Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Characterization of Firmiana danxiaensis chloroplast genomes and comparative analysis of Firmiana: insight into its adaptive evolution and phylogenetic relationships

Ya-li Li
Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences
Li-yun Nie
Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences
Shuang-wen Deng
Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences
Lei Duan
Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences
Zheng-feng Wang
Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems
Joseph L.M. Charboneau
University of Michigan–Ann Arbor
Boon-Chuan Ho
Singapore Botanic Gardens, National Parks Board
Hong-feng Chen (h.f.chen@scbg.ac.cn)
Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences

Research Article

Keywords: Adaptive evolution, Chloroplast genome, Comparative genomics, Firmiana danxiaensis, Phylogenetic relationships

Posted Date: June 8th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2918955/v1

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

Abstract

Background

Firmiana danxiaensis is a critically endangered and ecologically important tree currently only found in four locations in Danxia or Karst habitats in northern Guangdong Province, China. The specialized habitat preference makes it an ideal model species for study of adaptive evolution. Therefore, we sequenced its complete chloroplast (cp) genome from four locations and conducted comparative genomics analyses at both interspecific and intrageneric levels.

Results

The *F. danxiaensis* cp genomes are about 160,972 bp in size, with 112 unique genes encoded. The genomes revealed higher biased codon preferences in Karst habitat than those in Danxia habitats. Eighteen and 11 divergent hotpots were identified at interspecific and intrageneric levels, respectively. Selection pressure analysis revealed that Ka/Ks values of *F. danxiaensis* cp genomes were less than one at the interspecific level. Seven (*clpP*, *accD*, *ccsA*, *ndhH*, *rpl20*, *rpoC2*, and *rps4*) positively selected genes were identified when comparing *F. danxiaensis* cp genomes to *Sterculia monosperma* and other *Firmiana* species, respectively. Phylogenetic analysis revealed that *F. danxiaensis* is sister to *F. major* and *F. simplex*. However, our cp genomes did not cluster phylogenetically according to their habitat types.

Conclusions

The *F. danxiaensis* cp genomes reveal both possible environmental associated adaptation and random genetic effects in the species, which encourages further integration of geographical distances, environmental factors, and SNPs on the adaptive evolution study of *F. danxiaensis*.

Introduction

Firmiana Marsili, a genus in the Sterculioideae subfamily of Malvaceae Juss., contains 12–18 deciduous trees or shrubs [1–4], nine of which are native to China [5–7]. Among these *Firmiana* species in China, only *F. simplex* (L.) W. Wight is not endangered [8]. Small population sizes, restricted range, and habitat specificity make them susceptible to losing genetic diversity and to a high risk of extinction. *Firmiana danxiaensis* H. H. Hsue & H. S. Kiu is a tree with significant ornamental (Fig. 1: B, C)and ecological value, which has been listed as a critically endangered (CR) species by IUCN [9]. Hitherto, *F. danxiaensis* is only found in four locations restricted to northern Guangdong Province, China: Danxia Mountain (DX) [10], Nanxiong (NX) City [11], Yingde (YD) City [12], and Shixing (SX) County [13]((Fig. 1: A). The species is naturally distributed on Danxia (DX, NX, SX) (Fig. 1: D)and Karst (YD) landforms (Fig. 1: E), both of which are considered fragile ecosystems [14]. The two landforms are edaphically isolated terrestrial habitat islands of southern China [15]. Soils derived from Danxia and limestone Karst are shallow and wash off

easily [16, 17], while both are rich in mineral elements, such as calcium (Ca), magnesium (Mg) [16] and cadmium (Cd) (Table S1) (unpublished data). In addition, soils of Danxia habitats display lower concentrations of C, N, and P when compared to Karst habitats [16]. Plants in both environments are prone to drought stress due to the poor water storage capacities of Danxia and Karst soils. Harsh environmental conditions may intensify selective forces on driving evolution in plants [15].

Challenging environments may form selective genetic pressures, which leave sign of natural selection in genes involved in adaptive evolution [19]. As an essential organelle, the chloroplast (cp) plays vital roles in plant physiology and development, such as photosynthesis, oxygen release, and amino acid and nucleotide synthesis [20]. In contrast to mitochondrial and nuclear genomes, the cp genome is mainly maternally inherited in plants and has small size, slow evolutionary rate, highly conserved sequence, and many mutation events. These characteristics of the cp genome, combined with modern high-throughput technology, make it an ideal model for species identification, phylogenetic reconstruction, and adaptive evolution analysis [20–23].

Earlier studies of *F. danxiaensis* mainly focus on its distribution, growth [24–27], and cultivation [28–30]. In recent years, some studies have performed phylogenetic [31, 32] and comparative genomics analyses [33] based on *F. danxiaensis* cp genomes. However, only the cp genome of *F. danxiaensis* individuals from the DX location have been sequenced. Little is known about the other three locations (NX, YD, and SX). Moreover, the interspecific and intrageneric phylogenetic relationships of *F. danxiaensis* in four locations are unclear, as is whether *F. danxiaensis* has undergone adaptive evolution in adapting to Danxia and Karst habitats also remains ambiguous. Thus, a comparative genomics study primarily from its cp genomes is urgently needed.

Here, we first sequenced *F. danxiaensis* cp genomes from all four of its known locations and *Firmiana hainanensis*, and combined other available cp genomes of *Firmiana* from GenBank to conduct comparative genomic analyses. The aims of this study were to (1) compare structures of *F. danxiaensis* cp genomes; (2) detect divergent hotpots both within *F. danxiaensis* and among *Firmiana* species; (3) analyze adaptive evolution of *F. danxiaensis* and construct a phylogenetic tree among four locations. These results will provide insights into adaptive evolution, which have significance for further research on population genetics and conservation of genetic diversities.

Results

Chloroplast genome features of *Firmiana danxiaensis*

The *F. danxiaensis* cp genomes (GenBank accession numbers ON872508, ON872509, ON872510 and ON872511 for SX, NX, DX and YD, respectively) showed a typical quadripartite and conserved structure with almost identical length, which ranged from 160,832 to 161,206 bp (Table 1). They are composed of four parts, with the LSC region and SSC region separated by two inverted repeats, IRa and IRb (Fig. 2, Table 1). The overall GC contents of the four cp genomes are the same (36.9%). While the IR

regions contained higher GC content than the LSC and SSC regions, they showed few differences among the four cp genomes (Table 1). The number of genes encoded by all four cp genomes was identical. A total of 129 genes were encoded, of which 112 were unique, including 78 CDS genes, four rRNA genes, and 37 tRNA genes (Table 1, Table 2).

The 112 unique genes were classified into four main categories (Table 2) based on their functions. Seventeen genes were duplicated in IR regions. There were 18 genes containing introns, including 12 proteins encoding genes (*rps12*, *rps16*, *rpl2*, *rpl16*, *petB*, *petD*, *ndhA*, *ndhB*, *clpP*, *ycf3*, *rpoC1,* and *atpF*) and 6 tRNAs (*trnG*-UCC, *trnL*-UAA, *trnK*-UUU, *trnV*-UAC, *trnA*-UGC, and *trnI*-GAU), of which 16 genes have a single intron and two genes (*clpP* and *ycf3*) have two introns (Table 2). Additionally, no genes have been lost or gained in *F. danxianesis* cp genomes relative to other *Firmiana* species.

SSRs and long repeats identification

We found almost identical numbers of SSRs in four *F. danxiaensis* cp genomes. SSRs were located mainly in LSC regions (69.54%), followed by SSC regions (15.85%) and IR regions (14.61%) (Fig. 3A). Additionally, 63.83% of SSRs were located in intergenic spacers, 17.24% in introns and 18.39% in coding regions (Fig. 3B). Mononucleotide repeats (p1) were the most dominant type of repeats (60.92%) among six types of SSRs (Fig. 3C). These mononucleotide repeats were mainly composed of polyadenine (poly A) and polythymine (poly T) repetitions (Table S2).

There were 49 long repeat sequences in each of the *F. danxiaensis* cp genomes from four populations. Palindromic repeats (P) were the most common repeat type (46.94%), followed by forward repeats (F) (46.42%) and reverse repeats (R) (10.20%). Only the YD cp genome had one complementary repeat (C) (Fig. 3D). Among the long repeats, those 30–39 bp in size were the most abundant (79.59%), followed by those 40–49 bp (20.40%). Only the DX cp genome had one repeat in size of 50–60 bp (the longest was 60 bp) (Fig. 3E). Additionally, 61.22% of the long repeats were located in nine coding genes, of which y*cf2* had the most repeats (Fig. 3F).

Codon usage

For the four *F. danxiaensis* cp genomes, a total of 64 types of codons encode 20 amino acids. These genomes from different populations each used an almost identical total number of codons (~53,656). The highest frequency of amino acid was leucine (~ 5,345 per cp genome), and the lowest frequency amino acid was cysteine (~ 1,145 per cp genome). In total, 33 kinds of codons had RSCU > 1 among all four cp genomes (Table 3), and most of them ended with A/U. Additionally, ATG (methionine) and TGG (tryptophan) revealed no bias (RSCU=1).

