RESEARCH ARTICLE

# Comparison of chloroplast genomes and phylogenomics in the *Ficus sarmentosa* complex (Moraceae)

**Zhen Zhang[1], De-Shun Zhang[1], Lu Zou[2], Chi-Yuan Yao**[ORCID][1]*

1 College of Architecture and Urban Planning, Tongji University, Shanghai, China, 2 School of Life Sciences, East China Normal University, Shanghai, China

* cyyao@tongji.edu.cn

## Abstract

Due to maternal inheritance and minimal rearrangement, the chloroplast genome is an important genetic resource for evolutionary studies. However, the evolutionary dynamics and phylogenetic performance of chloroplast genomes in closely related species are poorly characterized, particularly in taxonomically complex and species-rich groups. The taxonomically unresolved *Ficus sarmentosa* species complex (Moraceae) comprises approximately 20 taxa with unclear genetic background. In this study, we explored the evolutionary dynamics, hotspot loci, and phylogenetic performance of thirteen chloroplast genomes (including eleven newly obtained and two downloaded from NCBI) representing the *F. sarmentosa* complex. Their sequence lengths, IR boundaries, repeat sequences, and codon usage were compared. Both sequence length and IR boundaries were found to be highly conserved. All four categories of long repeat sequences were found across all 13 chloroplast genomes, with palindromic and forward sequences being the most common. The number of simple sequence repeat (SSR) loci varied from 175 (*F. dinganensis* and *F. howii*) to 190 (*F. polynervis*), with the dinucleotide motif appearing the most frequently. Relative synonymous codon usage (RSCU) analysis indicated that codons ending with A/T were prior to those ending with C/T. The majority of coding sequence regions were found to have undergone negative selection with the exception of ten genes (*accD*, *clpP*, *ndhK*, *rbcL*, *rpl20*, *rpl22*, *rpl23*, *rpoC1*, *rps15*, and *rps4*) which exhibited potential positive selective signatures. Five hypervariable genic regions (*rps15*, *ycf1*, *rpoA*, *ndhF*, and *rpl22*) and five hypervariable intergenic regions (*trnH-GUG-psbA*, *rpl32-trnL-UAG*, *psbZ-trnG-GCC*, *trnK-UUU-rps*16 and *ndhF-rpl32*) were identified. Overall, phylogenomic analysis based on 123 *Ficus* chloroplast genomes showed promise for studying the evolutionary relationships in *Ficus*, despite cyto-nuclear discordance. Furthermore, based on the phylogenetic performance of the *F. sarmentosa* complex and *F. auriculata* complex, the chloroplast genome also exhibited a promising phylogenetic resolution in closely related species.

## Introduction

The genus *Ficus* L. (Moraceae) is a species-rich taxon which contains at least 800 species and is widely distributed across tropical and subtropical regions [1–3]. Due to insufficient genetic differentiation, the genus *Ficus* is taxonomically complex and contains many sympatric species, including the *F. pedunculosa* group, *F. punctata* group, *F. chartacea* group, and *F. subulata* group, among others [1, 3, 4]. Meanwhile, widely not-strict one-to-one obligate mutualism between fig trees and fig wasps has resulted in frequent hybridization and introgression among *Ficus* species [5–8], which has so far hindered research on the taxonomy and evolutionary history [9]. Our current understanding of the *Ficus* phylogenetic framework is the result of research on a few nuclear loci, such as ITS, ETS, *G3pdh*, GBSSI, and *waxy* [3, 4, 10–14]. Although nuclear genome data have been used to reconstruct the *Ficus* phylogeny [15–17], these studies represent less than ten percent of *Ficus* species, which is unlikely to accurately represent the evolutionary relationship in the genus *Ficus*.

While the use of nuclear genome data has advantages for the detection of hybridization and introgression, organellar genomic resources are also of great importance to evolutionary research due to maternal inheritance as a single unit [18, 19]. In the last decade, due to the rapidly decreasing cost of whole-genome sequencing (WGS) and the development of chloroplast genomic assembling pipelines, such as GerOrganelle [20], Fast-Plast (https://github.com/mrmckain/Fast-Plast), NOVOPlasty [21], and ORG.asm (https://git.metabarcoding.org/org-asm/org-asm), plastome-based evolutionary research has become easier and more cost-effective [22–24]. However, chloroplast genomes are publicly available for less than five percent of species in *Ficus* [25]. Bruun-Lund et al. [26] published a novel and innovative *Ficus* phylogenetic framework based on 59 newly obtained chloroplast genomes, and the results of which were obviously inconsistent with phylogenies based on the nuclear genome. Unfortunately, many of the genome sequences used by Bruun-Lund et al. contained gaps (an average of 7% missing data with a maximum of 65%), leading to difficulty in comparative chloroplast genomics and possible phylogenetic artifacts [27]. Even so, the chloroplast framework of Brunn-Lund et al. has been replicated by subsequent research generally with an extended dataset [17]. Overall, according to studies in both *Ficus* and the other taxa [28–30], it has been verified to be effective for chloroplast genomes to reconstruct the phylogenetic reference at infra-, intergeneric, and higher ranks. However, little is known about the phylogenetic performance of complete chloroplast genomes in studies of closely related species or even infra-species, particularly in the taxonomically complex genera such as *Ficus*.

The *Ficus* subg. *Synoecia* sect. *Rhizocladus* subsect. *Plagiostigma* comprises approximately 10 species and 11 varieties which are widely distributed across east Asia, parapatric to the distribution center of *Ficus* southeast Asia [1, 3, 13, 16, 31]. Subsect. *Plagiostigma* is genetically unclear and taxonomically unresolved, forming the *Ficus sarmentosa* species complex [4]. After the first systematic treatment of the *F. sarmentosa* complex by Corner in 1960 and 1965 [32, 33], few studies have attempted to unravel the taxonomic complexity of this group with the exception of descriptions of several controversial taxa in the 1980s [1, 34–37] (such as *F. dinganensis*, *F. guizhouensis*, and *F. polynervis*) and the rank elevation of some varieties (such as *F. pubigera* var. *anserina* and *F. sarmentosa* var. *thunbergii*) [38–40]. More recently, some of this phylogenetic ambiguity was resolved through the molecular work of Zhang et al. [4]. Zhang's work resulted in the rank elevation of *F. pubigera* var. *anserina* and the discovery that *F. sarmentosa* is not monophyletic, complicating the relationship between the complex and these previously described species in the 1980s. However, because only three loci (ITS, ETS, and *G3pdh*) were used to resolve the genetic background of the complex, the results were neither stable nor highly resolved [4]. The inclusion of more variable genetic loci or genome data

should be helpful to resolve the taxonomic uncertainty of the *F. sarmentosa* complex. In this study, we utilized both comparative chloroplast genomics and phylogenomics to characterize 1) the diversity of hotspot loci, 2) the variation among chloroplast genomes, and 3) the potential of chloroplast genomes to resolve the evolutionary relationships between closely related species of the F. sarmentosa species complex.

## Materials and methods

### Sample collection, DNA extraction and resequencing, and genome assembly and annotation

Healthy, young leaves were collected from the field in 2015–2021, and each was sealed in silica gel. Based on the phylogenetic relationships outlined in Zhang et al. [4], we sampled eleven taxa within the *F. sarmentosa* species complex in order to maximize genetic coverage. Detailed sample information is shown in S1 Table. All voucher specimens were stored at the herbarium of East China Normal University (HSNU).

Total DNA was extracted from 100 mg of dry leaf tissue using the CTAB method [41]. After quality detection with NanoDrop and Qubit 2.0, purified DNA samples were randomly ultrasonicated into ~350 bp segments, which were subsequently used to construct paired-end libraries. Whole-genome resequencing (WGS) was carried out using the Illumina NovaSeq 6000 platform, according to the PE150 sequencing strategy. Raw reads were filtered and cleaned according to the following criteria: reads containing > 10% unidentified nucleotides, > 50% low-quality bases (Q≤5), or adapter sequences were omitted for further analyses. Finally, the chloroplast genome sequences of two more taxa in the *F. sarmentosa* complex, *F. sarmentosa* var. *henryi* (GenBank accession no. OL415083) and *F. dinganensis* (GenBank accession no. OK375500), were included for further comparative analyses.

Clean data were used to assemble the complete circular chloroplast genome with GetOrganelle v1.6.4 [20], utilizing the "embplant_pt" model. In the case that the output sequence was not circular, the R (number of runs) and w (word size) parameters were adjusted until circularity was achieved. Chloroplast genome annotation was carried out using PGA [42], with the default parameters. The chloroplast genome was visualized using the online OGDRAW tool (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) [43]. For consistency, the two supplementary plastome sequences (OL415083 and OK375500) were re-annotated according to the same routine.

### Analysis of chloroplast genome structure

All 13 *F. sarmentosa* complex genomes were analyzed to determine the lengths and GC contents of the whole genomes, four quadripartite regions (large single copy (LSC), small single copy (SSC), and two inverted repeats (IRs)), and coding sequence (CDS) regions.

### Analysis of IR contraction and expansion

The online R Shiny application IRscope (https://irscope.shinyapps.io/irapp/) [44] was used to examine and visualize the boundary variation of LSC/IR/SSC of all 13 *F. sarmentosa* complex genomes.

### Analysis of long repeat sequences and SSRs

The four categories of long repeat sequences, forward (F), reverse (R), complement (C), and palindromic (P), were analyzed using the REPuter online tool (https://bibiserv.cebitec.uni-bielefeld.de/reputer) [45], with 50 maximum computed repeats and a minimal repeat size of 8.

One of the inverted repeat regions (IRb) was removed in the REPuter analysis to avoid repeatable results. Simple sequence repeats (SSRs) were detected using MISA (https://webblast.ipk-gatersleben.de/misa/) [46], with the following parameters: 8 repeat units for mononucleotide SSRs, 5 repeat units for dinucleotide SSRs, 4 repeat units for trinucleotide SSRs, and 3 repeat units for tetra-, penta-, and hexanucleotide SSRs. The maximal sequence length between two SSRs was set to 100 bp.

## Comparison of complete chloroplast genomes and diversity hotspot analysis

All 13 *F. sarmentosa* complex genomes were compared using the mVISTA online tool (https://genome.lbl.gov/vista/mvista/submit.shtml) [47], with the global multiple alignment model (LAGAN). The *F. anserine* chloroplast genome was used as the reference and the RankVISTA probability threshold was set to 0.5. All genes and intergenic regions were extracted from the Genbank annotation files in batches using Perl scripts created by Xiao-Jian Qu (https://github.com/quxiaojian/Bioinformatic_Scripts). Alignments of the genic and intergenic loci were carried out using MUSCLE v5.1 [48], with default parameters. After alignment, the nucleotide diversity (π) was calculated for all genic and intergenic loci.

## Codon usage analysis

To compare codon usage patterns of all the CDS sequences across 13 *F. sarmentosa* complex genomes, the relative synonymous codon usage (RSCU) was calculated using DAMBE v7.3.11 [49].

## Selective pressure analysis

The ratio of nonsynonymous (Ka) to synonymous (Ks) substitution can be used to quantify evolutionary selective pressure. The Ka/Ks ratio was calculated for all 79 unique CDSs using TBtools v1.09876 [50]. Positive selection was indicated by Ka/Ks > 1, negative (purifying) selection was indicated by Ka/Ks < 1, and neutral selection was indicated by Ka/Ks = 1. Based on the phylogenetic framework outlined in Zhang et al. [4], *F. simplicissima* was chosen as the reference to calculate Ka and Ks between *F. simplicissima* and our sampled representatives of the *F. sarmentosa* complex. For visualization purposes, the NaN value (i.e., both Ks and Ka = 0) was manually set as "1" to denote neutral selection. Finally, the infinity value (Ks > 0 and Ka = 0) was counted alone.

## Phylogenetic analysis

The 13 *F. sarmentosa* complex chloroplast genomes were combined with 36 *Ficus* chloroplast genomes from GenBank and 59 *Ficus* genomes published by Bruun-Lund et al. [26]. Additionally, 18 genomes from the China National GeneBank (accession number: CNP0001337) and Genome Sequence Archive (accession number: PRJCA002187) [15, 51], including 8 samples belonging to the *F. sarmentosa* complex, were assembled to further explore the *Ficus* phylogenomics and the potential of chloroplast genomes to resolve the evolutionary relationships among the closely related species. Seven samples from the Olmedieae tribe were chosen as the outgroup, according to previous studies [26, 52]. After discarding four samples found to be more than half (> 50%) missing data (three in the genus *Ficus* and one in the outgroup), a total of 129 chloroplast genomes were used to construct the phylogenetic tree (detailed samples information shown in S1 Table). Additionally, one of the two IRs was removed. The 129 genomes were aligned using MAFFT v7.490 [53], with the "auto" model. The aligned

sequences were trimmed with trimAl and sites with > 10% gaps were removed [54]. The maximum likelihood (ML) tree was constructed using IQ-TREE 2 with ultrafast bootstrap (-bb) and aLRT test (-alrt) numbers set at 10000. The optimal nucleotide substitution model was chosen with ModelFinder [55].

