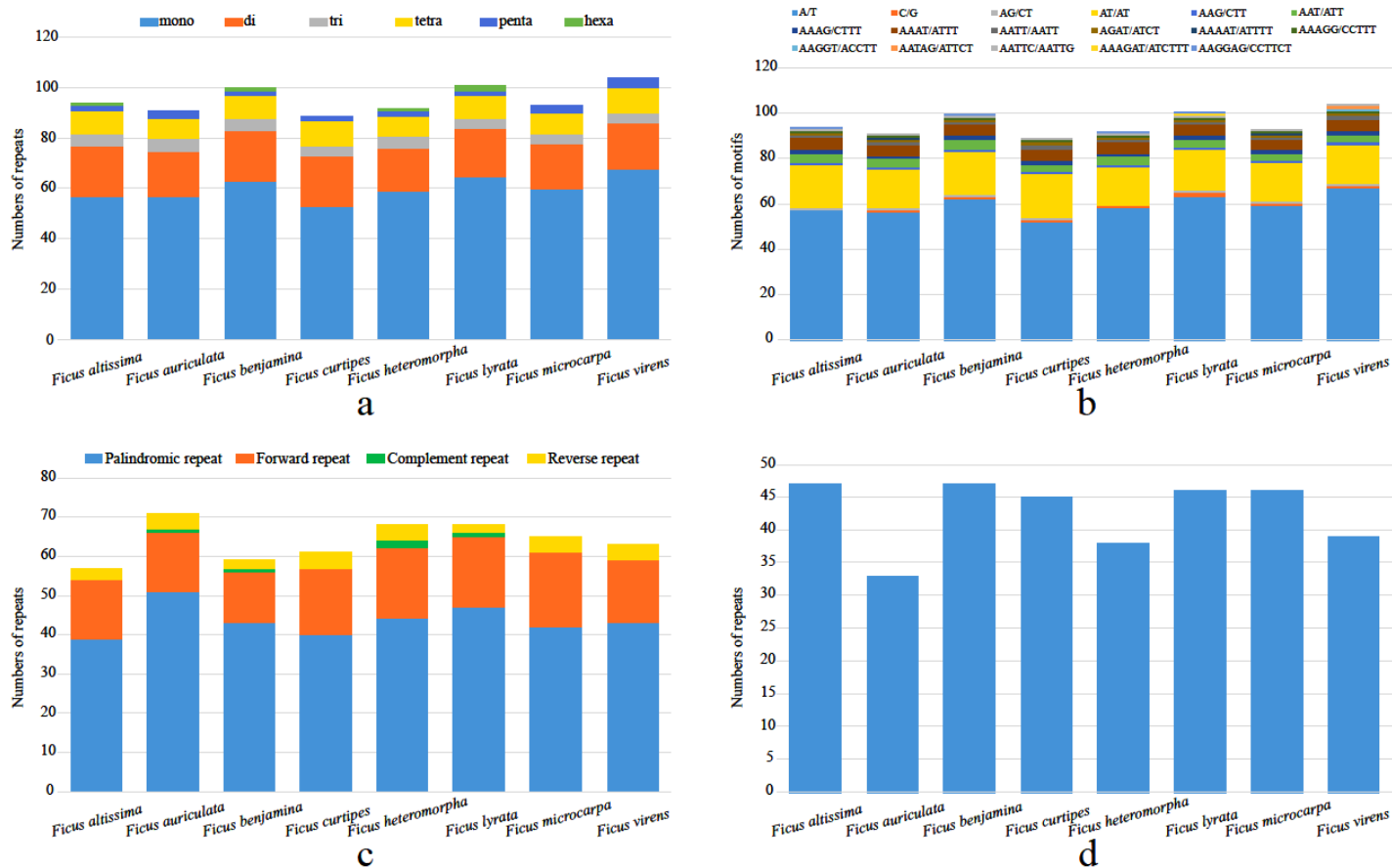


Figure 2

The number and type of simple sequence repeats (SSRs) and repeat sequences in the chloroplast genomes of eight *Ficus* species. (A) Frequency of eight SSR types, (B) Frequency of SSR motifs in different repeat class types, (C) Frequency of five repeat types, (D) Frequency of dispersed repeat sequences by length.