The highest and lowest RSCU values were recorded for AGA (1.9850) and CGC (0.4780), respectively, both encoding arginine in *F. danxiaensis* cp genome of YD. Meanwhile, the genome from YD had the greatest

number of high RSCU values (42.42%) (Table S3), which had highly preferred codons for arginine (AGA), aspartic acid (GAT), glutamine (CAA), and leucine (TTA).

SNP and indels detection

SNPs and indels in four *F. danxiaensis* cp genomes were compared pairwise with each other. The highest number of SNPs and indels were both found when NX cp genome was compared to DX genome, and the lowest were found when SX cp genome was compared to NX genome (Fig. 4). Additionally, there were 13 genes that presented variations, in which nine genes (*matK*, *atpF*, *rpoC2*, *rpoC1*, *rps3*, *ndhE*, *ndhF*, *ycf1,* and *psaA*) presented non-synonymous variations in at least one pair (Table S4).

Comparative chloroplast genome analysis

Multiple alignments of 11 cp genomes were compared by mVISTA with *F. danxiaensis* from SX as a reference. The results revealed a high degree of sequence similarity across these cp genomes (Fig. 5), which suggested greatly conserved evolution in *Firmiana*. In general, the coding regions showed less divergence than the non-coding regions.

For four *F. danxiaensis* cp genomes, *rps19*, *psbJ*, and *clpP* were divergent sequences in coding regions. There were six divergent sequences (*trnG*-UCC, *atpF*, *trnT*-UGU, *trnN*-GUU, *trnD*-GUC, *trnG*-GCC) in non-coding regions.

When *F. danxiaensis* cp genomes were compared to *F. colorata* (Roxb.) R. Br., *F. kwangsiensis* H. H. Hsue, and *F. pulcherrima* H. H. Hsue,, *rps19* and *rps12* were the divergent sequences in coding regions, and *trnK*-UUU, *trnG*-GCC, and *trnL*-UAG were in non-protein-coding regions. Additionally, when compared to *F. hainanensis* Kosterm.*, F. major* (W. W. Sm.) Hand.-Mazz., and *F. simplex*, only *trnN*-GUU and *trnG*-GCC were divergent sequences in non-protein-coding regions.

Nucleotide diversity (π) among the four new *F. danxiaensis* cp genomes calculated in 100-bp windows ranged from 0 to 0.00222, with a mean of 0.00010. Nucleotide diversity was higher in the SSC region (π = 0.00025) than the LSC region (π = 0.00001), and there was no variation within the IR region (π = 0). We detected five variable regions with relatively high π values. Four of them were in LSC region, including *atpE* (0.00167), *rps3* (0.00153), *atpF* (0.00121), and *rps4* (0.00111); only *ndhE* (0.00222) was in SSC region (Fig. 6).

Nucleotide diversity in 100-bp windows among the 10 *Firmiana* cp genomes ranged from 0 to 0.38272, with a mean of 0.01332. The LSC region had higher nucleotide diversity (π = 0.01751), followed by the SSC region (π = 0.00293) and the IR (0.00031). We detected five variable regions with high π values (≥0.1). All of them were in LSC region, including *trnH*-GUG (0.38272), *trnG*-UCC (0.31339), *trnM*-CAU (0.22857), *trnW*-CCA (0.22222), and *trnL*-UAA (0.12698) (Fig. 6).

Overall, all 10 variable regions found within *F. danxiaensis* and among the *Firmiana* were found in the LSC or SSC regions rather than in the IR, which indicated the IR is more conserved than single-copy regions.

The IR/SC boundaries in four *F. danxiaensis* cp genomes (Fig. 7) were highly consistent except for the boundary between IRb and the SSC (JSB), in which the distance between *ndhF* and and the end of IRb varied from 173–213 bp. The other three IR/SC boundaries showed there were no other expansions or contractions of the IR. In all four *F. danxiaensis* cp genomes, the first 6 bp of *rps19* are found in the IR at the junction between the LSC and IRb (JLB), the first 39 bp of *ycf1* are in the IR at the junction between the SSC and IRa (JSA), and the last two 2 bp of *trnH*-GUG extend into the IR at the junction between IRa and the LSC (JLA).

The IR/SC boundaries in 10 *Firmiana* cp genomes were highly consistent except in *F. colorata* (Fig. 7). At JLB, *rps19* is located entirely within the LSC and its 5' end is 30 bp away from the start of IRb. At JSA, the first 1,077 bp of *ycf1* are in the IR. In *F. colorata*, there is a 33 bp gap between the start of *rpl2* and the outer end of the IR as well as a 125 bp gap between the 3' end of *ndhF* and the inner end of IRb (both distances are shorter than those in the other cp genomes). Among the cp genomes other than *F. colorata*, the distance between the JSB and *ndhF* varied from 135-213 bp. The other five cp genomes also all had the same amount of *rps19* and *trnH*-GUG within the IR as the *F. danxiaensis* cp genomes. *F. pulcherrima* and *F. major* have 36 and 37 bp of *ycf1*, respectively, in the IR.

Selective pressure analysis

When analyses of non-synonymous and synonymous substitution frequencies (Ka/Ks) were performed across eight *Firmiana* cp genomes, Ka/Ks values could be obtained from 54 of 78 protein coding genes (Fig. S1). When only considering four *F. danxiaensis* cp genomes, Ka/Ks values were merely available in four genes (*matK, rpoC2*, *rps4,* and *ycf1*) and they were all less than one (0.11-0.27).

In intrageneric comparisons, the results indicated that only *clpP* show obvious signatures of positive selection (Ka/Ks=3.31, $p<0.05$) when *F. danxiaensis* four populations were pairwise compared to *S. monosperma* (Fig. 8: D). Six other genes (*accD*, *ccsA*, *ndhH*, *rpl20*, *rpoC2,* and *rps4*) evolved faster in *Firmiana* (Ka/Ks>1, $p>0.05$) (Fig. 8). Among them, the Ka/Ks values of four genes (*accD*, *ccsA*, *clpP,* and *rpl20*) were same when *F. danxiaensis* from four populations were pairwise compared to *F. colorata*, *F. kwangsiensis*, *F. major*, and *S. monosperma.* However, the Ka/Ks values of other 3 genes (*ndhH, rps4, and rpoC2*) were different (Fig. 8: A, B, C).

Phylogenetic analyses

The phylogenetic trees inferred from ML and BI based on 13 whole cp genomes shared an identical topology, while BI showed higher support values. All branches of of the BI tree (Fig. 9) had 100% Bayesian

posterior probabilities. According to the topology, 10 *Firmiana* cp genomes were divided into two clades, this first containing *F. colorata, F. kwangsiensis*, *and F. pulcherrima* that is sister to the second clade with the rest of species. Moreover, the four *F. danxiaensis* populations form a clade, in which SX and NX are sister to each other and DX and YD are sister to each other.

Discussion

Features of cp genomes

The four *F. danxiaensis* cp genomes exhibited a typical quadripartite structure and showed slight differences in terms of genome sizes, structure, and GC contents. Similar results were found in six other *Firmiana* plants [5, 33–36] (Table S5), indicating *Firmiana* cp genomes are highly conserved [5].

The GC contents of *F. danxiaensis* cp genomes were relatively low compared to *F. colorata*, *F. hainanensis*, *F. kwangsiensis*, *and F. pulcherrima*, showing same contents with *F. hainanensis* and *F. simplex* (Table S5), which were accordant with the phylogenetic relationship within *Firmiana* (Fig. 9). The closer the phylogenetic relationship with *F. danxiaensis*, the lower GC contents, relatively. On one hand, among seven *Firmiana* species, only *F. danxiaensis* grows on specialized landforms, hence the different GC contents may be caused by different selection [37]. On the other hand, the relatively lower GC contents of *F. danxiaensis* may be related to random mutation [38]. Thus, the relatively lower GC contents of *F. danxiaensis* in *Firmiana* provide insights into the further study of adaptive evolution on *F. danxiaensis*.

Gene gain or loss could be found in cp genomes during evolution [39]. No genes have been lost or gained in *F. danxianesis* cp genomes relative to other *Firmiana* species. Although *infA* was annotated as a protein-coding gene in *F. kwangsiensis* (MN786867) and *F. simplex* (NC_041438), this appears to be in error. In *F. major* (NC_037242), *infA* is annotated as a pseudogene, and the sequences in that region have only a small number of differences in other *Firmiana* cp genomes. The *infA* annotation in the three cp genomes all have a start codon of ATT, and the unannotated *Firmiana* cp genomes have the same sequence at that point in the alignment. ATT is a permitted start codon(Table S6)(Bacteria, Archaea, and plant plastids; https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi), but it is not commonly found (if ever) in plastid genomes. *infA* has been lost multiple times across angiosperms, including in nearly all rosids, and nuclear copies of the gene have been found in several plant genomes [40]. *ycf15* was annotated as a protein-coding gene in *F. kwangsiensis* and *F. major*, but this region is now not considered to be a protein-coding gene [41] and is frequently unannotated in cp genomes. Other *Firmiana* cp genomes have a nearly identical sequence in the *ycf15* region as *F. kwangsiensis* and *F. major*.