## Results

### Summary of all 13 complete chloroplast genomes in the *F. sarmentosa* complex

For the eleven newly resequenced samples, a total of 79G bases were obtained from 19,379,830 (*F. sarmentosa* var. *sarmentosa*) to 23,945,593 (*F. sarmentosa* var. *henryi*) clean paired-end reads. The ratios of chloroplast paired reads to whole reads ranged from 0.87% (*F. sarmentosa* var. *henryi*) to 4.05% (*F. guizhouensis*). The average kmer-coverage values ranged from 41.5 (*F. sarmentosa* var. *henryi*) to 128.2 (*F. sarmentosa* var. *thunbergii* and *F. pubigera*). Detailed information on high-throughput sequencing data can be found in S2 Table.

All *F. sarmentosa* complex genomes, including 11 newly obtained genomes and two genomes downloaded from GenBank, exhibited a typical quadripartite structure (Fig 1), containing one LSC, one SSC, and two IRs. The lengths of the complete chloroplast genomes ranged from 160,018 (*F. dinganensis*) to 160,385 bp (*F. sarmentosa* var. *lacrymans*) (Table 1). The LSCs accounted for 55.12–55.20% of the total genome size and ranged from 88,200 (*F. dinganensis*) to 88,535 bp (*F. sarmentosa* var. *lacrymans*) in length. The SSCs accounted for 12.51–12.55% of the total genome size and ranged from 20,064 (*F. sarmentosa* var. *lacrymans*) to 20,108 bp (*F. pubigera*) in length. The IRs accounted for 16.14–16.17% of the total genome size and varied from 25,866 (*F. dinganensis*) to 25,898 bp (*F. sarmentosa* var. *nipponica* and *F. sarmentosa* var. *thunbergii*). The CDS regions ranged from 79,149 (*F. dinganensis*) to 79,308 bp (*F. sarmentosa* var. *impressa*) in accumulated lengths. The GC content of the whole genome ranged from 35.94 to 35.99%, with *F. sarmentosa* var. *henryi* having the highest GC content, and *F. anserine* and *F. pubigera* having the lowest. The GC contents of LSCs, SSCs, and IRs were also calculated. Overall, the IRs had the highest GC content, which ranged from 42.63% (*F. anserina*, *F. howii*, and *F. pubigera*) to 42.65% (*F. pumila*, *F. sarmentosa* var. *lacrymans*, and *F. sarmentosa* var. *sarmentosa*). SSCs had the lowest GC content, which ranged from 28.94% (*F. sarmentosa* var. *lacrymans* and *F. sarmentosa* var. *sarmentosa*) to 29.09% (*F. sarmentosa* var. *henryi*, *F. sarmentosa* var. *impressa*, and *F. sarmentosa* var. *nipponica*). The same analyses applied to all 13 genomes, 131 genes were annotated, including 86 coding genes, 37 transfer RNA (tRNA) genes, and 8 ribosomal RNA (rRNA) genes. Finally, it was discovered that the *infA* gene was intensively pseudogenized [17].

### IR contraction and expansion

Across all 13 *F. sarmentosa* complex genomes, four junctions among IR, LSC, and SSC regions were compared (Fig 2). Invariably and consistently across all 13 genomes, the boundary between the LSC and IRb regions (JLB) was located within the *rps19* gene, 171 bp away from its starting base and 108 bp away from its ending base. The boundary between the IRb and SSC regions (JSB) was located within the *ndhF* gene, either 17 or 25 bp (*F. pumila*) away from its starting base and 2,236 (*F. anserina*, *F. pubigera*, *F. sarmentosa* var. *henryi*, and *F. sarmentosa* var. *sarmentosa*), 2,237 (*F. pumila*), or 2,245 bp (all other taxa) away from its ending base. The boundary between the SSC and IRa regions (JSA) was located within the *ycf1* gene, either 4,713, 4,722, or 4,724 bp away from its starting base and 1,024 (*F. pumila*) or 1,026 bp away from its ending base. For the IRa and LSC regions, the junction between them was located
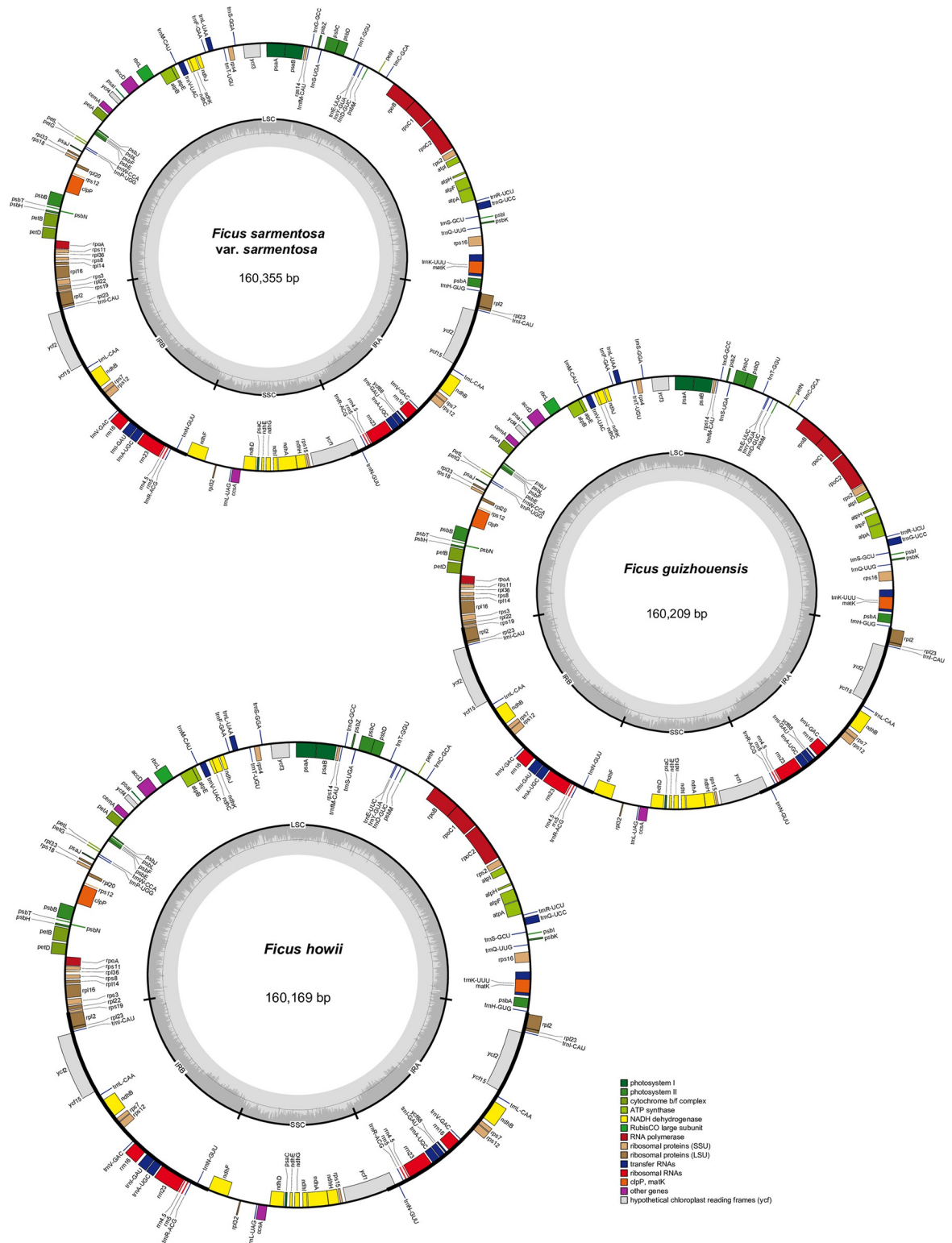
**Fig 1. Chloroplast gene maps of *Ficus sarmentosa* var. *sarmentosa*, *Ficus guizhouensis*, and *Ficus howii*.** Genes drawn inside are transcribed clockwise and genes drawn outside are counterclockwise. Genes belonging to different functional groups are color coded. In the inner circle, dark gray and light gray indicate the GC content and AT content, respectively. The boundaries of the large single copy (LSC), small single copy (SSC), and two inverted regions (IRa, IRb) are also shown in the inner circle.

**Table 1. Summary of complete chloroplast genomes of all thirteen taxa in the *F. sarmentosa* species complex.**

| Taxon | Total Length (bp) | LSC Length (bp) | SLC Length (bp) | IR Length (bp) | CDS Length (bp) | Total GC (%) | LSC GC (%) | SSC GC (%) | IR GC (%) |
|---|---|---|---|---|---|---|---|---|---|
| *F. anserina* | 160267 | 88395 | 20106 | 25883 | 79173 | 35.94 | 33.61 | 28.97 | 42.63 |
| *F. dinganensis* | 160018 | 88200 | 20086 | 25866 | 79149 | 35.98 | 33.67 | 29.00 | 42.64 |
| *F. guizhouensis* | 160209 | 88330 | 20103 | 25888 | 79272 | 35.98 | 33.66 | 29.02 | 42.64 |
| *F. howii* | 160169 | 88328 | 20076 | 25882 | 79167 | 35.98 | 33.65 | 29.03 | 42.63 |
| *F. polynervis* | 160356 | 88507 | 20063 | 25893 | 79155 | 35.95 | 33.62 | 28.95 | 42.64 |
| *F. pubigera* | 160299 | 88425 | 20108 | 25883 | 79167 | 35.94 | 33.61 | 28.98 | 42.63 |
| *F. pumila* | 160229 | 88369 | 20100 | 25880 | 79152 | 35.98 | 33.64 | 29.08 | 42.65 |
| *F. sarmentosa* var. *henryi* | 160183 | 88307 | 20080 | 25898 | 79170 | 35.99 | 33.66 | 29.09 | 42.64 |
| *F. sarmentosa* var. *impressa* | 160293 | 88417 | 20080 | 25898 | 79308 | 35.97 | 33.63 | 29.09 | 42.64 |
| *F. sarmentosa* var. *lacrymans* | 160385 | 88535 | 20064 | 25893 | 79155 | 35.95 | 33.63 | 28.94 | 42.65 |
| *F. sarmentosa* var. *nipponica* | 160314 | 88436 | 20082 | 25898 | 79170 | 35.97 | 33.63 | 29.09 | 42.64 |
| *F. sarmentosa* var. *sarmentosa* | 160355 | 88501 | 20068 | 25893 | 79155 | 35.95 | 33.62 | 28.94 | 42.65 |
| *F. sarmentosa* var. *thunbergii* | 160282 | 88397 | 20089 | 25898 | 79179 | 35.98 | 33.65 | 29.06 | 42.64 |

https://doi.org/10.1371/journal.pone.0279849.t001

between the *rpl2* and *trnH* genes, either 62 or 63 bp away from the starting base of the *trnH* gene.

## Long repeat sequences and simple sequence repeats (SSRs)

Across all 13 *F. sarmentosa* complex genomes, a total of 373 long repeat sequences were identified, representing all four repeat categories: forward (F), reverse (R), complement (C), and palindromic (P) (Fig 3A). All four categories of repeats were detected within all 13 chloroplast genomes. Four taxa, *F. anserina*, *F. dinganensis*, *F. polynervis*, and *F. sarmentosa* var. *sarmentosa*, contained the greatest number of long repeat sequences (31), while *F. sarmentosa* var. *thunbergii* contained the least (26). Among all four repeats, P repeats were the most common across all 13 genomes, ranging from 14 (*F. guizhouensis*) to 16. There were relatively fewer R and C repeats, with *F. polynervis*, *F. sarmentosa* var. *lacrymans*, and *F. sarmentosa* var. *sarmentosa* containing the most (4). For the lengths of long repeat sequences, 30–39 bp is the most common ranging from 21 (*F. sarmentosa* var. *thunbergii*) to 26 (*F. anserina* and *F. dinganensis*) times (Fig 3B). The long repeat sequences over 60 bp in length were the least, appearing only once. The maximum length among all the long repeat sequences was 64 bp.