Inverted Repeats

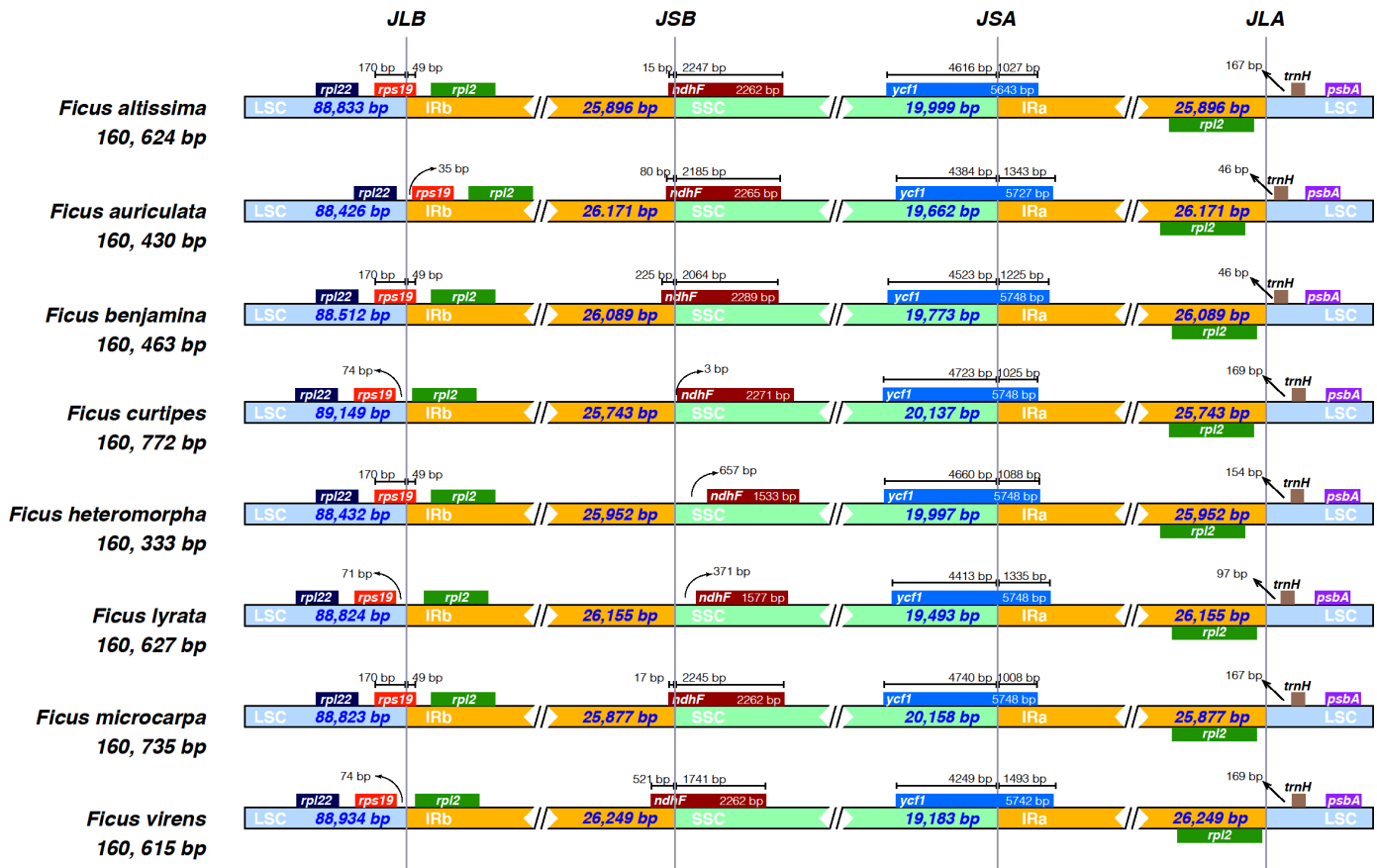


Figure 3

Comparison of the borders of LSC, IR and SSC in the chloroplast genomes of eight *Ficus* species.