Codon usage bias and sequence variation

Codon usage analysis is vital to understand the evolutionary process and selection pressure on genes [43]. For four *F. danxiaensis* cp genomes, 78 genes encoded about 53,000 codons, twice as many as some species [44, 45], which may be related to codon hydrophilicity, natural selection, or gene expression rate [46]. The favored codons (RSCU > 1) tended to end in an A or U base, consistent with a previous study

[47]. Additionally, the most abundant amino acids in four *F. danxiaensis* cp genomes were leucine and cysteine. The same results were also reported in other species [44]. Moreover, it was worth mentioning that RSCU values of YDdiffered greatly from other three locations. The highest RSCU value of any codon was found in this cp genome (1.9850 for AGA encoding for arginine), accounting for the largest proportion (21.67%) of highest RSCU values among 20 amino acids as well (Table S3). Meanwhile, the YD location cp genome was the only one that showed bias in the codons most commonly used for aspartic acid, glutamic acid, glutamine, and lysine. *F. danxiaensis* at the YD location is found on Karst landforms, while the other three populations are all on Danxia landforms. The soil factors concentration in two landforms are different [16], and both landforms are subjected to frequent droughts due to the poor water storage capacities of the soils. Hash environmental conditions may intensify selective forces on driving evolution in plants [15]. Thus, the different RSCU values in YD cp genomes might shed light on further research on adaptive evolution of *F. danxiaensis* to Karst habitat.

Regions enriched with SNPs and indels are considered hypervariable regions in the cp genome [48]. In our study, there were more hypervariable regions when comparing the NX and DX cp genomes than when comparing SX and NX, indicating SX and NX are more genetically closely related than the other two locations, which was accordant with the results of phylogenetic analysis (Fig. 9). While YD and DX also form a clade, the relatively closer relationship than SX and NX may be related to the river near the location. The Jinjiang River flows through Danxia

Mountain and flows into the Zhenjiang River. Then, the Zhenjiang River flows southward and flows into the Beijiang River near Yingde City. The globose seed of *F. danxiaensis* may disperse along the river, which form the gene flow between the DX and YD populations, which provides insights into the study of poplulation genetics on *F. danxiaensis*.

The alignments of *Firmiana* cp genomes indicated that non-coding regions were more divergent than coding regions, correspondent with previous cp genomics studies [49, 51]. For four *F. danxiaensis* cp genomes, nine divergent sequences (*rps19*, *trnG*-UCC, *atpF*, *trnT*-UGU, *trnN*-GUU, *trnD*-GUC, *trnG*-GCC, *psbJ*, and *clpP*) could be good candidates for intraspecific identification. At the genus level, the seven divergent sequences (*rps19*, *rps12*, *trnK*-UUU, *trnG*-GCC, *trnN*-GUU, *trnG*-GCC, and *trnL*-UAG), some of which were also to be highly divergent in other species [42], which could be good candidates for *Firmiana* intrageneric identification.

Localized areas of high nucleotide diversity (π) can be useful molecular markers in population genetics [51]. In our study, the SC regions, especially the LSC, showed higher nucleotide diversity than the IR region., which was consistent with previous studies [44, 70]. For four *F. danxiaensis* cp genomes, nucleotide diversity was low ($\pi < 0.004$) even in the five most divergent regions (*atpE*, *rps3*, *atpF, ndhE*, and *rps4*), indicating the four cp genomes were evolutionary conserved, and probably to be expected for four conspecific populations. Nucleotide diversity was higher in the five most divergent regions in *Firmiana* cp genomes, and two these (*trnL*-UAA and *atpE*) coincide with divergent regions identified

across Malvaceae [51, 53]. Together, all the ten divergent regions could be applicable for further analysis of species identification and population genetics.

Adaptive evolution

Ka/Ks values have been widely used to evaluate natural selection pressure and evolutionary rates of nucleotides in genes [54]. For four *F. danxiaensis* cp genomes, only four genes (*matK, rpoC2*, *rps4*, and *ycf1*) had Ka/Ks values that could be calculated, and all of these were less than one (0.11–0.27). It might be related to the fact that most selection is against disadvantageous non-synonymous mutations [56]. Thus, purifying selection played a vital role in shaping the sequences of *F. danxiaensis* cp genomes.

For *Firmiana* cp genomes, positive selection ($Ka/Ks = 3.31$, $p < 0.05$) was identified in *clpP* when *F. danxiaensis* cp genomes were compared to *Sterculia monosperma* Vent.. Similar results were reported in other *Firmiana* or Malvaceae species (*Althaea officinalis* L., *F. major*, and *Heritiera parvifolia* Merr.) [51], and other plant families [45] as well. The *clpP* encodes the CLP protease, which plays a role in cp protein transformation [57]. It may also influence plant adaptation to high levels of carbon dioxide [58]. The difference in carbon dioxide concentrations bewteen the living habitat of *F. danxiaensis* and *S. monosperma* could be attributable to the positive adaption of the *clpP* gene.

Additionally, six genes under positive selection ($Ka/Ks > 1$, $p > 0.05$) (*accD*, *ccsA*, *ndhH*, *rpl20*, *rpoC2*, and *rps4*) were found when four *F. danxiaensis* cp genomes were compared to other related species. The same Ka/Ks values of *accD*, *ccsA*, and *rpl20* genes indicated that *F. danxiaensis* cp genomes underwent the same degree of selection pressure in different habitats. *accD* may be related to the stress tolerance and resistance to special habitat [59], *ccsA* may be related to substitution of amino-acid, indel presence, or prematurity in stop codon [60], and *rpl20*, which has been evidenced to be essential for the development of cp ribosomes in plants [61]. All the three genes also have been found to play roles in adaptation in other species [36, 45].

The Ka/Ks values of *ndhH*, *rpoC2*, and *rps4* genes were different in different pairwise comparisons. NADH-dehydrogenase genes groups (*ndh*) play vital roles in photosynthesis and they are essential in the use of light energy and the electron transfer chain to produce ATP [62, 64]. Moreover, *ndh* genes could allow adaptation to severe environmental stress conditions (drought, strong light or photo-oxidative stress) in the way of optimizing photosynthesis [65]; *rpo* genes encode proteins that play roles in modification of transcription and post-transcription [65]; *rps4* plays a role in encoding the ribosomal protein S4 [61]. Together, these positively selected and fast-evolving genes may play roles in adaptation to hash environments in *F. danxiaensis*.

Phylogenetic relationship

The phylogenetic trees (ML, BI) strongly supported intrageneric relationships of *Firmiana* consistent with those reported in previous studies [5, 8, 66], namely the two major clades within *Firmiana* recovered. Moreover, we first reported *F. danxiaensis* cp genomes from the NX, SX, and YD populations, revealing the interspecific relationships of *F. danxiaensis* from all four locations and helped clarify the systematic

relationship of *F. hainanensis* within *Firmiana*. Unexpectedly, we found the YD population clustered with the DX population, which is not consistent with the habitat types. The DX, NX and SX populations are located in Danxia landforms and the YD population is on Karst landforms. Moreover, the geographic distance is large between the YD and DX populations compared to the distance between DX and the other two populations. The reason for the YD and DX cluster is currently unknown. Wang et al. [15] indicated that both habitats and climatic factors, independent of geography, can drive genomic differentiation among populations, and the effect of climatic factors can be larger than habitat type. This unexpected relationship between the YD and DX populations requires further study to clarify this.

Materials and methods

Sampling, DNA extraction and genome sequencing

Fresh leaves of *F. danxiaensis* were collected from four populations in Guangdong Province, China: Shixing County (SX; 114°04' E, 25°01' N), Nanxiong City (NX; 114°12' E, 25°07' N), Danxia Mountain (DX; 113°40' E, 24°59' N), and Yingde City (YD; 113°21' E, 24°20' N). Leaves of *F. hainanensis* were collected from an individual cultivated in the South China Botanical Garden (SCBG), Chinese Academy of Sciences (CAS). Total genomic DNA was extracted from silica-gel dried leaves using modified CTAB methods [67]. Approximately 4 Gb of sequence data were generated for each species using the Illumina HiSeq 2000 platform with paired-end (PE) reads 150 bp in length.