Simple sequence repeats (SSRs) consisting of the 1- to 6- nucleotide motifs were surveyed across the 13 *F. sarmentosa* complex genomes (Fig 4A). Mononucleotide SSRs were the most abundant (78.71%) across all 13 genomes. Dinucleotide SSRs, the most commonly used motif in population genetics and phylogenetics, were also relatively abundant, appearing from 19 (*F. howii*, *F. polynervis*, *F. sarmentosa* var. *lacrymans*, and *F. sarmentosa* var. *sarmentosa*) to 22 (*F. pumila* and *F. sarmentosa* var. *thunbergii*) times. It is noteworthy that tetranucleotide SSRs were more common than trinucleotide SSRs (Fig 4A), considering the former appeared from 9 (*F. guizhouensis*, *F. polynervis*, *F. sarmentosa* var. *lacrymans*, and *F. sarmentosa* var. *sarmentosa*) to 11 (*F. pumila*, *F. sarmentosa* var. *henryi*, *F. sarmentosa* var. *impressa*, and *F. sarmentosa* var. *nipponica*) times. Hexanucleotide SSRs were absent from all genomes. The total number of SSRs varied from 175 (*F. dinganensis* and *F. howii*) to 190 (*F. polynervis*). Of the repeat
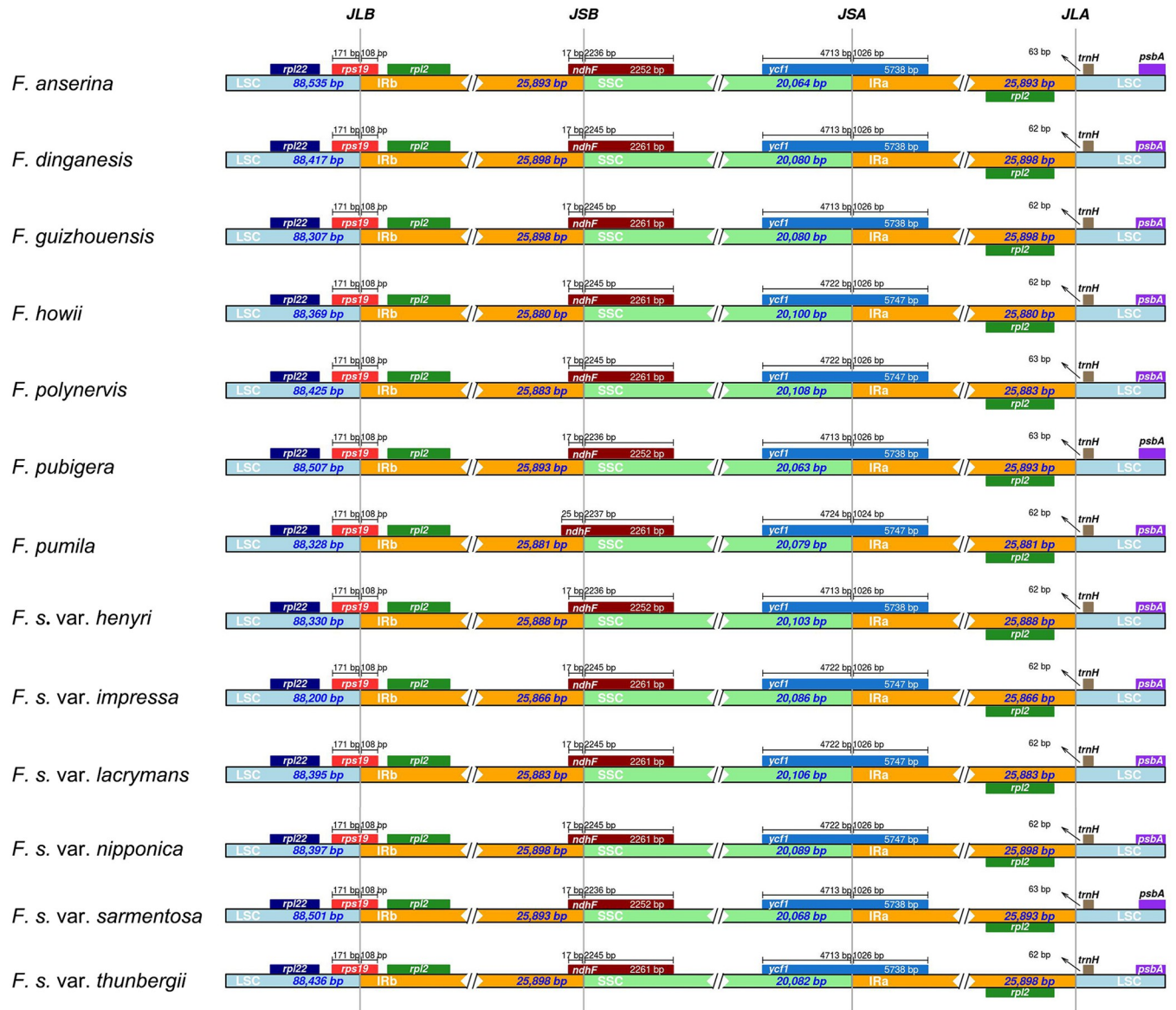
**Fig 2. Comparison of the boundaries between the large single copy (LSC), small single copy (SSC), and two inverted repeat regions (IRs) among 13 chloroplast genomes in the *F. sarmentosa* complex.** The numbers around the vertical lines indicate the distances between the boundaries and the starting or ending bases of their nearest genes.

https://doi.org/10.1371/journal.pone.0279849.g002

motifs, A/T was the most common mononucleotide motif, appearing from 132 (*F. guizhouensis*) to 149 (*F. polynervis* and *F. sarmentosa* var. *sarmentosa*) times (Fig 4B). AT/AT was the second most common motif, appearing between 18 (*F. howii*, *F. polynervis*, *F. sarmentosa* var. *lacrymans*, and *F. sarmentosa* var. *sarmentosa*) and 21 (*F. pumila* and *F. sarmentosa* var. *thunbergii*) times. The majority of the remaining motifs appeared only once to four times, with the exception of AATT/AATT appearing up to seven times (Fig 4B).

## Codon usage

Across all 13 *F. sarmentosa* complex genomes, the 79 unique protein-coding CDS regions were encoded by between 22,867 (*F. polynervis* and *F. sarmentosa* var. *lacrymans*) and 22,918 (*F.*
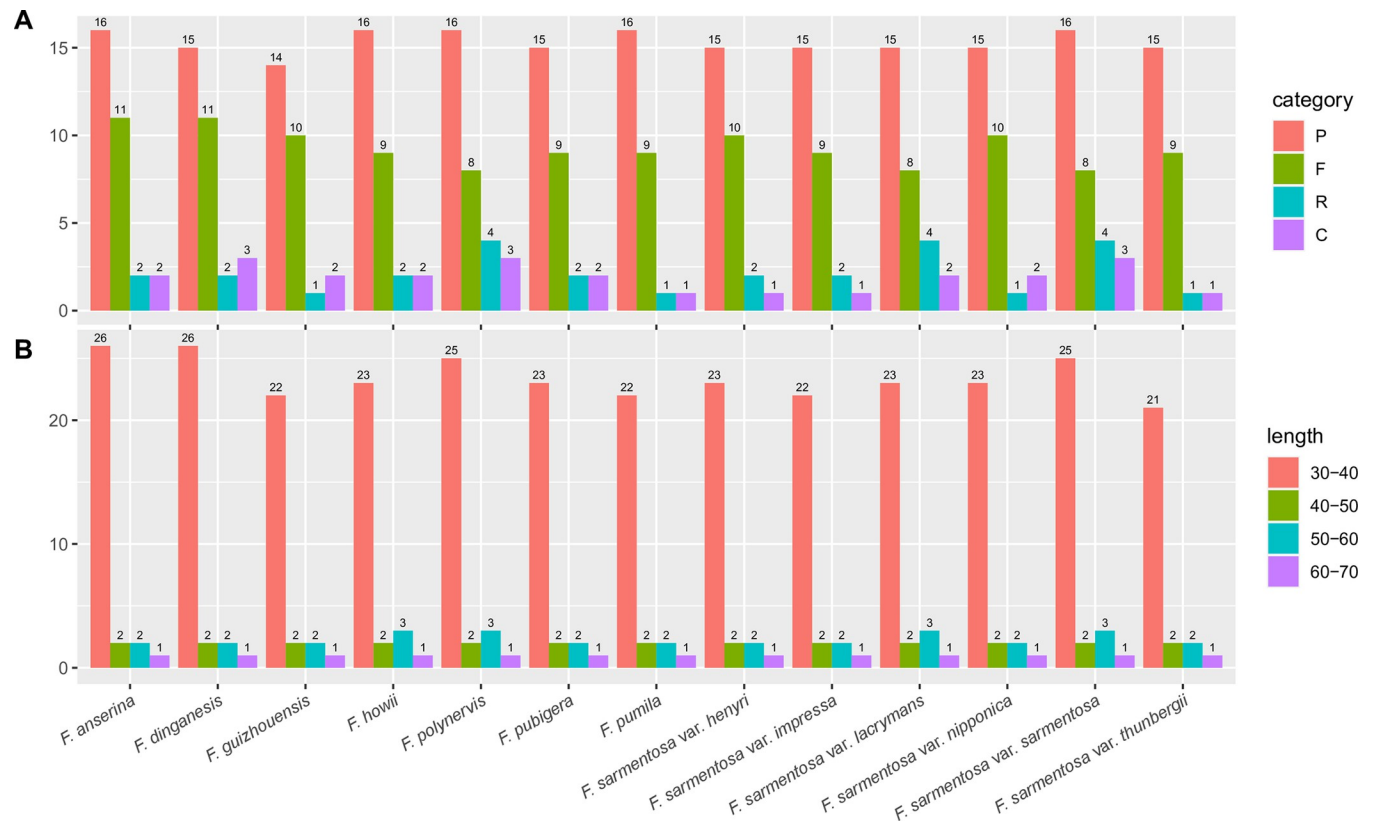
**Fig 3. Comparison of long repeat sequences among 13 *F. sarmentosa* complex genomes.** A, The number of each of four long repeat types (P, palindromic; F, forward; R; reverse; C complement); B, The number of long repeat sequences of different lengths.

*sarmentosa* var. *impressa*) codons. The codon usage among all 81 protein-coding genes is summarized in Table 2. Among these codons, CAU encoding histidine (H) was the most frequent, appearing 12,800 times across all 13 taxa. Except the stop codon, UGU encoding L-Cysteine (C) was the next rarest, appearing 718 times only. According to the RSCU analysis, GCU and CUU had the highest average values of 1.858 and 1.818, respectively (Table 2), whereas UAC and CGC had the lowest average values of 0.364 and 0.400, respectively. Among all three stop codons, UAA was the most common (53.64%). Thirty out of the 64 codons with RSCU > 1 ended with either A or U, while 32 out of 64 codons with RSCU < 1 ended with either G or C, with the exception of the AUA codon.

## Genomic divergence and hotspot regions

The divergence of whole sequence among the 13 *F. sarmentosa* complex genomes was analyzed using the mVISTA online platform with *F. anserina* as a reference. The results showed that the full-length chloroplast genomes were largely conserved across all 13 taxa. The majority of variable sites were located in intergenic spacer regions (marked red in Fig 5). Interestingly, IRs were found to be more conserved than either LSCs or SSCs.