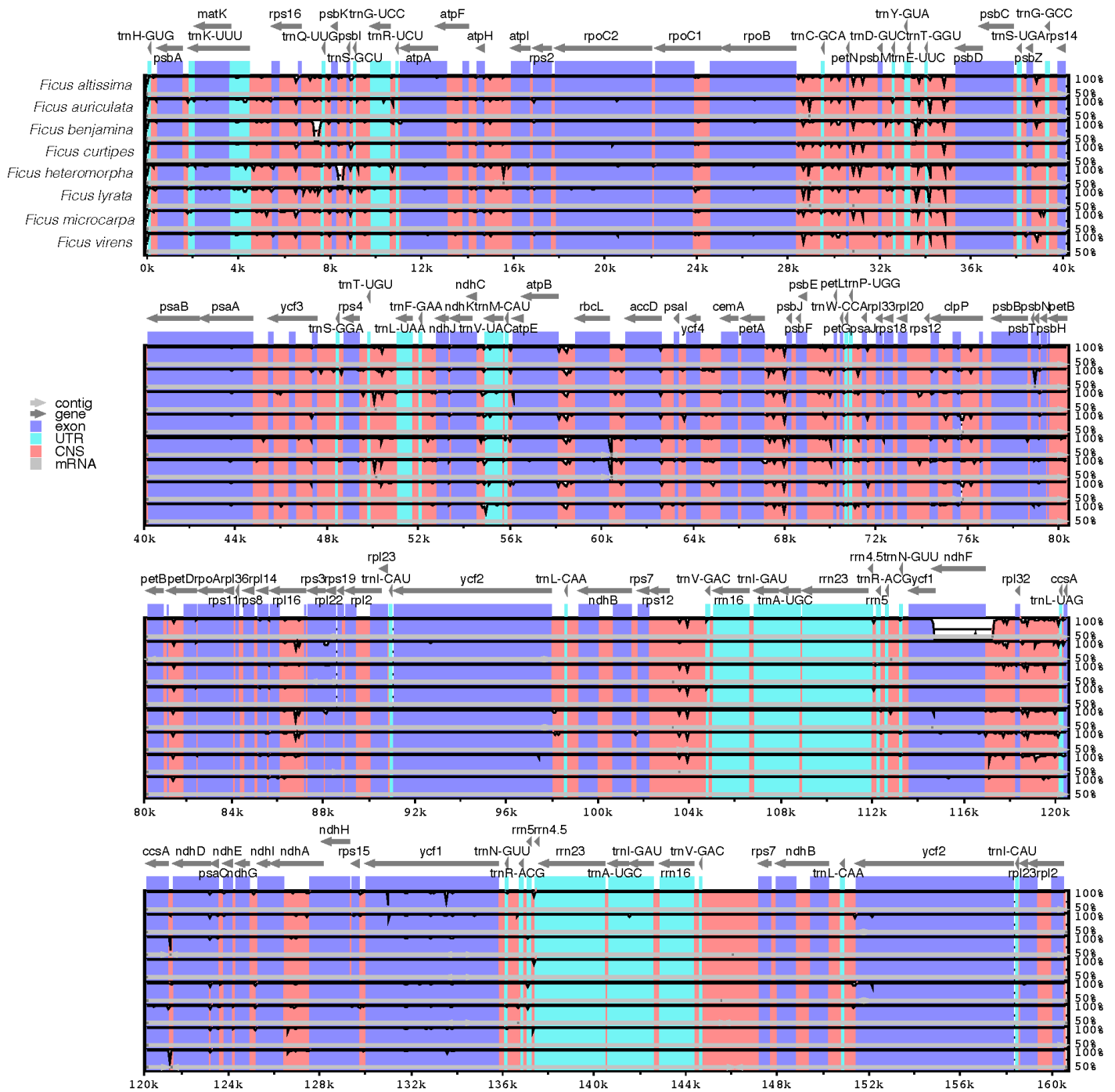


Figure 4

Sequence alignment of the chloroplast genomes of eight *Ficus* species, with *Ficus religiosa* as a reference. The X-axis represents the coordinates in the chloroplast genome, and the Y-axis represents the percentage of homogeneity, ranging from 50% to 100%.