Chloroplast genome assembly and annotation

FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to evaluate sequence read quality. Trimmomatic v0.39 [68] was used to remove low-quality and adapter-containing reads. The clean reads were then assembled using GetOrganelle v1.7.5 [69] with k-mers of 21, 45, 65, 85, 105, 115, 127 and max-rounds of 20. The assemblies were blasted to *F. danxiaensis* (MN720649) and *F. major* (NC_037242) genomes using Geneious R9.0.2 [70] to obtained circled genomes. CPGAVAS2 [71] was used to annotate the assembled genomes. The Chloroplot online tool [72] was used to visualize the genomes and their genes. Newly assembled genomes have been submitted to GenBank under accession numbers ON872508, ON872509, ON872510, ON872511, and ON872512 (Table S7).

Repeat sequence analysis

Simple sequence repeats (SSRs) in the genomes were identified by MISA online software [73] with the following parameters: at least eight repeat units for mono-nucleotides, five repeat units for di-nucleotides, four repeat units for tri-nucleotides, and three repeat units for tetra-, penta-, and hexa-nucleotides. Repeat sequences (forward, reverse, complement, and palindromic repeats) were analyzed by an online REPuter software [74] with minimal repeat size of 30 bp and Hamming distance of three.

Codon usage analysis SNP and indel detection

Relative Synonymous Codon Usage (RSCU) was analyzed by the cusp online tool (https://www.bioinformatics.nl/emboss-explorer/) with unique protein-coding genes. As reported, when RSCU ≤ 1.0, it indicates no preference for that codon for the amino acid it codes for, 1.0 < RSCU < 1.2 indicates low preference, 1.2 ≤ RSCU ≤ 1.3 indicates moderate preference, and RSCU > 1.3 indicates high preference [75]. SNPs and indels were pairwise detected among four *F. danxiaensis* cp genomes using snippy [76].

Chloroplast genome comparison

The mVISTA tool [77] was used to compare the structures of *F. danxiaensis* cp genomes using the Shuffle-LAGAN mode. Four *F. danxiaensis* cp genomes were compared to *F. colorata*, *F. hainanensis*, *F. kwangsiensis*, *F. major*, *F. pulcherrima*, *F. simplex*, and *Sterculia monosperma*. The IRscope online program [78] was used to compare the LSC/IRb/SSC/IRa region borders with *F. danxiaensis* cp genome from SX as a reference, plus for species mentioned above in *Firmiana*. DnaSP6 software [79] was used to discover mutation hotspots with window length and step size fixed at 100 bp and 25 bp, respectively.

Selective pressure analysis

We extracted the CDS sequences of the protein-coding genes from eight sequences (*F. danxiaensis* from four locations, *F. colorata*, *F. kwangsiensis*, *F. major*, and *S. monosperma*), which resulted in 78 CDS matrices. MAFFT [80] was used to generate CDS alignments. ALTER [81] online tool was used to convert .fasta format to .aln format. The KaKs_Calculator 2.0 [82] was used to estimate the nonsynonymous (Ka)/ synonymous (Ks) values for each gene using the γYN method. Ka/Ks < 1, Ka/Ks = 1, and Ka/Ks > 1, indicate purifying, neutral, and positive selection, respectively.

Phylogenetic analysis

We constructed a phylogenetic tree with the newly assembled cp genomes of *F. danxiaensis* and *F. hainanensis* combined eight cp genomes downloaded from NCBI (Table S7). PhyloSuite [83] software was used to create Maximum Likelihood (ML) tree and Bayesian (BI) trees. ML analysis was performed using IQ-TREE [84] with an edge-linked partition model (K81u + I + G4 + F) and 5000 standard non-parametric bootstrap replicates. BI analysis was performed using MrBayes3.2.6 [85] with default model, two independent parallel chains and 5,000,000 generations and sampling once every 1,000 generations. The first 25% of trees from all runs were discarded as burn-in. *Tilia amurensis* Rupr. was used as outgroup. Finally, ITOL [86] software was used to visualize and refine the tree.

Abbreviations

cp: Chloroplast; LSC: Large single-copy region; SSC: Small single-copy region; IR: Inverted Repeat; CDS: Coding sequence; RSCU: Relative synonymous codon usage; SSR: Simple sequence repeats; π: Nucleotide diversity; ML: Maximum likelihood; BI: Bayesian inference; NCBI: National Center for Biotechnology Information; SX: Shixing location; NX: Nanxiong location; DX: Danxia Mountain location; YD: Yingde location.

Declarations

Acknowledgments

We are very grateful to Fang Chen, Zai-xiong Chen of Danxia Mountain National Nature Reserve, Ping-sheng Zhong of Nanxiong City Forestry Bureau, Shang-fu Chen of Shixing Forestry Bureau, and Yuan-qiu Li of Guangdong Shimentai National Nature Reserve for assistance with sample collection.

Authors' contributions

Hong-feng Chen & Ya-li Li: Conceptualization, Methodology; Hong-feng Chen, Ya-li Li & Shuang-wen Deng: Field investigation, Materials collection; Ya-li Li, Li-yun Nie & Shuang-wen Deng: Data analyses and visualization; Ya-li Li: manuscript preparation; Lei Duan, Zheng-feng Wang, Joseph L. M. Charboneau, Boon-Chuan Ho & Hong-feng Chen: manuscript revision. All authors have read and approved the manuscript.

Funding

This work was supported by the Project for Research and Development of Key Field in Guangdong (Grant No. 2020B1111530004), and the Project for Wild Plant Conservation and Management of State Forestry and Grassland Administration (Grant No. 20190730).

Availability of data and materials

The sequence data of *Firmiana danxiaensis* and *Firmianan hainanensisi* chloroplast genomes involved in this study have been deposited in GenBank with accession numbers ON872508 (SX), ON872509 (NX), ON872510 (DX), ON872511 (YD), and ON872512. All relevant data can be found within the manuscript and its supporting materials.

Ethics approval and consent to participate

The authors confirm that all methods comply with relevant institutional, national, and international guidelines and legislations. The collection of plant material was carried out with the permission.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author details

[1]Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China. [2]University of Chinese Academy of Sciences, Beijing 100049, China. [3]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 510650, China. [4]Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, Key Laboratory of Carbon Sequestration in Terrestrial Ecosystem, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China. [5]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA. [6]Singapore Botanic Gardens, National Parks Board, 1 Cluny Road, Singapore 259569, Republic of Singapore.

References

1. Tang Y, Michael GG, Laurence JD. Sterculiaceae. Wu CY, Raven PH, Hong DY, editors. Beijing: Science Press. 2007.
2. Mabberley DJ. The plant-book: a portable dictionary of the vascular plants. 2nd ed. New York, USA: Cambridge University Press. 1997.
3. Kostermans AJGH. Notes on *Firmiana* Marsili (Sterculiaceae). Blumea. 1989; 34:117-8.
4. Kostermans AJGH. The genus *Firmiana* Marsili (Sterculiaceae). Reinwardtia. 1954; 4:281-310.
5. Lu QF, Huang ZH, Luo WH. Characterization of complete chloroplast genome in *Firmiana kwangsiensis* and *F. danxiaensis* with extremely small populations. Biodiv Sci. 2021;29:586-95.
6. Huang YS, Wu WH, Xu WB, Liu Y. *Firmiana calcarea* sp. nov. (Malvaceae) from limestone areas in Guangxi, China. Nord. J. Bot. 2011;29:608-10.
7. Zhang GL, Cai L, Duan JQ, Wang T, Xiang JY. *Firmiana daweishanensis* sp. nov. (Malvaceae)from Southeast Yunan China. Phytotaxa. 2020;456:215-8.
8. Li RZ, Cai J, Yang JB, Zhang ZR, Li DZ, Yu WB. Plastid phylogenomics resolving phylogenetic placement andgenera phylogeny of Sterculioideae (Malaceae). Guihaia. 2022;42:25-38.
9. Qin HN, Yang Y, Dong SY, He Q, Jia Y, Zhao LN, et al. Threatened species list of China's higher plants. Biodiv Sci. 2017;25:696-744.
10. Xu XH, Qiu HX, Xu SJ. New species and variety of Sterculiaceae from China. J South China Agric Univ. 1987;8:1-5.
11. Liao CH, Zhang QX. *Firmiana danxiaensis* grows in Nanxiong Quan'an Cangshi Village. Shaoguan Daily. 2014;A01.
12. Miao SY, Huang HZ, Li YQ, Tao WQ, Zeng YJ, Chen ZX, et al. Resource Survey and Protection of the Key National Protected Species *Firmiana danxiaensis* Endemic to Guangdong, China. Subtrop Plant Sci. 2020;49:71-5.
13. Li YL, Fu L, Lei YY, Li YQ, Chen HF. New Distribution and Modification of Morphological Characteristics of *Firmiana danxiaensis*, an Endemic Species of China. J Trop Subtrop Bot. 2022;30:735-41.