Nucleotide diversity (π) was calculated for each gene and intergenic region to evaluate genetic differentiation and detect hyper-variable segments. Among 60 genes > 200 bp in length, *rps15*, *ycf1*, *rpoA*, *ndhF*, and *rpl22* exhibited the highest nucleotide diversity: 0.00535, 0.00462, 0.00456, 0.00427, and 0.00377, respectively (Fig 6). The alignment lengths of these
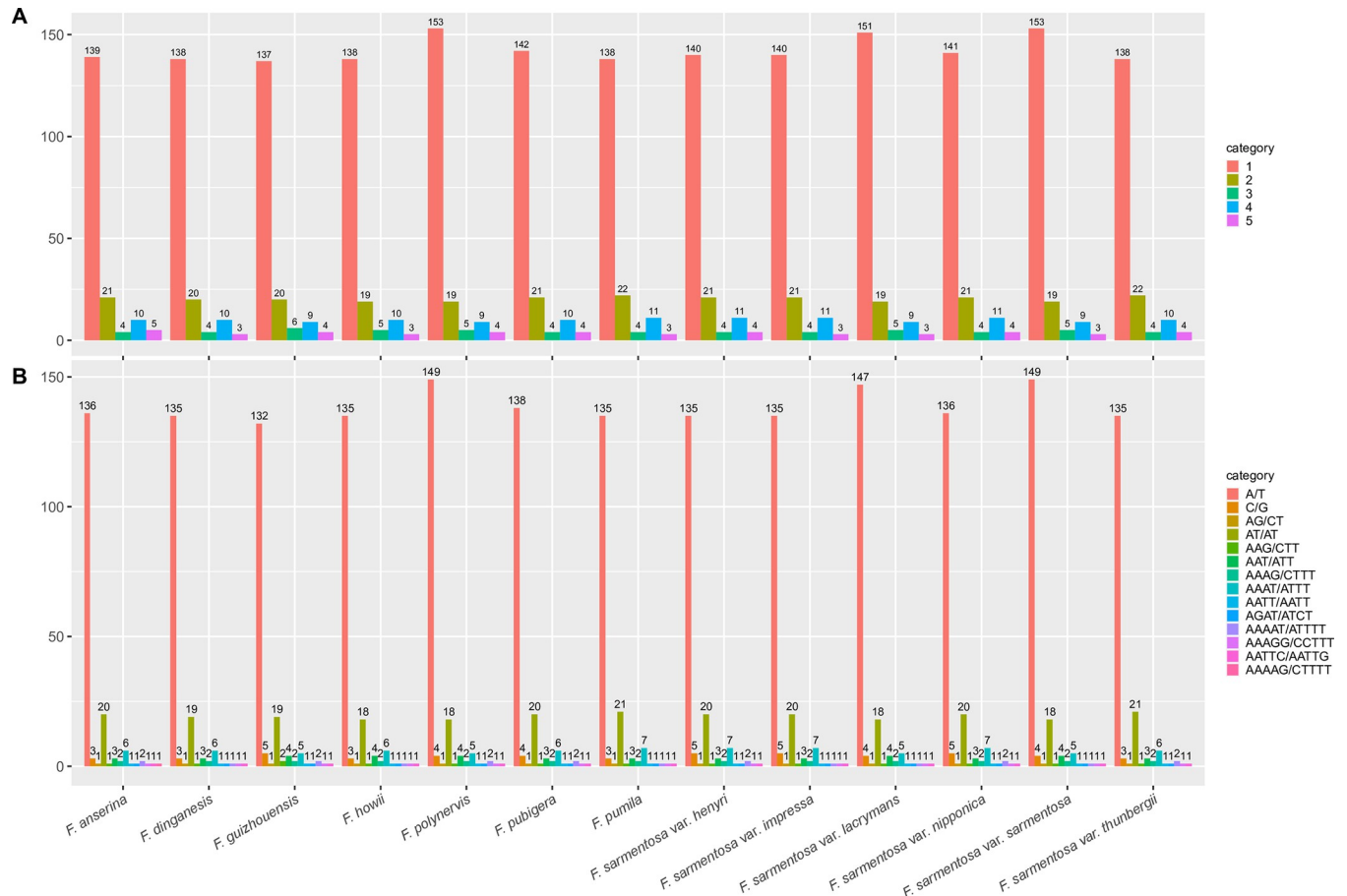
**Fig 4. Comparison of simple sequence repeats (SSRs) among 13 *F. sarmentosa* complex genomes.** A, The number of SSRs containing one- to five-nucleotide motifs; B, The number of different SSR motifs.

five genes ranged from 282 to 5,748 bp. Seven genes (*atpI*, *psbE*, *rpl33*, *rps18*, *psbH*, *rrn23*, and *psaC*) were found to be identical, with a nucleotide diversity of zero. The average diversity of the 60 genes was 0.001676, while the average diversity of intergenic spacer regions was 0.00435, approximately 2.6 times that of the genes (Fig 6). Among the intergenic spacer regions, the highest nucleotide diversity was exhibited by *trnH-GUG-psbA* (0.01458), followed by *rpl32-trnL-UAG* (0.01225), *psbZ-trnG-GCC* (0.01148), *trnK-UUU-rps16* (0.01144), and *ndhF-rpl32* (0.01112). The alignment lengths of these regions ranged from 366 to 1,829 bp (Fig 6).

## Selective pressure analysis

The ratio of nonsynonymous (Ka) to synonymous (Ks) substitutions was calculated to quantify the evolutionary selective pressure on the *F. sarmentosa* complex, with *F. simplicissima* used as the reference genome. Overall, Ka/Ks ratios of most genes were < 1 (Fig 7). Additionally, twenty genes were found to contain no substitutions, i.e., both Ka and Ks are zero (shown as "1" in Fig 7). However, five genes (*accD*, *matK*, *ndhF*, *rpoA*, and *ycf1*) had partial Ka/Ks ratios over 1, which are potential signals of positive selection. An infinite Ka/Ks ratio (Ka > 0 and Ks = 0) existed in 27 genes (Fig 7). Notably, ten genes (*accD*, *clpP*,

**Table 2. The relative synonymous codon usage (RSCU) of all 64 codons.** The taxa are represented by the numbers indicated in Table 1.

| Codon | AA | RSCU (relative synonymous codon usage) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| UGA | * | 0.662 | 0.662 | 0.662 | 0.662 | 0.662 | 0.662 | 0.662 | 0.662 | 0.662 | 0.662 | 0.662 | 0.662 | 0.662 |
| UAG | * | 0.701 | 0.701 | 0.74 | 0.701 | 0.74 | 0.701 | 0.74 | 0.74 | 0.74 | 0.74 | 0.74 | 0.74 | 0.74 |
| UAA | * | 1.636 | 1.636 | 1.597 | 1.636 | 1.597 | 1.636 | 1.597 | 1.597 | 1.597 | 1.597 | 1.597 | 1.597 | 1.597 |
| GCU | A | 1.853 | 1.853 | 1.853 | 1.853 | 1.862 | 1.853 | 1.854 | 1.861 | 1.863 | 1.869 | 1.861 | 1.862 | 1.857 |
| GCG | A | 0.413 | 0.413 | 0.413 | 0.413 | 0.415 | 0.413 | 0.412 | 0.404 | 0.404 | 0.411 | 0.404 | 0.415 | 0.413 |
| GCC | A | 0.594 | 0.594 | 0.6 | 0.597 | 0.589 | 0.594 | 0.592 | 0.591 | 0.591 | 0.589 | 0.591 | 0.589 | 0.59 |
| GCA | A | 1.14 | 1.14 | 1.135 | 1.136 | 1.134 | 1.14 | 1.142 | 1.144 | 1.143 | 1.132 | 1.144 | 1.134 | 1.141 |
| UGU | C | 1.57 | 1.575 | 1.569 | 1.569 | 1.561 | 1.561 | 1.569 | 1.569 | 1.569 | 1.561 | 1.569 | 1.561 | 1.569 |
| UGC | C | 0.43 | 0.425 | 0.431 | 0.431 | 0.439 | 0.439 | 0.431 | 0.431 | 0.431 | 0.439 | 0.431 | 0.439 | 0.431 |
| GAU | D | 1.596 | 1.596 | 1.596 | 1.596 | 1.596 | 1.598 | 1.594 | 1.595 | 1.595 | 1.597 | 1.595 | 1.598 | 1.593 |
| GAC | D | 0.404 | 0.404 | 0.404 | 0.404 | 0.404 | 0.402 | 0.406 | 0.405 | 0.405 | 0.403 | 0.405 | 0.402 | 0.407 |
| GAG | E | 0.462 | 0.467 | 0.464 | 0.464 | 0.459 | 0.463 | 0.464 | 0.465 | 0.466 | 0.461 | 0.465 | 0.459 | 0.463 |
| GAA | E | 1.538 | 1.533 | 1.536 | 1.536 | 1.541 | 1.537 | 1.536 | 1.535 | 1.534 | 1.539 | 1.535 | 1.541 | 1.537 |
| UUU | F | 1.368 | 1.367 | 1.374 | 1.367 | 1.373 | 1.369 | 1.372 | 1.37 | 1.37 | 1.372 | 1.371 | 1.373 | 1.37 |
| UUC | F | 0.632 | 0.633 | 0.626 | 0.633 | 0.627 | 0.631 | 0.628 | 0.63 | 0.63 | 0.628 | 0.629 | 0.627 | 0.63 |
| GGU | G | 1.34 | 1.34 | 1.338 | 1.338 | 1.341 | 1.334 | 1.342 | 1.337 | 1.338 | 1.342 | 1.337 | 1.341 | 1.342 |
| GGG | G | 0.609 | 0.606 | 0.61 | 0.612 | 0.612 | 0.609 | 0.609 | 0.612 | 0.61 | 0.613 | 0.612 | 0.612 | 0.609 |
| GGC | G | 0.404 | 0.402 | 0.404 | 0.407 | 0.397 | 0.407 | 0.402 | 0.405 | 0.404 | 0.396 | 0.405 | 0.397 | 0.402 |
| GGA | G | 1.647 | 1.652 | 1.648 | 1.643 | 1.651 | 1.649 | 1.647 | 1.647 | 1.648 | 1.649 | 1.647 | 1.651 | 1.647 |
| CAC | H | 0.463 | 0.468 | 0.46 | 0.465 | 0.457 | 0.463 | 0.465 | 0.464 | 0.464 | 0.457 | 0.464 | 0.457 | 0.464 |
| CAU | H | 1.537 | 1.532 | 1.54 | 1.535 | 1.543 | 1.537 | 1.535 | 1.536 | 1.536 | 1.543 | 1.536 | 1.543 | 1.536 |
| AUU | I | 1.471 | 1.471 | 1.471 | 1.477 | 1.47 | 1.473 | 1.471 | 1.468 | 1.467 | 1.47 | 1.468 | 1.47 | 1.47 |
| AUA | I | 0.965 | 0.967 | 0.959 | 0.965 | 0.965 | 0.966 | 0.962 | 0.964 | 0.963 | 0.963 | 0.964 | 0.965 | 0.964 |
| AUC | I | 0.564 | 0.562 | 0.57 | 0.558 | 0.565 | 0.561 | 0.567 | 0.568 | 0.57 | 0.567 | 0.568 | 0.565 | 0.566 |
| AAA | K | 1.531 | 1.534 | 1.534 | 1.533 | 1.528 | 1.53 | 1.534 | 1.536 | 1.537 | 1.527 | 1.536 | 1.528 | 1.539 |
| AAG | K | 0.469 | 0.466 | 0.466 | 0.467 | 0.472 | 0.47 | 0.466 | 0.464 | 0.463 | 0.473 | 0.464 | 0.472 | 0.461 |
| CUA | L | 1.162 | 1.16 | 1.169 | 1.166 | 1.145 | 1.169 | 1.158 | 1.155 | 1.154 | 1.146 | 1.155 | 1.145 | 1.16 |
| CUC | L | 0.521 | 0.522 | 0.528 | 0.518 | 0.52 | 0.521 | 0.518 | 0.516 | 0.516 | 0.52 | 0.516 | 0.52 | 0.515 |
| CUG | L | 0.504 | 0.501 | 0.5 | 0.501 | 0.513 | 0.504 | 0.504 | 0.506 | 0.505 | 0.513 | 0.506 | 0.513 | 0.505 |
| CUU | L | 1.813 | 1.816 | 1.802 | 1.814 | 1.823 | 1.806 | 1.82 | 1.823 | 1.825 | 1.821 | 1.823 | 1.823 | 1.819 |
| UUA | L | 1.249 | 1.251 | 1.25 | 1.249 | 1.254 | 1.249 | 1.25 | 1.254 | 1.254 | 1.251 | 1.254 | 1.252 | 1.25 |
| UUG | L | 0.751 | 0.749 | 0.75 | 0.751 | 0.746 | 0.751 | 0.75 | 0.746 | 0.746 | 0.749 | 0.746 | 0.748 | 0.75 |
| AUG | M | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| AAC | N | 0.446 | 0.444 | 0.442 | 0.446 | 0.452 | 0.446 | 0.444 | 0.443 | 0.443 | 0.454 | 0.443 | 0.453 | 0.445 |
| AAU | N | 1.554 | 1.556 | 1.558 | 1.554 | 1.548 | 1.554 | 1.556 | 1.557 | 1.557 | 1.546 | 1.557 | 1.547 | 1.555 |
| CCA | P | 1.166 | 1.16 | 1.165 | 1.163 | 1.154 | 1.166 | 1.162 | 1.162 | 1.165 | 1.158 | 1.162 | 1.154 | 1.16 |
| CCC | P | 0.68 | 0.681 | 0.679 | 0.677 | 0.678 | 0.676 | 0.68 | 0.684 | 0.684 | 0.678 | 0.684 | 0.678 | 0.684 |
| CCU | P | 1.592 | 1.596 | 1.586 | 1.598 | 1.596 | 1.597 | 1.592 | 1.592 | 1.591 | 1.596 | 1.592 | 1.596 | 1.591 |
| CCG | P | 0.562 | 0.563 | 0.57 | 0.562 | 0.573 | 0.562 | 0.566 | 0.562 | 0.561 | 0.568 | 0.562 | 0.573 | 0.565 |
| CAA | Q | 1.585 | 1.588 | 1.58 | 1.586 | 1.587 | 1.585 | 1.582 | 1.585 | 1.583 | 1.584 | 1.585 | 1.587 | 1.582 |
| CAG | Q | 0.415 | 0.412 | 0.42 | 0.414 | 0.413 | 0.415 | 0.418 | 0.415 | 0.417 | 0.416 | 0.415 | 0.413 | 0.418 |
| AGA | R | 1.496 | 1.5 | 1.5 | 1.497 | 1.502 | 1.494 | 1.496 | 1.499 | 1.501 | 1.502 | 1.499 | 1.502 | 1.497 |
| AGG | R | 0.504 | 0.5 | 0.5 | 0.503 | 0.498 | 0.506 | 0.504 | 0.501 | 0.499 | 0.498 | 0.501 | 0.498 | 0.503 |
| CGA | R | 1.6 | 1.596 | 1.604 | 1.596 | 1.602 | 1.601 | 1.596 | 1.591 | 1.594 | 1.602 | 1.591 | 1.602 | 1.596 |
| CGC | R | 0.405 | 0.399 | 0.398 | 0.399 | 0.403 | 0.399 | 0.399 | 0.399 | 0.398 | 0.403 | 0.399 | 0.403 | 0.399 |
| CGG | R | 0.426 | 0.425 | 0.419 | 0.425 | 0.423 | 0.43 | 0.414 | 0.419 | 0.419 | 0.423 | 0.419 | 0.423 | 0.414 |