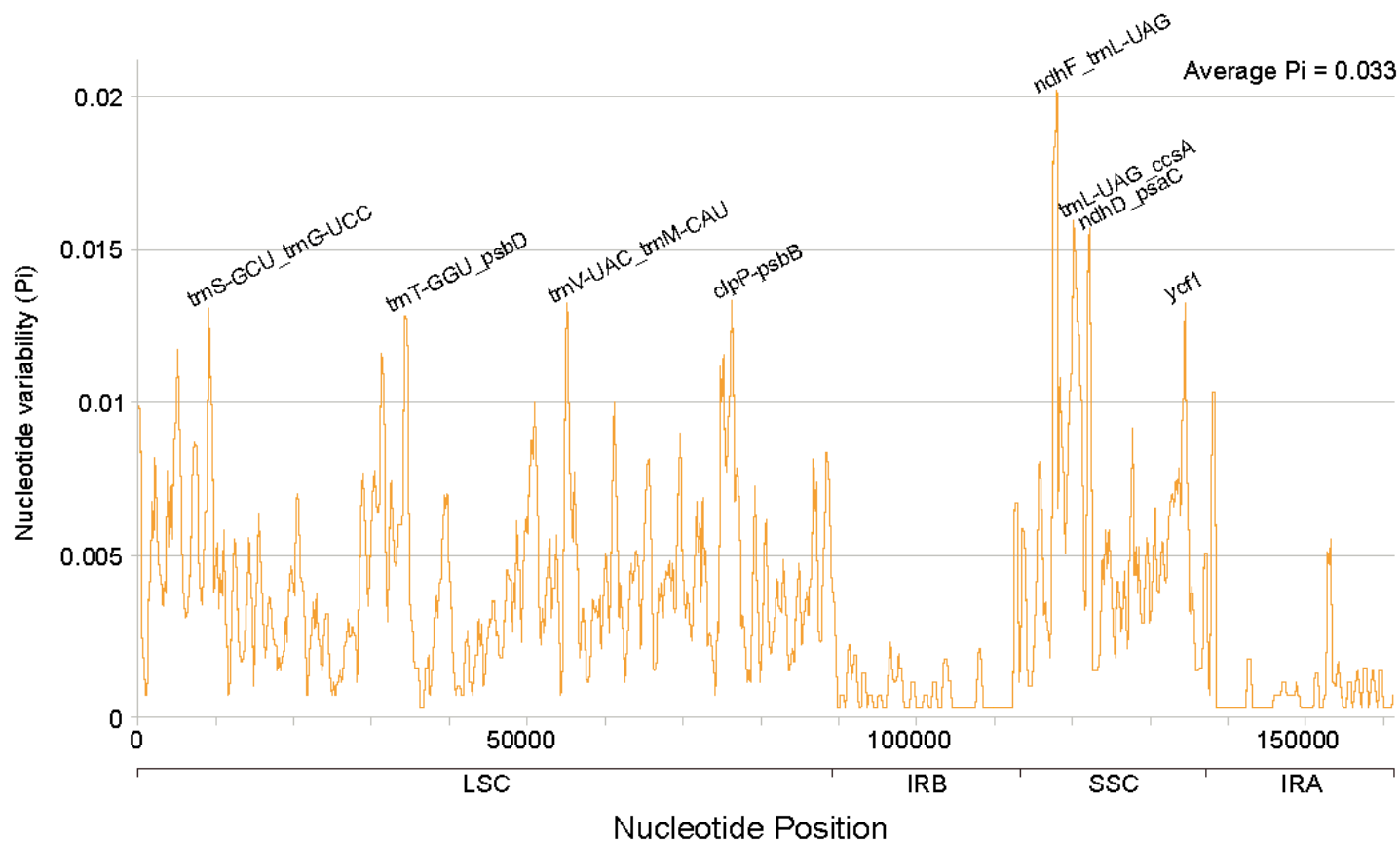


Figure 5

Nucleotide variability (Pi) in coding region of eight *Ficus* species.