14. Li YB. Landscape ecological characteristics and ecological construction of karst mountain areas in southwest China. Geography. 2004;13:702-6.
15. Wang J, Feng C, Jiao TL, Von Wettberg EB, Kang M. Genomic signature of adaptive divergence despite strong nonadaptive forces on edaphic islands: a case study of *Primulina juliae*. Genome Biol Evol. 2017;9:3495-508.
16. Hao Z, Kuang Y, Kang M, Niu S. Untangling the influence of phylogeny, soil and climate on leaf element concentrations in a biodiversity hotspot. Funct Ecol. 2015;29:165-76.
17. Jiao TL. Studies on the adaptive evolution of *Primulina juliae* endemic in special habitat and analysis of chloroplast genomics among its closely related species. South China Botanical Garden, Chinese Academy of Sciences. 2018.
18. Liu F, Wang SJ, Luo HB, Liu YS, Liu HY. Micro-habitats in Karst forest ecosystem and variability of soils. Acta Pedol Sin. 2008;45:1055-62.
19. Wu ZH, Liao MR, Yang TG, Dong X, Lan DQ, Qin R, et al. Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae). BMC Genom. 2020;21:621.
20. Daniell H, Lin C, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17:134.
21. Dobrogojski J, Adamiec M, Luci ´nski R. The chloroplast genome: A review. Acta Physiol Plant. 2020;42:1-13.
22. Liu YC, Lin BY, Lin JY, Wu WL, Chang CC. Evaluation of chloroplast DNA markers for intraspecific identification of *Phalaenopsis equestris* cultivars. Sci Hortic. 2016;203:86-94.
23. Lee HL, Jansen RK, Chumley TW, Kim KJ. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping Inversions. Mol Biol Evol. 2007;24:1161-80.
24. Ouyang J, Zhuang CW, Luo XY, Qiao Y, Duan YN, Zheng SF, et al. Data collection and visualization of *Firmiana danxiaensis* H. H. Hsue & H. S. Kiu in the Zhanglao Peak, Danxia Mountain, Guangdong Province, China. Sci Geog Sin. 2020;40:1181-90.
25. Wu JR, Wei BJ, Hu XJ, Zhu ML, Li RZ, Feng XC. Correlation of spatial distribution and habitat factors of *Firmiana danxiaensis* based on Geo-Detector. J Appl Ecol. 2020;31:2671-9.
26. Ouyang J, Peng H, Luo XY, Chen ZX, Zhang AX, Ma YX. Environmental features of the micro-landforms of the spatial distribution of the national rare species of *Firmiana danxiaensis* on the Danxiashan Mountain. Sci Geog Sin. 2017;37:1585-92.
27. Peng SL, Liao WB, Li Z, Jia FL, Wang YY, Chang H, et al. Integrated biological surveys on Mount Danxia, Guangdong. Beijing: Sciences Press. 2011.
28. Zhang W, Huang YT, Zhong PS, He XY, Chen HF. Study on the characteristics of seed germination of the endangered plant *Firmiana danxiaensis*. Forestry and Environ Sci. 2018;31:51-5.
29. Zhang QM, Luo XY, Chen ZX. Conservation and reintroduction of *Firmiana danxiaensis*, a rare tree species endemic to southern China. ORYX. 2014;48:485.

30. Gu LH, Zhou HJ. The first artificial insemination success of *Firmiana danxiaensis*. Southern Daily. 2010;A10.
31. Lin XY, Chen ZH, Yang YY, Mo SQ, Li DL. The complete chloroplast genome sequence of *Firmina danxiaensis*. Mitochondrial DNA B Resour. 2020;5:908-9.
32. Lu QF, Luo WH, Huang ZH. The complete chloroplast genome of *Firmiana danxiaensis*, an endangered species endemic to Danxia landform in Southern China, Mitochondrial DNA B Resour. 2019;4:4071-2.
33. Abdullah, Shahzadi, I., Mehmood, F., Ali, Z., Malik, M. S., Waseem, S., Mirza, B., Ahmed, I., Waheed, M. T. Comparative analyses of chloroplast genomes among three *Firmiana* species: Identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. Plant Gene. 2019;19:100199.
34. Wang JH, Cai YC, Zhao KK, Zhu ZX, Zhou RC, Wang HF. Characterization of the complete chloroplast genome sequence of *Firmiana pulcherrima* (Malvaceae). Conservation Genet Resour. 2018;10:445-8.
35. Ya JD, Yu ZX, Yang YQ, Zhang SD, Zhang ZR., Cai J, et al. Complete chloroplast genome of *Firmiana major* (Malvaceae), a critically endangered species endemic to southwest China. Conservation Genet Resour. 2018;10:713-5.
36. Wang JH, Moore MJ, Wang H, Zhu ZX, Wang HF. Plastome evolution and phylogenetic relationships among Malvaceae subfamilies. Gene. 2021;765:145103.
37. Foerstner KU, von Mering C, Hooper SD, Bork P. Environments shape the nucleotide composition of genomes. EMBO Rep. 2005;6:1208-13.
38. Terakami S, Matsumura Y, Kurita K, Kanamori H, Katayose Y, Yamamoto T, et al. Complete sequence of the chloroplast genome from pear (*Pyrus pyrifolia*): genome structure and comparative analysis. Tree Genet Genomes. 2012;8:841-54.
39. Liu XF, Zhu GF, Li DM, Wang XJ. Complete chloroplast genome sequence and phylogenetic analysis of *Spathiphyllum*'Parrish'. PLoS One. 2019;14:e0224038.
40. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, et al. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell. 2001;13:645-58.
41. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genom. 2007; 8:174.
42. Mo ZH, Lou WR, Chen YQ, Jia XD, Zhai M, Guo ZR, et al. The Chloroplast Genome of *Carya illinoinensis*: Genome Structure, Adaptive Evolution, and Phylogenetic Analysis. Forests, 2020;11:207-21.
43. Yan X, Luo X, Cai X. Analysis of codon usage pattern in *Taenia saginata* based on a transcriptome dataset. Parasit. Vectors. 2014;7:1-11.
44. Liang D, Wang H, Zhang J, Zhao Y, Wu F. Complete chloroplast genome sequence of *Faguslongipetiolata* Seemen (Fagaceae): genome structure, adaptive evolution, and phylogenetic

relationships. Life. 2022;12:92.

45. Sheikh-Assadi M, Naderi R, Kafi M, Fatahi R, Salami SA, Shariati V. Complete chloroplast genome of *Lilium ledebourii* (Baker) Boiss and its comparative analysis: lights into selective pressure and adaptive evolution. Sci Rep. 2022;12:9375.
46. Bierne N, Eyre-Walker A. The problem of counting sites in the estimation of the synonymous and nonsynonymous substitution rates: Implications for the correlation between the synonymous substitution rate and codon usage bias. Genetics. 2003;165:1587-97.
47. Zhou M, Long W, Li X. Patterns of synonymous codon usage bias in chloroplast genomes of seed plants. For. Stud. China. 2008;10:235-42.
48. Duan H, Guo J, Xuan L, Wang Z, Li M, Yin Y, et al. Comparative chloroplast genomics of the genus *Taxodium*. BMC Genom. 2020;21:114.
49. Jiang H, Tian J, Yang J, Dong X, Zhong Z, Mwachala G, et al. Comparative and phylogenetic analyses of six *Kenya Polystachya* (Orchidaceae) species based on the complete chloroplast genome sequences. BMC Plant Biol. 2022;22:177.
50. Moghaddam M, Ohta A, Shimizu M, Terauchi R, Kazempour-Osaloo S. The complete chloroplast genome of *Onobrychis gaubae* (Fabaceae-Papilionoideae): comparative analysis with related IR-lacking clade species. BMC Plant Biol. 2022;22:75.
51. Abdullaha, Mehmooda F, Shahzadia I, Waseemb S, Mirzaa B, Ahmed I, et al. Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): Comparative analyses and identification of mutational hotspots. Genomics. 2020;112:581-91.
52. Ding SX, Dong X, Yang JX, Guo CC, Cao BB, Guo Y, et al. Complete chloroplast genome of *Clethra fargesii* Franch., an original sympetalous plant from central China: comparative analysis, adaptive evolution, and phylogenetic relationships. Forests. 2021;12:441.
53. 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.
54. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol. 2000;17:32-43.
55. Kimura M. The neutral theory of molecular evolution and the world view of the neutralists. Genome. 1989;31:24-31.
56. Zhang J. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J Mol Evol. 2000;50:56-68.
57. Kuroda, H., Maliga, P. The plastid *clpP1* protease gene is essential for plant development. Nature. 2003;425:86-9.
58. Majeran W, Wollman FA, Vallon O. Evidence for a role of *clpP* in the degradation of the chloroplast cytochrome b6 f complex. Plant Cell. 2000;12:137-50
59. Sablok G, Mudunuri SB, Edwards D, Ralph PJ. Chloroplast genomics: Expanding resources for an evolutionary conserved miniature molecule with enigmatic applications. Curr Plant Biol. 2016;7:34-8.