*(Continued)*

**Table 2.** (Continued)

| Codon | AA | RSCU (relative synonymous codon usage) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| CGU | R | 1.569 | 1.581 | 1.579 | 1.581 | 1.571 | 1.57 | 1.591 | 1.591 | 1.589 | 1.571 | 1.591 | 1.571 | 1.591 |
| AGC | S | 0.447 | 0.456 | 0.455 | 0.46 | 0.451 | 0.45 | 0.454 | 0.454 | 0.454 | 0.45 | 0.454 | 0.447 | 0.454 |
| AGU | S | 1.553 | 1.544 | 1.545 | 1.54 | 1.549 | 1.55 | 1.546 | 1.546 | 1.546 | 1.55 | 1.546 | 1.553 | 1.546 |
| UCA | S | 1.111 | 1.111 | 1.115 | 1.114 | 1.116 | 1.107 | 1.111 | 1.109 | 1.108 | 1.116 | 1.11 | 1.116 | 1.109 |
| UCC | S | 0.817 | 0.817 | 0.809 | 0.817 | 0.803 | 0.814 | 0.806 | 0.812 | 0.816 | 0.8 | 0.813 | 0.803 | 0.808 |
| UCG | S | 0.449 | 0.449 | 0.448 | 0.452 | 0.448 | 0.452 | 0.45 | 0.448 | 0.447 | 0.448 | 0.448 | 0.448 | 0.446 |
| UCU | S | 1.624 | 1.622 | 1.628 | 1.617 | 1.632 | 1.627 | 1.632 | 1.631 | 1.629 | 1.635 | 1.629 | 1.632 | 1.636 |
| ACC | T | 0.701 | 0.699 | 0.701 | 0.701 | 0.701 | 0.701 | 0.702 | 0.695 | 0.697 | 0.702 | 0.695 | 0.701 | 0.699 |
| ACA | T | 1.187 | 1.189 | 1.182 | 1.185 | 1.177 | 1.185 | 1.176 | 1.177 | 1.179 | 1.179 | 1.177 | 1.177 | 1.174 |
| ACG | T | 0.417 | 0.416 | 0.422 | 0.417 | 0.417 | 0.417 | 0.422 | 0.426 | 0.425 | 0.418 | 0.426 | 0.417 | 0.426 |
| ACU | T | 1.695 | 1.696 | 1.695 | 1.697 | 1.706 | 1.697 | 1.7 | 1.701 | 1.698 | 1.702 | 1.701 | 1.706 | 1.701 |
| GUU | V | 1.463 | 1.462 | 1.461 | 1.468 | 1.466 | 1.466 | 1.461 | 1.458 | 1.461 | 1.46 | 1.458 | 1.466 | 1.461 |
| GUG | V | 0.513 | 0.511 | 0.522 | 0.513 | 0.513 | 0.51 | 0.513 | 0.515 | 0.521 | 0.516 | 0.515 | 0.513 | 0.513 |
| GUC | V | 0.405 | 0.406 | 0.405 | 0.403 | 0.409 | 0.405 | 0.406 | 0.411 | 0.407 | 0.415 | 0.411 | 0.409 | 0.406 |
| GUA | V | 1.618 | 1.621 | 1.613 | 1.616 | 1.612 | 1.618 | 1.62 | 1.616 | 1.612 | 1.609 | 1.616 | 1.612 | 1.62 |
| UGG | W | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| UAC | Y | 0.365 | 0.364 | 0.364 | 0.365 | 0.364 | 0.361 | 0.366 | 0.366 | 0.365 | 0.364 | 0.366 | 0.364 | 0.363 |
| UAU | Y | 1.635 | 1.636 | 1.636 | 1.635 | 1.636 | 1.639 | 1.634 | 1.634 | 1.635 | 1.636 | 1.634 | 1.636 | 1.637 |

*ndhK*, *rbcL*, *rpl20*, *rpl22*, *rpl23*, *rpoC1*, *rps15*, and *rps4*) possessed infinite Ka/Ks ratios in more than half of the taxa.

## Phylogenetic analysis

The chloroplast phylogenomic ML tree illustrated a well-supported phylogenetic relationship in the genus *Ficus* (Fig 8). Subgenus *Pharmacosycea* sect. *Pharmacosycea* was strongly supported as a sister group to the rest of the genus *Ficus* (SH-aLRT = 100 and MLBS = 100) Additionally, a clade including subg. *Urostigma* sect. *Galoglychia* and *Americana*, subg. *Sycomorus* sect. *Sycomorus*, and a few species of subg. *Pharmacosycea* sect. *Oreosycea* (subser. *Ablbipilae* in Corner's system) (clade A) was found to be sister to the remainder of *Ficus* except sect. *Pharmacosycea*. Aside from these two clades, the remaining *Ficus* taxa formed three clades (clades B, C, and D in Fig 8), with unstable support (SH-aLRT = 64.9 and MLBS = 73). Clade B (SH-aLRT = 99.9 and MLBS = 100) was comprised of five different subgenera, while clade C (SH-aLRT = 100 and MLBS = 100) included all species of subg. *Sycidium* as well as members of each of the other five subgenera. Clade D (SH-aLRT = 100 and MLBS = 100) contained only two species in subg. *Pharmacosycea* sect. *Oreosycea*. On the whole, all six subgenera were found not to be monophyletic.

With the inclusion of eleven additional samples, a phylogenomic analysis was carried out on 24 samples representing the 13 taxa in the *F. sarmentosa* complex (Fig 8). The results showed that the *F. sarmentosa* complex failed to form a monophyletic group. Except three samples which are unexpectedly embedded in a distinct clade with other members of subg. *Synoecia* (clade E), two distinct lineages were recognized, with the majority of nodes within these two lineages being well-supported (clades F and G). The six *F. sarmentosa* varieties were found to be scattered across clades E, F, and G, and only two individuals of *F. sarmentosa* var. *lacrymans* clustered together. The six *F. pumila* samples were also found not to be a monophyletic group embedded by *F. sarmentosa* var. *thunbergii*.
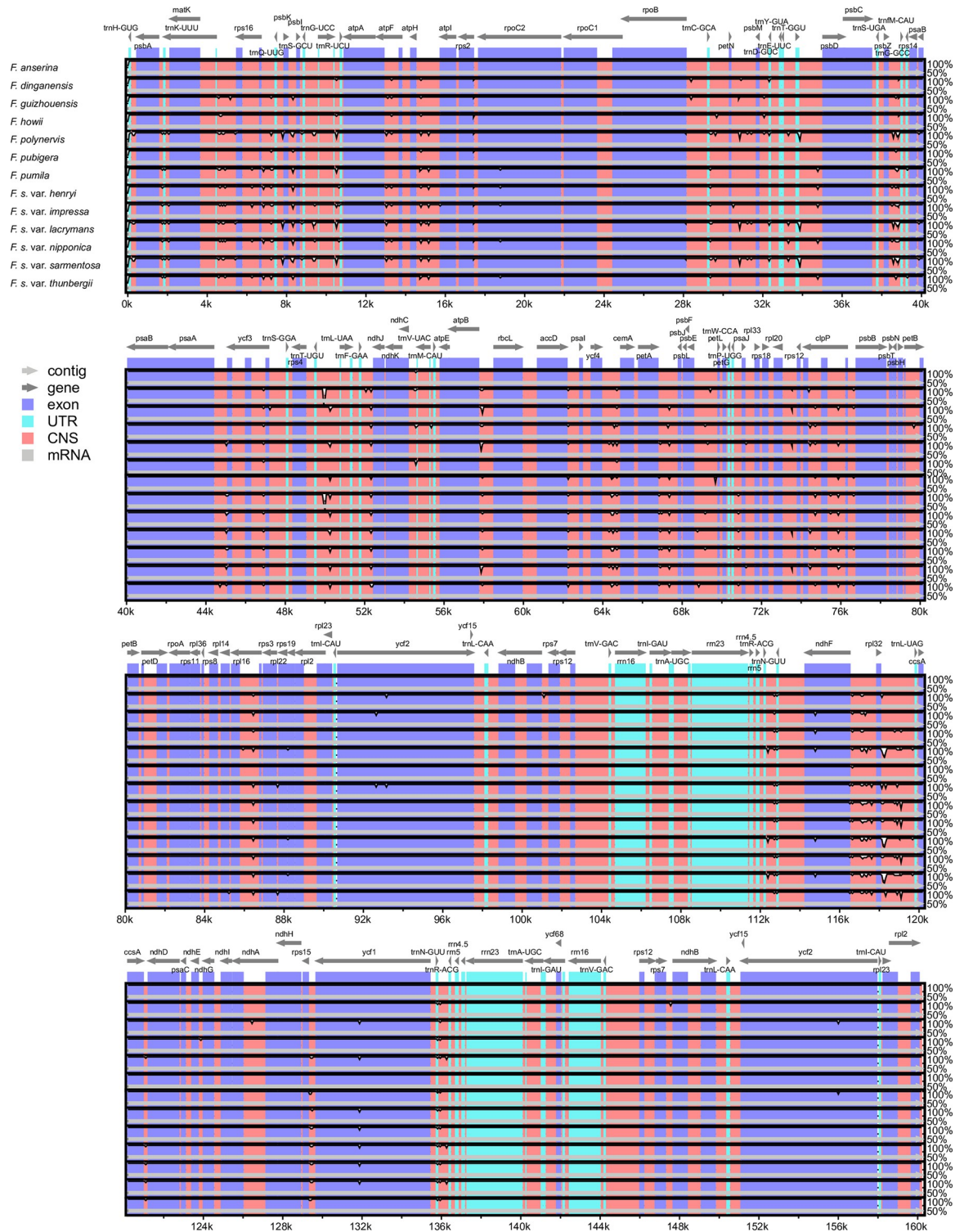
**Fig 5. Comparison of complete chloroplast genomes among 13 taxa in the *F. sarmentosa* complex with *F. anserina* as a reference.** Thick, gray arrows above the alignment indicate the orientation and position of each gene. A cut-off of 70% identity was chosen for the plots. The Y-axis represents the identity percentage, ranging from 50 to 100%.