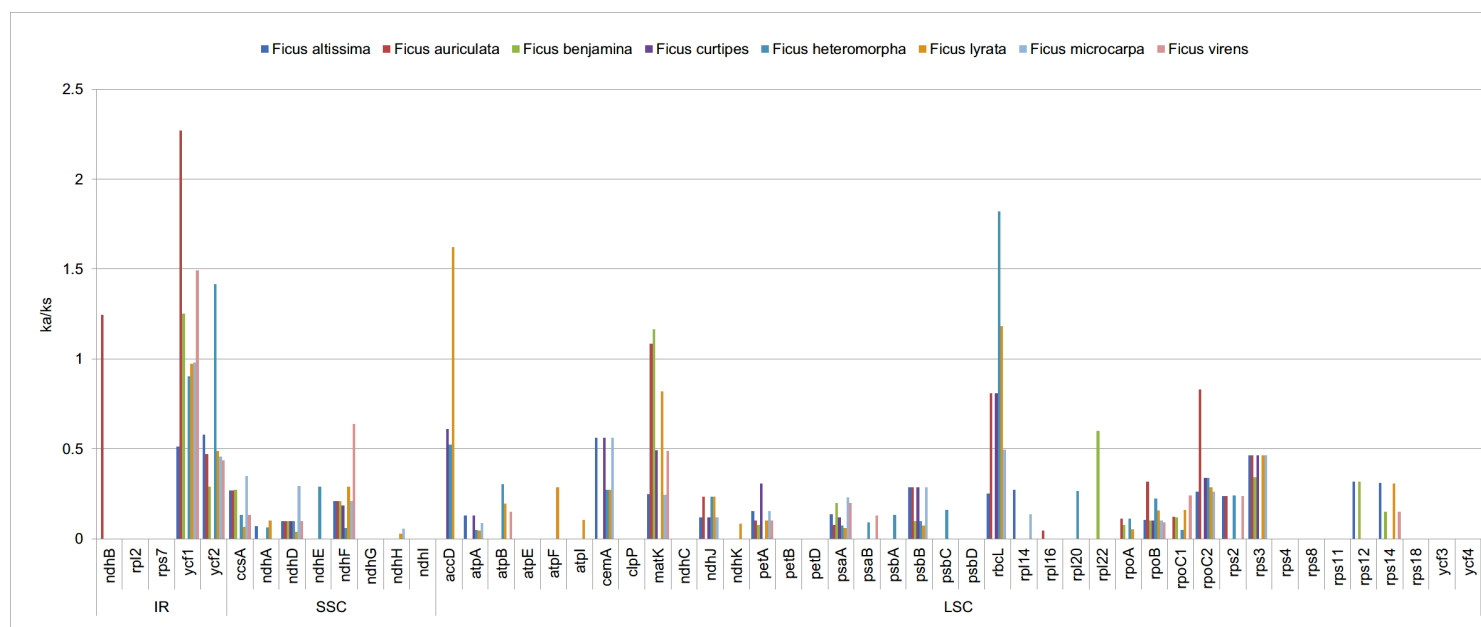


Figure 6

The Ka/Ks ratio of 53 protein-coding genes of eight cp genomes for comparison with *Ficus religiosa*.