60. Raman G, Park S. The complete chloroplast genome sequence of the *Speirantha gardenii*: Comparative and adaptive evolutionary analysis. Agronomy. 2020;10:1405.
61. Rogalski M, Ruf S, Bock R. Tobacco plastid ribosomal protein S18 is essential for cell survival. Nucleic Acids Res. 2006;34:4537-45.
62. Peltier G, Aro EM, Shikanai T. NDH-1 and NDH-2 plastoquinone reductases in oxygenic photosynthesis. Annu Rev Plant Biol. 2016;67:55-80.
63. Yamori W, Shikanai T. Physiological functions of cyclic electron transport around photosystem I in sustaining photosynthesis and plant growth. Annu Rev Plant Biol. 2016;67:81-106.
64. Rumeau D, Peltier G, Cournac L. Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. Plant, Cell & Environ. 2007;30:1041-51.
65. Jalal A, Schwarz C, Schmitz-Linneweber C, Vallon O, Nickelsen J, Bohne AV. A small multifunctional pentatricopeptide repeat protein in the chloroplast of *Chlamydomonas reinhardtii*. Mol Plant. 2015;8:412-26.
66. Fan Q, Guo W, Chen SF, Liao WB. Phylogeny of *Firmiana* (Sterculiaceae) based on nrDNA ITS analysis. In: The 10th Youth Academic Seminar of National System and Evolutionary Botany (ed. Botanic Society of Yunnan), p.68. Kunming. 2011.
67. Doyle JJ. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19:11-5.
68. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114-20.
69. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, et al. Organelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:241.
70. Van Tan L, Thi Thu Hong N, My Ngoc N, Tan Thanh T, Thanh Lam V, Anh Nguyet L, et al. SARS-CoV-2 and co-infections detection in nasopharyngeal throat swabs of COVID-19 patients by metagenomics. J Infect. 2020;81:175-7.
71. Shi LC, Chen HM, Jiang M, Wang LQ, Wu X, Huang LF, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 2019;47:65-73.
72. Zheng SY, Poczai P, Hyvönen J, Tang J, Amiryousefi A. Chloroplot: an online program for the versatile plotting of organelle genomes. Front Genet. 2020;11:576124.
73. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33:2583-85.
74. 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;29:4633-42.
75. Zuo LH, Shang AQ, Zhang S, Yu XY, Ren YC, Yang MS, et al. The first complete chloroplast genome sequences of *Ulmus* species by de novo sequencing: Genome comparative and taxonomic position analysis. PLoS One. 2017;12:e0171264.

76. Seemann T. Snippy: Fast bacterial variant calling from NGS reads. Available online: https://github.com/tseemann/snippy (accessed on 11 June 2022). 2015.
77. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: Computational tools for comparative genomics. Nucleic Acids Res, 2004;32:W273-9.
78. Amiryousefi A, Hyvönen J, Poczai P. IRscope: An online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34:3030-1.
79. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34:3299-302.
80. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772-80.
81. Glez-Peña D, Gómez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada D. ALTER: program-oriented format conversion of DNA and protein alignments. Nucleic Acids Res. 2010;38: (Web Server issue): W14-8.
82. Wang DP, Zhang YB, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genom Proteom Bioinf. 2010;8:77-80.
83. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20:348-55.
84. Nguyen LT, Schmidt HA, Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268-74.
85. Ronquist FM, Teslenko P, van der Mark DL, Ayres A, Darling S, Höhna B, et al. MRBAYES 3.2: Efficient Bayesian phylogenetic inference and model selection across a large model space. Syst Biol. 2012;61:539-42.
86. Ivica L, Peer B. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res. 2019;47:W256-9.

Tables

Table 1 Chloroplast genomes of *Firmiana danxiaensis* from four populations

Item	SX	NX	DX	YD
Cp genome/ bp	160832	160836	161016	161206
LSC/ bp	89737	89739	89905	90135
SSC /bp	20049	20051	20067	20027
IR/ bp	25523		25522	
LSC GC content %	34.8	34.8	34.8	34.7
SSC GC content %	31.4	31.3	31.1	31.2
IR GC content %	42.8	42.9	43.0	43.0
GC content (%)	36.9			
Cp genes	129			
Unique CDS	78			
CDS	84			
tRNA	37			
rRNA	8			

SX: Shixing population. **NX**: Nanxiong population. **DX**: Danxia Mountain population. **YD**: Yingde population

Table 2 Gene content and functional classification of *Firmiana danxiaensis* chloroplast genome

Category	Gene Group	Gene Name
Photosynthesis	Subunits of photosystem I	*psaA*, *B*, *C*, *I*, *J*
	Subunits of photosystem II	*psbA*, *B*, *C*, *D*, *E*, *F*, *H*, *I*,*J*, *K*, *L*, *M*, *N*, *T*, *Z*
	Subunits of NADH dehydrogenase	*ndhA**, *B**(2), *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*, *K*
	Subunits of cytochrome b/f complex	*petA*, *B**, *D**, *G*, *L*, *N*
	Subunits of ATP synthase	*atpA*, *B*, *E*, *F**, *H*, *I*
	Large subunit of rubisco	*rbcL*
Self-replication	Proteins of large ribosomal subunit	*rpl22*, *14*, *16**, *2**(2), *20*, *23*(2), *32*, *33*, *36*
	Proteins of small ribosomal subunit	*rps16**, *11*, *12**(2), *14*, *15*, *18*, *19*, *2*, *3*, *4*, *7*(2), *8*
	Subunits of RNA polymerase	*rpoA*, *B*, *C1**, *C2*
	Ribosomal RNAs	*rrn16*(2), *23*(2), *4.5*(2), *5*(2)
	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, *trnfM*-CAU
Other genes	Maturase	*matK*
	Protease	*clpP***
	Envelope membrane protein	*cemA*
	Acetyl-CoA carboxylase	*accD*
	c-type cytochrome synthesis gene	*ccsA*
Genes of unknown function	Conserved hypothetical chloroplast ORF	*ycf1*, *ycf2*(2), *ycf3***, *ycf4*

Note: Gene (2), Multiple copy gene, the number of copies in parenthesis; Gene *, Gene with one intron; Gene **, Genes containing two introns.

Table 3 Statistics of relative synonymous codon usage (RSCU) values among chloroplast genomes of *Firmiana danxiaensis* from four populations

Sample	Codon	ENC	CAI	RSCU>1	1<RSCU≤1.2	1.2<RSCU≤1.3	RSCU>1.3
SX	53610	55.85	0.649	31	10	10	11
NX	53612	55.94	0.644	30	7	11	12
DX	53672	55.41	0.641	34	12	10	12
YD	53735	55.13	0.648	32	13	5	14

ENC = Effective number of codons, CAI = Codon adaptation index

Figures

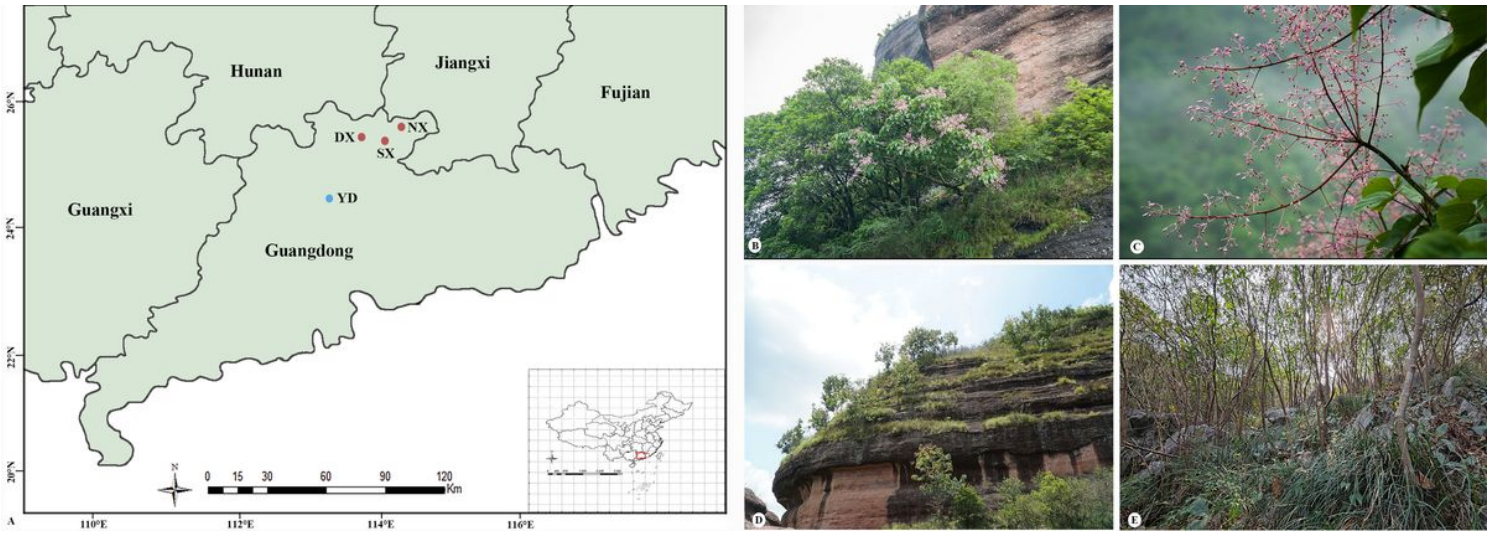

Figure 1

Sampling sites and habitats of *Firmiana danxiaensis*. **SX**:Shixing location. **NX**:Nanxiong location. **DX**:Danxia Mountain location. **YD**:Yingde location. **Red circles**: Danxia landform. **Blue circle**: Karst landform. **B)** *Firmiana danxiaensis* growing on the cliffs of Danxia habitat. **C)** Inflorescence of *Firmiana danxiaensis*. **D)** . Danxia habitat. **E)** Karst habitat.