https://doi.org/10.1371/journal.pone.0279849.g005
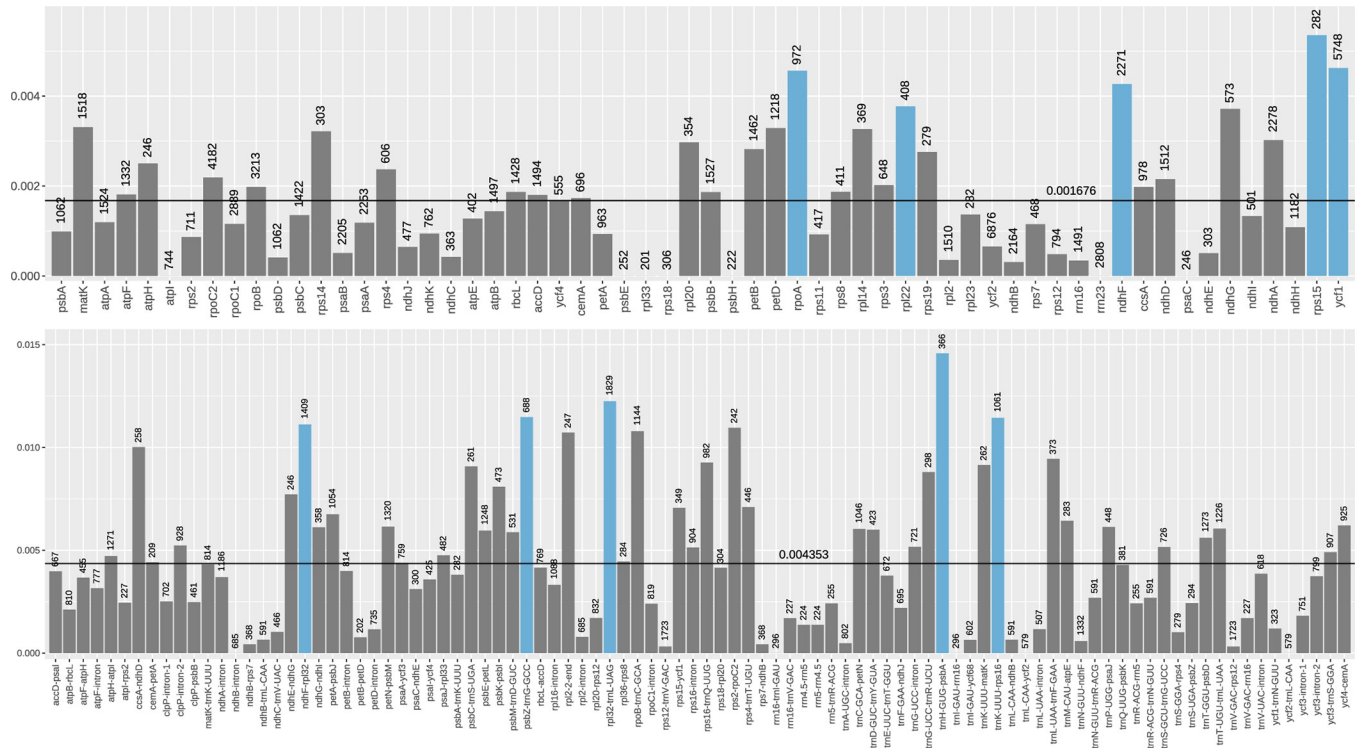
**Fig 6. Nucleotide diversity of genes and intergenic spacer regions among 13 taxa in the *F. sarmetosa* complex.** The alignment lengths are indicated on the bars. The horizontal lines indicate the average nucleotide diversity of genes and intergenic spacer regions, respectively. The top five genes or intergenic spacers with the highest nucleotide diversity are highlighted in blue.

https://doi.org/10.1371/journal.pone.0279849.g006

## Discussion

### The differentiation and diversity of the chloroplast genome in the *F. sarmentosa* complex

To date, vast comparative chloroplast genomic studies have been conducted in a wide range of taxonomic levels, such as order (such as Dipsacales [56] and Saxifragales [57]), family (such as
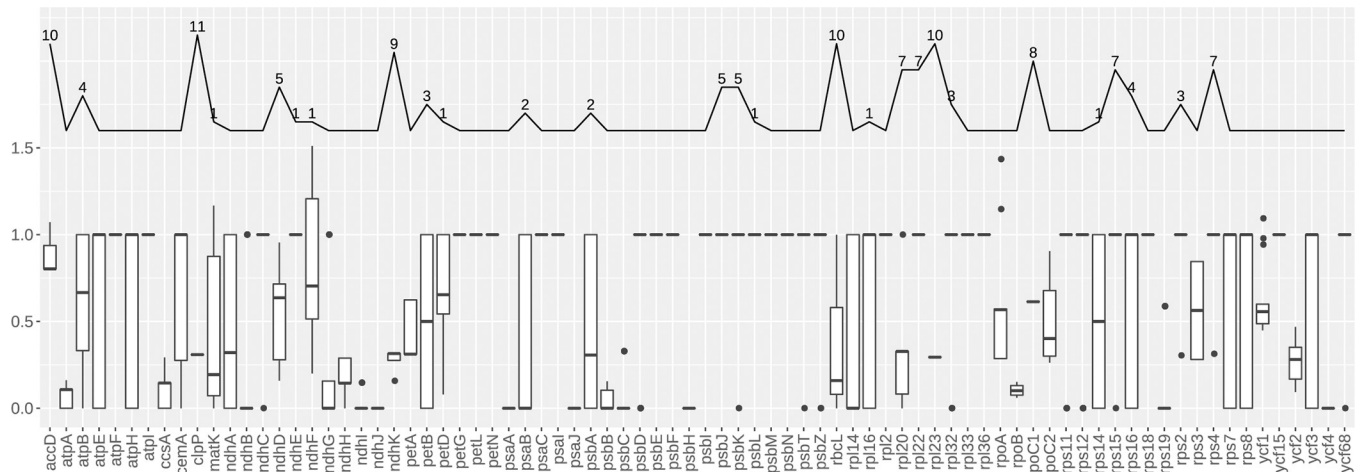


**Fig 7. Boxplot of Ka/Ks ratios for 80 unique CDS regions.** The value 1.0 represents the situation where both Ka and Ks equal zero. The line chart superimposed upon the boxplot demonstrates the frequency of infinite Ka/Ks ratios (Ka > 0 and Ks = 0), with detailed numbers labeled simultaneously.

https://doi.org/10.1371/journal.pone.0279849.g007
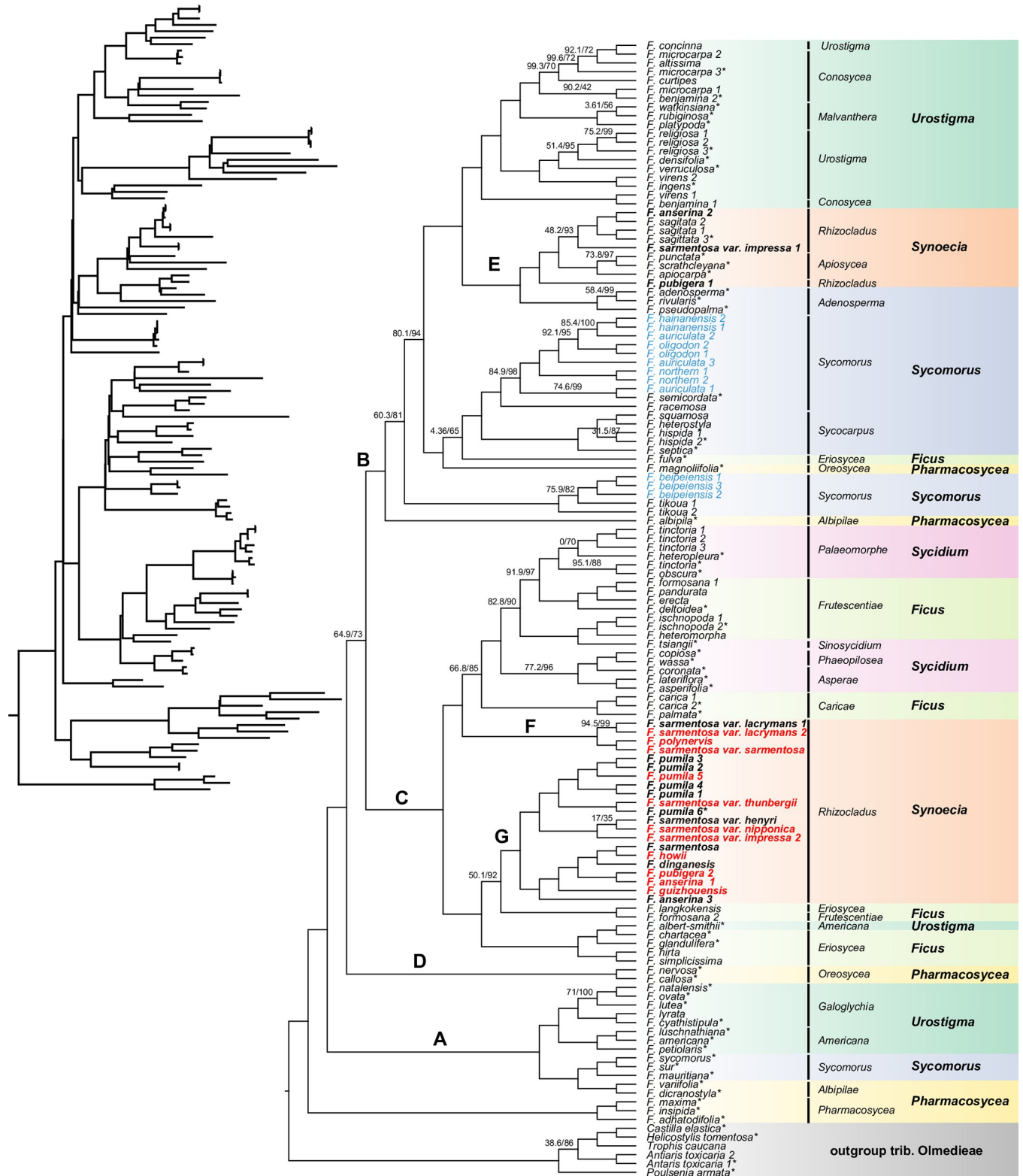
**Fig 8. The maximum likelihood (ML) phylogenetic tree of 123 chloroplast genomes in *Ficus* with six Olmedieae genomes as the outgroup.** Only the branches with either SH-aLRT or ultrafast bootstrap < 95% were annotated by corresponding values. The starred tip names indicate genomes obtained from Bruun-Lund et al. [26]; the red names indicate genomes obtained in this study; the blue names indicate members of the *F. auriculata* complex; and the bold names indicate members of the *F. sarmentosa complex*. The subgenus and section division of *Ficus* are annotated to the right of tip names. The topology of the ML tree is shown in the upper left corner (excluding the outgroup).

Orchidaceae [58] and Zingiberaceae [59]), and genus (such as *Camellia* [60], *Lindera* [61], *Gossypium* [62], and *Ficus* [17, 51, 63]). However, less research has focused on the comparative genomics of taxa undergoing recent speciation, such as the species complex [29]. Although the structural conservation of the chloroplast genome at low taxonomic levels is well-characterized [64–67], the detailed patterns of genomic differentiation and diversity among closely related species remain largely unknown. Therefore, comprehensive comparisons between closely related species are necessary to improve our understanding of the mechanisms, rates, or directionality of genome evolution during the early stages after speciation [29].

In this study, we investigated the evolutionary dynamics of thirteen high-quality chloroplast genomes from the *F. sarmentosa* complex. Overall, the lengths of both whole-genome and quadripartite regions were quite similar among taxa, with only 0.2288%, 0.3784%, 0.2188%, and 0.1236% variation among whole genomes, LSCs, SSCs, and IRs, respectively (Table 1, Fig 1). Furthermore, the number, content, and orientation of annotated genes among all 13 genomes were identical. The IR boundaries (JLB, JSB, JSA, and JLA) were also relatively coincident among the 13 genomes, being located at the same loci with only slight variation in the distance to the starting or ending bases. For long repeat sequences, all four repeat units were shared among all 13 plastomes, and the number of repeat units and their proportions exhibited only slight differentiation. For example, the proportion of palindromic sequences ranged from 45.16 to 61.54% (Fig 3). Similarly, in SSR regions, the proportion of the mononucleotide repeat units ranged from 72.63 to 85.96% (Fig 4).

The high conservation exhibited across the *F. sarmentosa* complex is consistent with other studies on closely related taxa. For example, a study of 22 closely related *Oryza* species indicated that conservation was common at lower taxonomic levels [29]. Even in morphologically diverse shrub willows (*Salix*), such high conservation is still exhibited [68]. A study of four peanut varieties serves as a more extreme example, reporting perfectly identical IR boundary junction positions [69]. Although comparative chloroplast genomic studies have rarely focused on closely related species, the high conservation of chloroplast genome among closely related species is recognized based on our work and other related research.

## Chloroplast genomic evolutionary hotspots

Although chloroplast genomes exhibit high conservation among closely related species, discrepancy and heterogeneity have also been widely observed across the whole genomes. Overall, the variable sites of the two single-copy regions (LSC and SSC) are more abundant than IR regions in both the genus *Ficus* (Fig 5) and most other plant groups [56, 70, 71]. Chloroplast regions with different mutation rates are appropriate for a range of evolutionary research. In general, conservative coding genes are suitable for deep phylogenetic inferences at the family-level [28, 72, 73] or higher [74–76], whereas highly variable regions are appropriate for studies of biogeography, species delimitation, population genetics, and phylogenetic reconstruction at lower infra-generic levels [77, 78].

To date, more than 20 regions have been recommended as alternative loci for phylogenetics, species delimitation, and barcoding, including *matK*, *rbcL*, *trnH-psbA*, *ycf1*, *ycf1-ndhF*, among others [78–80]. However, these loci had few informative sites in the *F. sarmentosa* complex (Fig 6), suggesting that evolutionary heterogeneity of chloroplast loci may be relatively common among different plant groups. In this study, we mined five hyper-variable intergenic regions at the level of the species complex, i.e., *trnH-GUG-psbA*, *rpl32-trnL-UAG*, *psbZ-trn*G-*GCC*, *trnK-UUU-rps16* and *ndhF-rpl32*. However, other studies of comparative chloroplast genomics in *Ficus* identified entirely different hyper-variable intergenic regions. For example, Xia et al. [63] identified *trnS-GCU-trnG-UCC*, *trnT-GGU-psbD*, *trnV-UAC-trnM-CAU*, *clpP-*

*psbB*, *ndhF-trnL-UAG*, *trnL-UAG-ccsA*, *ndhD-psaC*, and *ycf1*, and Zhang et al. [17] identified t*rnL-UAG-rpl32*, *trnE-UUC-psbD*, *trnK-UUU-rps16*, *rpoB-trnC-GCA*, and *petN-psbM*. These disparate results suggest that the evolutionary dynamics of chloroplast genomes varies across both groups and taxonomic ranks. Therefore, we suggest that chloroplast loci should be chosen cautiously according to the research objective, plant group, and taxonomic level.