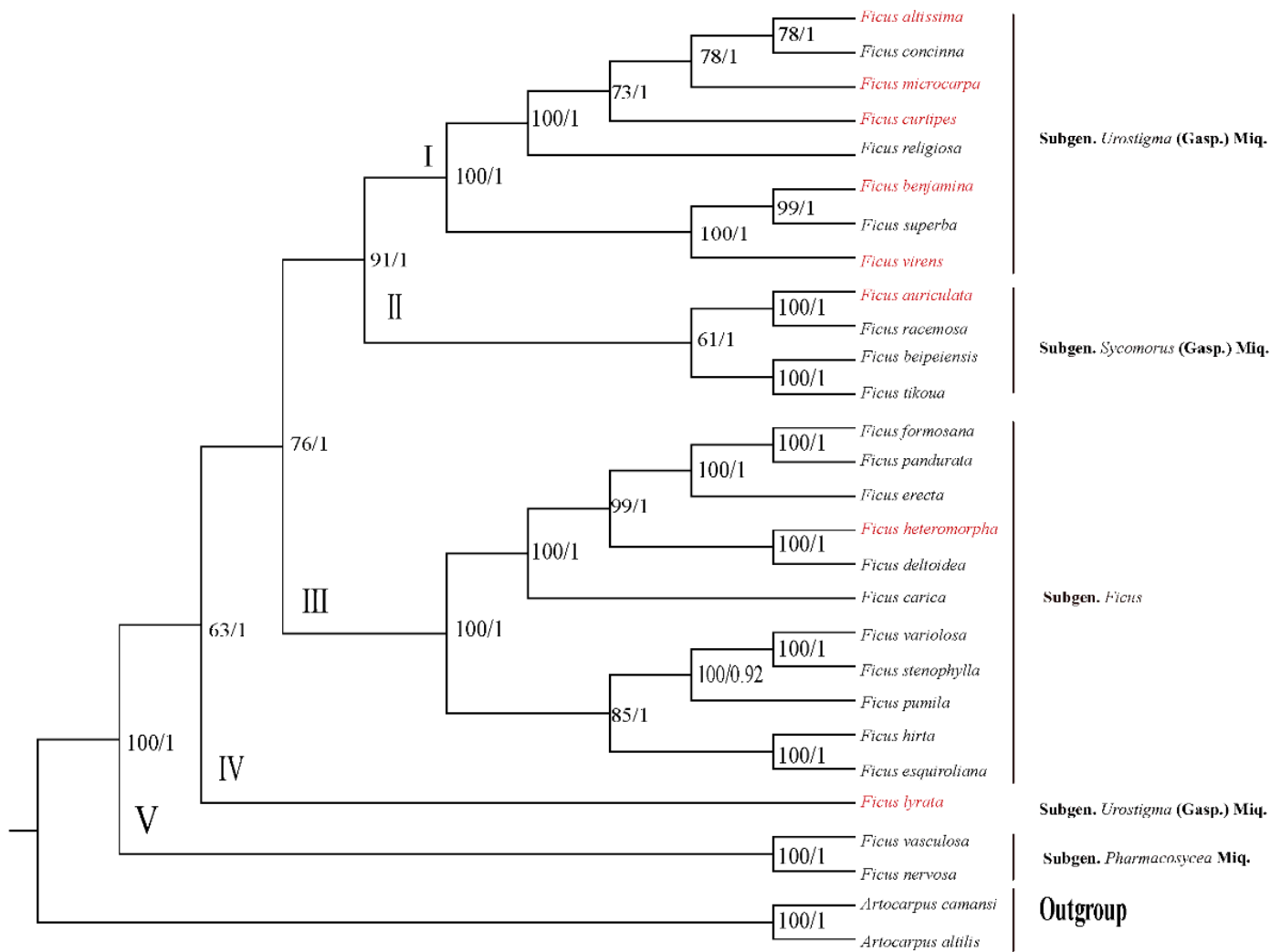


Figure 7

Phylogenetic tree inferred from maximum likelihood (ML) and Bayesian inference (BI) based on the 28 taxa complete chloroplast genomes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)