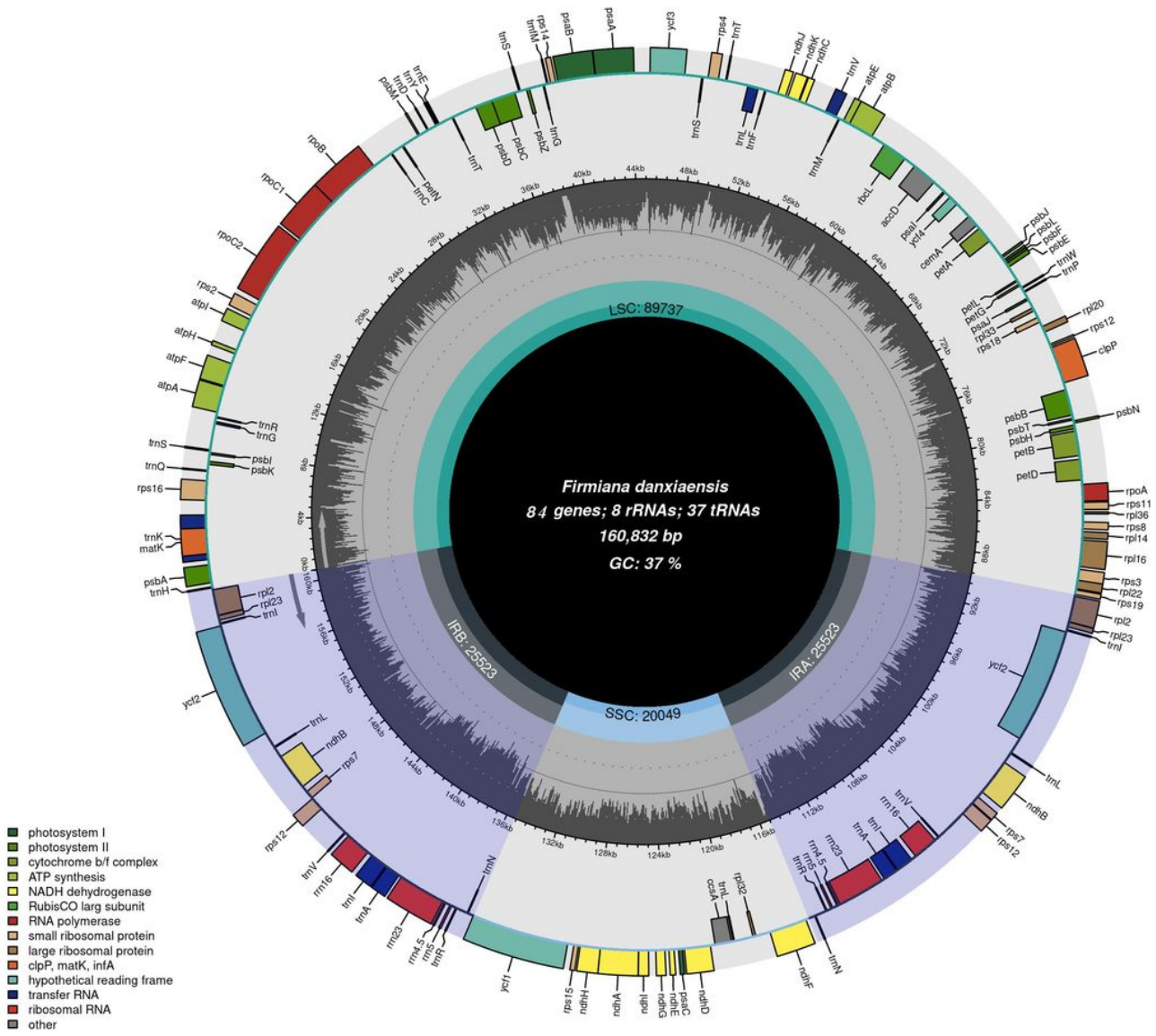

Figure 2

Chloroplast genome of *Firmiana danxiaensis* from the Shixing (SX) population. Genes are color-coded according to legend shown on the bottom left. GC content is plotted in the interior circle.

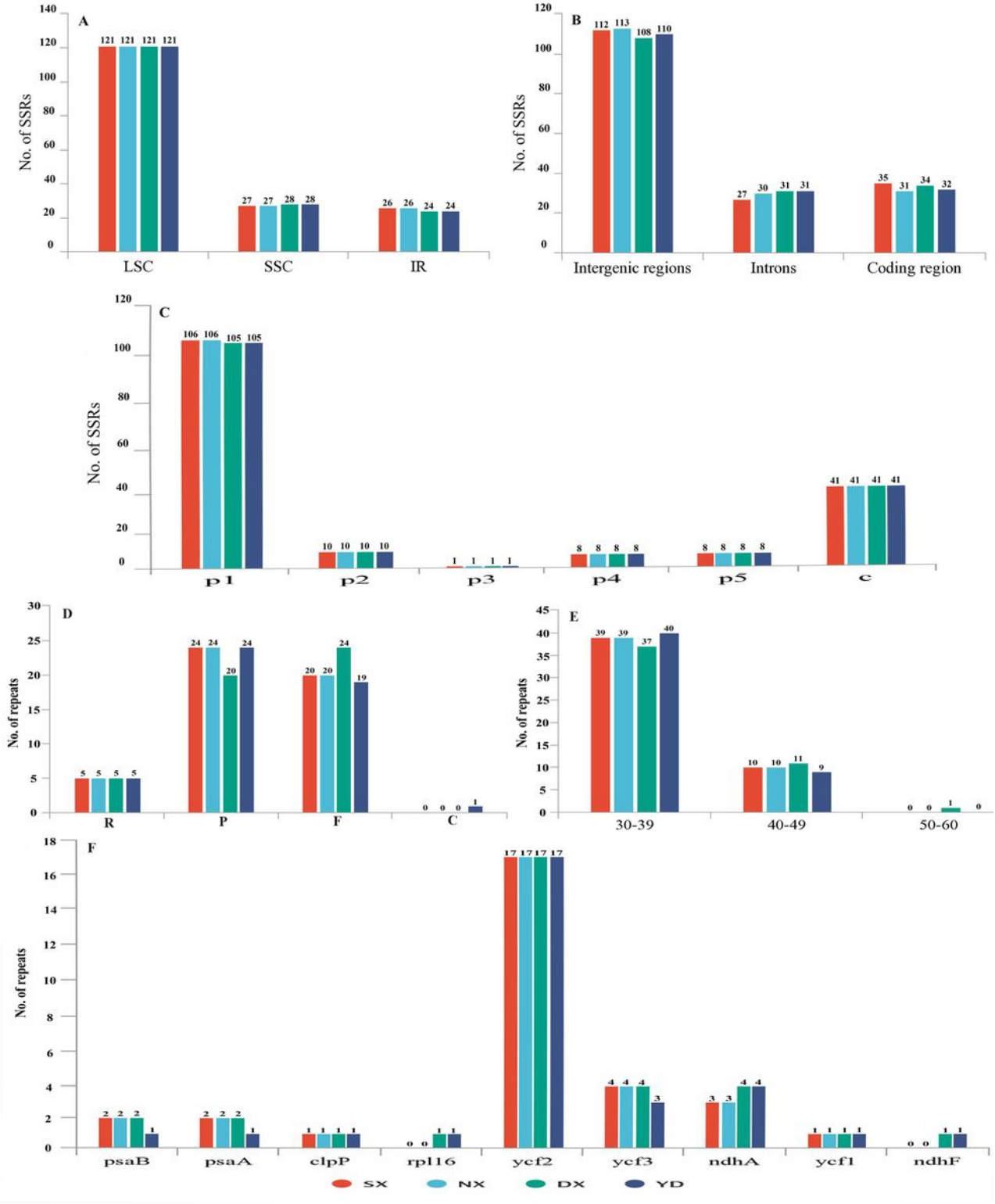

Figure 3

Analysis of repeated sequences of four *Firmiana danxiaensis* chloroplast genomes. **A)** The number of SSRs distributed in different copy regions. **B)**The number of SSRs distributed in different gene regions. **C)** The number of six SSR types. **p1**:mononucleotide repeats. **p2**:dinucleotide repeats. **p3**:trinucleotide repeats. **p4**:tetraucleotide repeats. **p5**:pentaucleotide repeats. **c**: compound repeats. **D)** The number of four long repeat types. **R**: reverse repeats. **P**: palindromic repeats. **F**: forward repeats. **C**: complementary

repeats. **E)** The length distribution of long repeats. **F)** The number of long repeats located in coding genes. **SX**:Shixing population. **NX**:Nanxiong population. **DX**:Danxia Mountain population. **YD**:Yingde population