## The plastome phylogeny of the genus *Ficus* and the phylogenetic performance in closely related species

Based on a compilation of nearly all available *Ficus* chloroplast genomes, we obtained a robust ML phylogenetic tree with the majority of nodes exhibiting high bootstrap and SH-aLRT values, particularly the deep nodes (Fig 8). Overall, our ML phylogenetic tree is largely consistent with previous research [17, 26], including the systematic position of subg. *Pharmacosycea* sect. *Pharmacosycea* and the mysterious displacement of certain individuals (*F. fulva*, *F. magnoliifolia*, *F. albipila*, *F. albert-smithii*, and *F. pumila*). However, analysis of our extended dataset highlighted an increase in non-monophyletic groups, such as subg. *Urostigma* sect. *Urostigma* and *Conosycea* (Fig 8). Notably, there were numerous incongruences between the chloroplast cladogram and previously published nuclear trees. For example, the chloroplast-based tree failed to support either *Ficus* subg. *Sycomorus* or subg. *Sycidium* as monophyletic groups, whereas these groups are well-confirmed in nuclear phylogenies [3, 12–14, 81]. Subg. *Synoecia* was also divided into three different clades (Fig 8). Further research into the displaced species, such as *F. fulva*, *F. ablert-smithii*, *F. magnoliilolia*, and *F. albipila*, may reveal host shifts or nonspecific pollination between fig trees and fig wasps [26]. Considering that a stable and comprehensively-sampled nuclear phylogenomic framework for the genus *Ficus* is still lacking, precise identification of these hybridization events will require more robust nuclear genome data as well as data from the associated fig wasps.

Compared to the nuclear phylogeny [4], we discovered a disparate evolutionary relationship among taxa in the *F. sarmentosa* complex, including three distinct clades (Fig 8, clades E, F, and G). Three samples from the complex dispersed into the clade E mixing with *F. sagittata* and the members of subg. *Synoecia* sect. *Apiosycea*. Unless more data support hybridization, misidentification may be an alternative explanation. Neither geographic nor morphological traits could be detected to support the split of clades F and G. Hybridization between *F. pumila* and *F. sarmentosa* var. *thunbergii* might exist, considering that the latter embedded into the former within clade G. Although our current samples are insufficient to fully resolve the taxonomy of the *F. sarmentosa* complex, the phylogenetic resolution provided by chloroplast genomes in the complex appears to be promising, as almost all the nodes are strongly supported. The chloroplast genome has also shown high discriminability across the *F. auriculata* complex (Fig 8, blue labels) [82–84]. The relationships between four taxa (*F. auriculata*, *F. oligodon*, *F. hainanensis*, and *F. beipeiensis*) in the complex were well-resolved, while *F. beipeiensis* shared a distinct phylogenetic relationship with the climbing fig tree *F. tikoua*. Moreover, a new linage (*F. northern*) was identified based on chloroplast genomics, suggesting a promising segue into further exploration of the cryptic species (Fig 8, blue labels) [17]. Chloroplast genomes have been used to reconstruct high-resolution phylogenetic trees in other closely related species groups, such as peanut [85], rice [29], willow [68], and orchardgrass [86]. Taken together, the chloroplast genome appears to be a promising tool for exploring the evolutionary relationships between closely related species and even species complexes, although cyto-nuclear discordances often exist.

## Conclusions

In this study, eleven *F. sarmentosa* complex chloroplast genomes were newly sequenced and characterized. Sequence lengths, IR boundaries, repeat sequences, and codon usage were

compared among these eleven, and two previously-reported, chloroplast genomes, indicating that these parameters were highly conserved across taxa. However, heterogeneity was found in both nucleotide diversity and selective pressure among segments. We characterized ten evolutionary hotspot regions (*rps15*, *ycf1*, *rpoA*, *ndhF*, *rpl22*, *trnH*-GUG-*psbA*, *rpl32-trnL-UAG*, *psbZ-trnG-GCC*, *trnK-UUU-rps*16 and *ndhF-rpl32*). Phylogenomic analysis indicated that chloroplast genomes show promise for inferring the phylogenetic relationships between closely related groups, despite cyto-nuclear discordance.

## Supporting information

**S1 Table. The collecting information of samples in the study.**
(XLSX)

**S2 Table. Detailed information for the high-throughput sequencing data in the study.**
(XLSX)

## Acknowledgments

## Author Contributions

**Conceptualization:** De-Shun Zhang, Chi-Yuan Yao.

**Data curation:** Lu Zou.

**Formal analysis:** Zhen Zhang.

**Funding acquisition:** Chi-Yuan Yao.

**Investigation:** Zhen Zhang.

**Methodology:** Zhen Zhang.

**Project administration:** Chi-Yuan Yao.

**Resources:** Chi-Yuan Yao.

**Software:** Zhen Zhang.

**Visualization:** De-Shun Zhang.

**Writing – original draft:** Zhen Zhang, Lu Zou.

**Writing – review & editing:** Zhen Zhang.

## References

1. Berg CC, Corner EJH. Moraceae (*Ficus*). In: Nooteboom HP, editor. Flora Malesiana. 17. Leiden: National Herbarium of the Netherlands; 2005. p. 1–730.

2. Clement WL, Weiblen GD. Morphological evolution in the mulberry family (Moraceae). Syst Bot. 2009; 34(3):530–52. https://doi.org/10.1600/036364409789271155

3. Clement WL, Bruun-Lund S, Cohen A, Kjellberg F, Weiblen GD, Rønsted N. Evolution and classification of figs (*Ficus*, Moraceae) and their close relatives (Castilleae) united by involucral bracts. Bot J Linn Soc. 2020; 193(3):316–39. https://doi.org/10.1093/botlinnean/boaa022

4. 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; 69(5):927–45. https://doi.org/10.1002/tax.12282

5. Machado CA, Robbins N, Gilbert MTP, Herre EA. Critical review of host specificity and its coevolutionary implications in the fig/fig-wasp mutualism. P Natl Acad Sci USA. 2005; 102(suppl 1):6558–65. https://doi.org/10.1073/pnas.0501840102 PMID: 15851680

6. Wachi N, Kusumi J, Tzeng HY, Su ZH. Genome-wide sequence data suggest the possibility of pollinator sharing by host shift in dioecious figs (Moraceae, *Ficus*). Mol Ecol. 2016; 25(22):5732–46. https://doi.org/10.1111/mec.13876 PMID: 27706883

7. Wang G, Cannon CH, Chen J. Pollinator sharing and gene flow among closely related sympatric dioecious fig taxa. P Roy Soc B-Biol Sci. 2016; 283(1828):20152963. https://doi.org/10.1098/rspb.2015.2963 PMID: 27075252

8. Yu H, Liao YL, Cheng YF, Jia YJ, Compton SG. More examples of breakdown the 1:1 partner specificity between figs and fig wasps. Bot Stud. 2021; 62(1):1–12. https://doi.org/10.1186/s40529-021-00323-8 PMID: 34626257

9. Yu H, Zhao NX, Yao J, Chen YZ. The influence of species-specific coevolution of figs and fig wasps on *Ficus* (Moraceae) classification. J Trop Subtrop Bot. 2006; 14(5):439–43. https://doi.org/10.3969/j.issn.1005-3395.2006.5.014

10. Rønsted N, Weiblen GD, Cook JM, Salamin N, Machado CA, Savolainen V. 60 million years of co-divergence in the fig–wasp symbiosis. P Roy Soc B-Biol Sci. 2005; 272(1581):2593–9. https://doi.org/10.1098/rspb.2005.3249 PMID: 16321781

11. Rønsted N, Weiblen GD, Clement W, Zerega N, Savolainen V. Reconstructing the phylogeny of figs (*Ficus*, Moraceae) to reveal the history of the fig pollination mutualism. Symbiosis. 2008; 45:45–55.

12. 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):546–57. https://doi.org/10.1111/j.1759-6831.2011.00155.x

13. Cruaud A, Rønsted N, Chantarasuwan B, Chou LS, Clement WL, Couloux A, et al. An extreme case of plant–insect codiversification: figs and fig-pollinating wasps. Syst Biol. 2012; 61(6):1029–47. https://doi.org/10.1093/sysbio/sys068 PMID: 22848088

14. Zhang Q, Onstein RE, Little SA, Sauquet H. Estimating divergence times and ancestral breeding systems in *Ficus* and Moraceae. Ann Bot. 2019; 123(1):191–204. https://doi.org/10.1093/aob/mcy159 PMID: 30202847

15. Zhang XT, Wang G, Zhang SC, Chen S, Wang YB, Wen P, et al. Genomes of the banyan tree and pollinator wasp provide insights into fig-wasp coevolution. Cell. 2020; 183(4):875–89. e17. https://doi.org/10.1016/j.cell.2020.09.043 PMID: 33035453

16. Rasplus JY, Rodriguez LJ, Sauné L, Peng YQ, Bain A, Kjellberg F, et al. Exploring systematic biases, rooting methods and morphological evidence to unravel the evolutionary history of the genus *Ficus* (Moraceae). Cladistics. 2021; 37(4):402–22. https://doi.org/10.1111/cla.12443 PMID: 34478193

17. Zhang ZR, Yang X, Li WY, Peng YQ, Gao J. Comparative chloroplast genome analysis of *Ficus* (Moraceae): Insight into adaptive evolution and mutational hotspot regions. Front Plant Sci. 2022; 13:1–17. https://doi.org/10.3389/fpls.2022.965335 PMID: 36186045

18. Soltis DE, Albert VA, Savolainen V, Hilu K, Qiu YL, Chase MW, et al. Genome-scale data, angiosperm relationships, and 'ending incongruence': a cautionary tale in phylogenetics. Trends Plant Sci. 2004; 9(10):477–83. https://doi.org/10.1016/j.tplants.2004.08.008 PMID: 15465682

19. Jansen RK, Cai ZQ, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. P Natl Acad Sci USA. 2007; 104(49):19369–74. https://doi.org/10.1073/pnas.0709121104 PMID: 18048330

20. Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020; 21(1):1–31. https://doi.org/10.1186/s13059-020-02154-5 PMID: 32912315

21. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017; 45(4):e18–e. https://doi.org/10.1093/nar/gkw955 PMID: 28204566

22. Yang JB, Tang M, Li HT, Zhang ZR, Li DZ. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. BMC Evol Biol. 2013; 13(1):1–12. https://doi.org/10.1186/1471-2148-13-84 PMID: 23597078

23. Li XW, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen SL. Plant DNA barcoding: from gene to genome. Biol Rev. 2015; 90(1):157–66. https://doi.org/10.1111/brv.12104 PMID: 24666563

**24.** Tonti-Filippini J, Nevill PG, Dixon K, Small I. What can we do with 1000 plastid genomes? Plant J. 2017; 90(4):808–18. https://doi.org/10.1111/tpj.13491 PMID: 28112435

**25.** Zhang Z, Zhang DS. The complete chloroplast genome sequence of *Ficus sarmentosa* (Moraceae, Rosales), a widely distributed fig tree in East Asia. Mitochondrial DNA B. 2022; 7(9):1597–8. https://doi.org/10.1080/23802359.2022.2115860 PMID: 36106191

**26.** 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). Mol Phylogenet Evol. 2017; 109:93–104. https://doi.org/10.1016/j.ympev.2016.12.031 PMID: 28042043

**27.** Roure B, Baurain D, Philippe H. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. Mol Biol Evol. 2013; 30(1):197–214. https://doi.org/10.1093/molbev/mss208 PMID: 22930702

**28.** Zhang SD, Jin JJ, Chen SY, Chase MW, Soltis DE, Li HT, et al. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. New Phytol. 2017; 214(3):1355–67. https://doi.org/10.1111/nph.14461 PMID: 28186635

**29.** Gao LZ, Liu YL, Zhang D, Li W, Gao J, Liu Y, et al. Evolution of *Oryza* chloroplast genomes promoted adaptation to diverse ecological habitats. Commun Biol. 2019; 2(1):1–13. https://doi.org/10.1038/s42003-019-0531-2 PMID: 31372517

**30.** Liu ZF, Ma H, Ci XQ, Li L, Song Y, Liu B, et al. Can plastid genome sequencing be used for species identification in Lauraceae? Bot J Linn Soc. 2021; 197(1):1–14. https://doi.org/10.1093/botlinnean/boab018

**31.** Berg CC. Flora Malesiana precursor for the treatment of Moraceae 4: *Ficus* subgenus *Synoecia*. Blumea. 2003; 48(3):551–71. https://doi.org/10.3767/000651903X489546

**32.** Corner EJH. Taxonomic notes on *Ficus* L., Asia and Australasia. Gard Bull Singapore. 1960; 18(1):1–69.

**33.** Corner EJH. Check-list of *Ficus* in Asia and Australasia with keys to identification. Gard Bull Singapore. 1965; 21:1–186.

**34.** Chang SS. Three new species of the Moraceae from China. Acta Phytotax Sin. 1982; 20(1):95–8.

**35.** Chang SS. New taxa of Moraceae from China. Guihaia. 1983; 3(4):295–306.

**36.** Chang SS. New Taxa of Moraceae from China and Vietnam. Acta Phytotax Sin. 1984; 22(1):64–76.

**37.** Chang SS, Wu CY, Cao ZY. Moraceae (*Ficus*). In: Chang SS, Wu CY, editors. Flora Reipublicae Popularis Sinicae. 23. Beijing: Science Press; 1998. p. 66–169.

**38.** Yamazaki T. Taxonomic review of the Moraceae from Japan, Korea, Taiwan and adjacent areas (2). J Phytogeo Taxon. 1983; 31(1):1–15.

**39.** Tsai L, Hayakawa H, Fukuda T, Yokoyama J. A breakdown of obligate mutualism on a small island: An interspecific hybridization between closely related fig species (*Ficus pumila* and *Ficus thunbergii*) in Western Japan. Amer J Pl Sci. 2015; 6(01):126–31. https://doi.org/10.4236/ajps.2015.61014

**40.** Berg CC. Precursory taxonomic studies on *Ficus* (Moraceae) for the Flora of Thailand. Thai Forest Bull. 2007; 35:4–28.

**41.** Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987; 19:11–5.

**42.** Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 2019; 15(1):1–12. https://doi.org/10.1186/s13007-019-0435-7 PMID: 31139240

**43.** Greiner S, Lehwark P, Bock R. Organellar Genome DRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019; 47(W1):W59–W64. https://doi.org/10.1093/nar/gkz238 PMID: 30949694

**44.** Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018; 34(17):3030–1. https://doi.org/10.1093/bioinformatics/bty220 PMID: 29659705

**45.** 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 Res. 2001; 29(22):4633–42. https://doi.org/10.1093/nar/29.22.4633 PMID: 11713313

**46.** Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017; 33(16):2583–5. https://doi.org/10.1093/bioinformatics/btx198 PMID: 28398459

**47.** Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004; 32(suppl_2):W273–W9. https://doi.org/10.1093/nar/gkh458 PMID: 15215394

**48.** Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32(5):1792–7. https://doi.org/10.1093/nar/gkh340 PMID: 15034147

**49.** Xia X. DAMBE 6: new tools for microbial genomics, phylogenetics, and molecular evolution. J Hered. 2017; 108(4):431–7. https://doi.org/10.1093/jhered/esx033 PMID: 28379490

**50.** Chen CJ, Chen H, Zhang Y, Thomas HR, Frank MH, He YH, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020; 13(8):1194–202. https://doi.org/10.1016/j.molp.2020.06.009 PMID: 32585190

**51.** Huang YY, Li J, Yang ZR, An WL, Xie CZ, Liu SS, et al. Comprehensive analysis of complete chloroplast genome and phylogenetic aspects of ten *Ficus* species. BMC Plant Biol. 2022; 22(1):1–15. https://doi.org/10.1186/s12870-022-03643-4 PMID: 35606691

**52.** Gardner EM, Garner M, Cowan R, Dodsworth S, Epitawalage N, Arifiani D, et al. Repeated parallel losses of inflexed stamens in Moraceae: phylogenomics and generic revision of the tribe Moreae and the reinstatement of the tribe Olmedieae (Moraceae). Taxon. 2021:946–88. https://doi.org/10.1002/tax.12526

**53.** Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30(4):772–80. https://doi.org/10.1093/molbev/mst010 PMID: 23329690

**54.** Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009; 25(15):1972–3. https://doi.org/10.1093/bioinformatics/btp348 PMID: 19505945

**55.** Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017; 14(6):587–9. https://doi.org/10.1038/nmeth.4285 PMID: 28481363

**56.** Fan WB, Wu Y, Yang J, Shahzad K, Li ZH. Comparative chloroplast genomics of Dipsacales species: insights into sequence variation, adaptive evolution, and phylogenetic relationships. Front Plant Sci. 2018; 9:1–13. https://doi.org/10.3389/fpls.2018.00689 PMID: 29875791

**57.** Dong WP, Xu C, Cheng T, Zhou SL. Complete chloroplast genome of Sedum sarmentosum and chloroplast genome evolution in Saxifragales. PloS One. 2013; 8(10):e77965. https://doi.org/10.1371/journal.pone.0077965 PMID: 24205047

**58.** Luo J, Hou BW, Niu ZT, Liu W, Xue QY, Ding XY. Comparative chloroplast genomes of photosynthetic orchids: insights into evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. PloS One. 2014; 9(6):e99016. https://doi.org/10.1371/journal.pone.0099016 PMID: 24911363

**59.** Gao BM, Yuan L, Tang TL, Hou J, Pan K, Wei N. The complete chloroplast genome sequence of *Alpinia oxyphylla* Miq. and comparison analysis within the Zingiberaceae family. PloS One. 2019; 14(6): e0218817. https://doi.org/10.1371/journal.pone.0218817 PMID: 31233551

**60.** Yang JB, Yang SX, Li HT, Yang J, Li DZ. Comparative chloroplast genomes of *Camellia* species. PloS One. 2013; 8(8):e73053. https://doi.org/10.1371/journal.pone.0073053 PMID: 24009730

**61.** Zhao ML, Song Y, Ni J, Yao X, Tan YH, Xu ZF. Comparative chloroplast genomics and phylogenetics of nine *Lindera* species (Lauraceae). Sci Rep. 2018; 8(1):1–11. https://doi.org/10.1038/s41598-018-27090-0 PMID: 29891996

**62.** Wu Y, Liu F, Yang DG, Li W, Zhou XJ, Pei XY, et al. Comparative chloroplast genomics of *Gossypium* species: insights into repeat sequence variations and phylogeny. Front Plant Sci. 2018; 9:376. https://doi.org/10.3389/fpls.2018.00376 PMID: 29619041

**63.** Xia X, Peng JY, Yang L, Zhao XL, Duan AA, Wang DW. Comparative Analysis of the Complete Chloroplast Genomes of Eight *Ficus* Species and Insights into the Phylogenetic Relationships of *Ficus*. Life. 2022; 12(6):848. https://doi.org/10.3390/life12060848 PMID: 35743879

**64.** Palmer JD. Comparative organization of chloroplast genomes. Annu Rev Genet. 1985; 19(1):325–54. https://doi.org/10.1146/annurev.ge.19.120185.001545 PMID: 3936406

**65.** Sugiura M. The chloroplast genome. Plant Mol Biol. 1992:149–68. https://doi.org/10.1007/BF00015612 PMID: 1600166

**66.** Jansen RK, Ruhlman TA. Plastid genomes of seed plants. In: Bock R, Knoop V, editors. Genomics of Chloroplasts and Mitochondria. Dordrecht: Springer; 2012. p. 103–26.

**67.** Mower JP, Vickrey TL. Structural diversity among plastid genomes of land plants. Adv Bot Res. 2018; 85:263–92. https://doi.org/10.1016/bs.abr.2017.11.013

**68.** Wagner ND, Volf M, Hörandl E. Highly diverse shrub willows (*Salix* L.) share highly similar plastomes. Front Plant Sci. 2021:1740. https://doi.org/10.3389/fpls.2021.662715 PMID: 34539686

69. Wang J, Li CJ, Yan CX, Zhao XB, Shan SH. A comparative analysis of the complete chloroplast genome sequences of four peanut botanical varieties. PeerJ. 2018; 6:e5349. https://doi.org/10.7717/peerj.5349 PMID: 30083466

70. Li XQ, Zuo YJ, Zhu XX, Liao S, Ma JS. Complete chloroplast genomes and comparative analysis of sequences evolution among seven *Aristolochia* (Aristolochiaceae) medicinal species. Int J Mol Sci. 2019; 20(5):1045. https://doi.org/10.3390/ijms20051045 PMID: 30823362

71. Guo XX, Qu XJ, Zhang XJ, Fan SJ. Comparative and phylogenetic analysis of complete plastomes among Aristidoideae species (Poaceae). Biology. 2022; 11(1):63. https://doi.org/10.3390/biology11010063 PMID: 35053061

72. Zhang R, Wang YH, Jin JJ, Stull GW, Bruneau A, Cardoso D, et al. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. Syst Biol. 2020; 69(4):613–22. https://doi.org/10.1093/sysbio/syaa013 PMID: 32065640

73. Liu LM, Du XY, Guo C, Li DZ. Resolving robust phylogenetic relationships of core Brassicaceae using genome skimming data. J Syst Evol. 2021; 59(3):442–53. https://doi.org/10.1111/jse.12666

74. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. P Natl Acad Sci USA. 2010; 107(10):4623–8. https://doi.org/10.1073/pnas.0907801107 PMID: 20176954

75. Gitzendanner MA, Soltis PS, Wong GKS, Ruhfel BR, Soltis DE. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. Am J Bot. 2018; 105(3):291–301. https://doi.org/10.1002/ajb2.1048 PMID: 29603143

76. Li HT, Luo Y, Gan L, Ma PF, Gao LM, Yang JB, et al. Plastid phylogenomic insights into relationships of all flowering plant families. BMC Biology. 2021; 19(1):1–13. https://doi.org/10.1186/s12915-021-01166-2 PMID: 34711223

77. Dong WP, Liu J, Yu J, Wang L, Zhou SL. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PloS One. 2012; 7(4):e35071. https://doi.org/10.1371/journal.pone.0035071 PMID: 22511980

78. Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. Am J Bot. 2014; 101(11):1987–2004. https://doi.org/10.3732/ajb.1400398 PMID: 25366863

79. Dong WP, Xu C, Li CH, Sun JH, Zuo YJ, Shi S, et al. *ycf1*, the most promising plastid DNA barcode of land plants. Sci Rep. 2015; 5(1):1–5. https://doi.org/10.1038/srep08348 PMID: 25672218

80. Amar MH. *ycf1-ndhF* genes, the most promising plastid genomic barcode, sheds light on phylogeny at low taxonomic levels in *Prunus persica*. J Genet Eng Biotechnol. 2020; 18(42):1–10. https://doi.org/10.1186/s43141-020-00057-3 PMID: 32797323

81. Pederneiras LC, Gaglioti AL, Romaniuc-Neto S, Mansano VF. The role of biogeographical barriers and bridges in determining divergent lineages in *Ficus* (Moraceae). Bot J Linn Soc. 2018; 187(4):594–613. https://doi.org/10.1093/botlinnean/boy036

82. Wei ZD, Kobmoo N, Cruaud A, Kjellberg F. Genetic structure and hybridization in the species group of *Ficus auriculata*: can closely related sympatric *Ficus* species retain their genetic identity while sharing pollinators? Mol Ecol. 2014; 23(14):3538–50. https://doi.org/10.1111/mec.12825 PMID: 24938182

83. Zhang LF, Zhang Z, Wang XM, Gao HY, Tian HZ, Li HQ. Molecular phylogeny of the *Ficus auriculata* complex (Moraceae). Phytotaxa. 2018; 362(1):039–54. https://doi.org/10.11646/phytotaxa.362.1.3

84. Zhang Z, Wang XM, Liao S, Tian HZ, Li HQ. Taxonomic treatment of the *Ficus auriculata* complex (Moraceae) and typification of some related names. Phytotaxa. 2019; 399(3):203–8. https://doi.org/10.11646/phytotaxa.399.3.4

85. Wang J, Li Y, Li C, Yan C, Zhao X, Yuan C, et al. Twelve complete chloroplast genomes of wild peanuts: great genetic resources and a better understanding of *Arachis* phylogeny. BMC Plant Biol. 2019; 19(1):1–18. https://doi.org/10.1186/s12870-019-2121-3 PMID: 31744457

86. Jiao YJ, Feng GY, Huang LK, Nie G, Li Z, Peng Y, et al. Complete Chloroplast Genomes of 14 Subspecies of *D. glomerata*: Phylogenetic and Comparative Genomic Analyses. Genes. 2022; 13(9):1621. https://doi.org/10.3390/genes13091621 PMID: 36140789