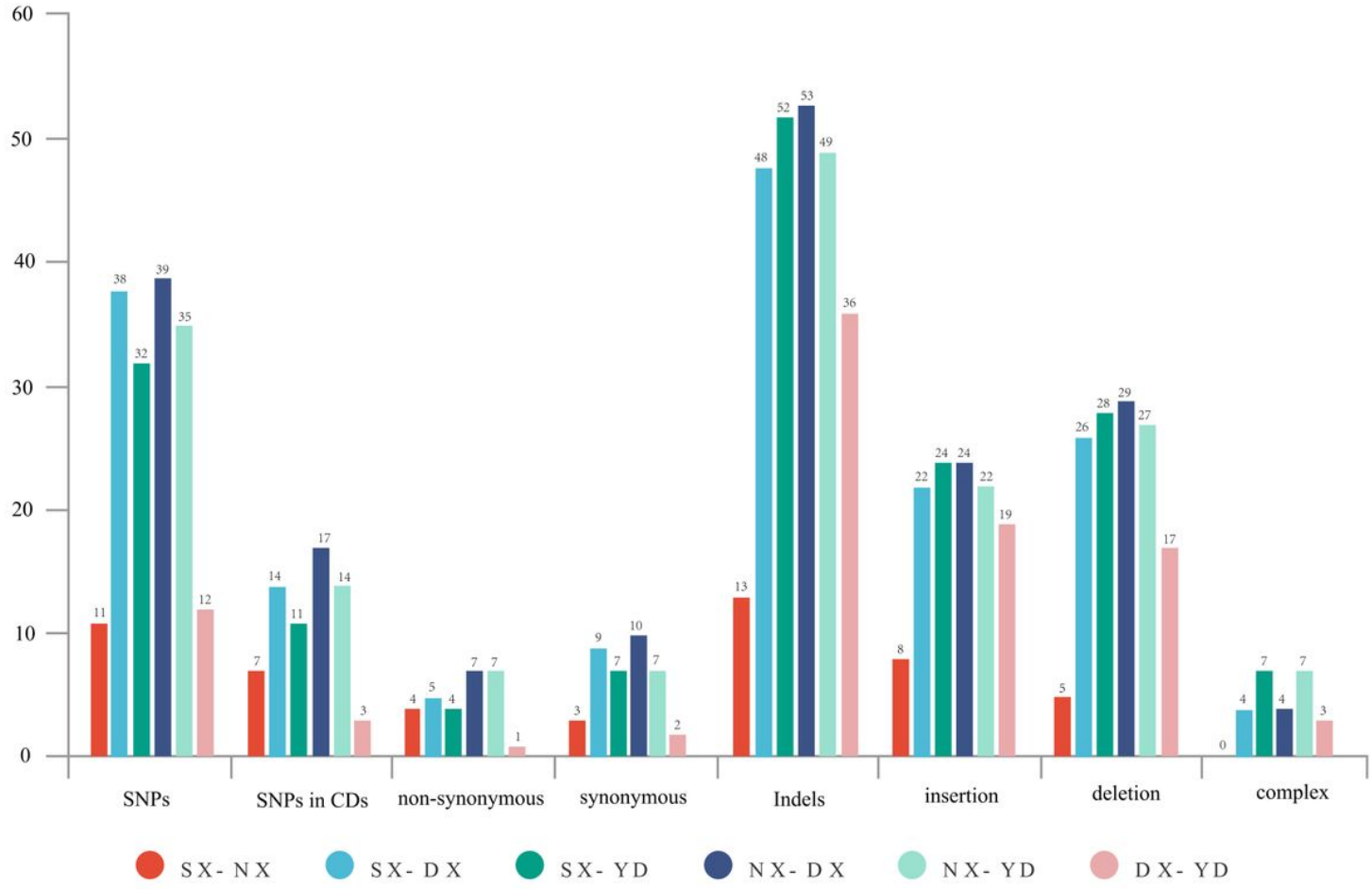

Figure 4

Counts of SNP and indels from pairwise comparisos of four *Firmiana danxiaensis* chloroplast genomes. **SX** Shixing population. **NX**Nanxiong population. **DX** Danxia Mountain population. **YD** Yingde population

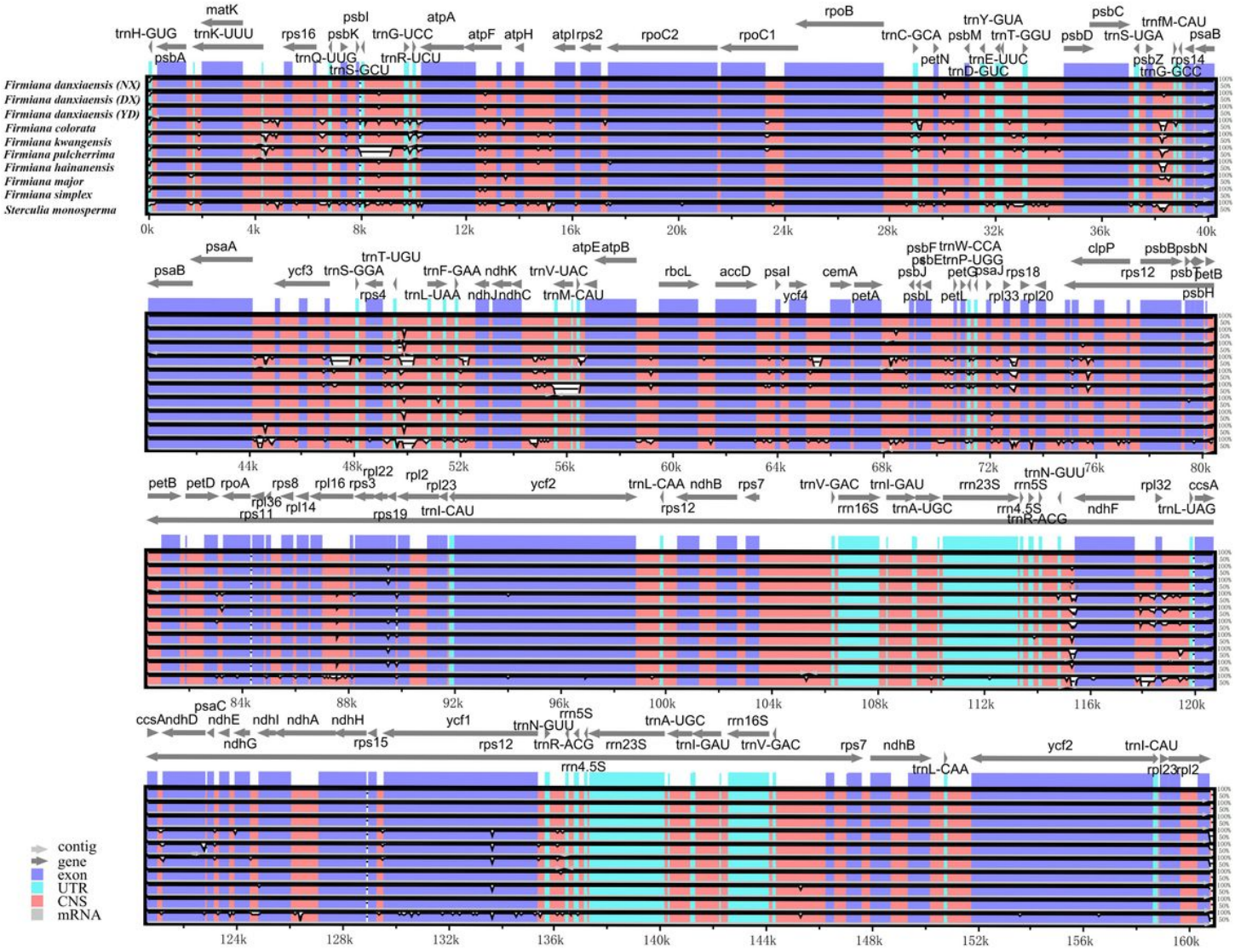

Figure 5

Comparisons of 11 Malvaceae chloroplastgenomes. Gray arrows above the alignment indicate gene orientation. Genome regions are color-coded as exons, rRNA or tRNA (UTR), and non-coding sequences (CNS). Vertical scale indicates the percentage of identity ranging from 50 to 100%. **SX**:Shixing population. **NX**:Nanxiong population. **DX**:Danxia Mountain population. **YD**:Yingde population

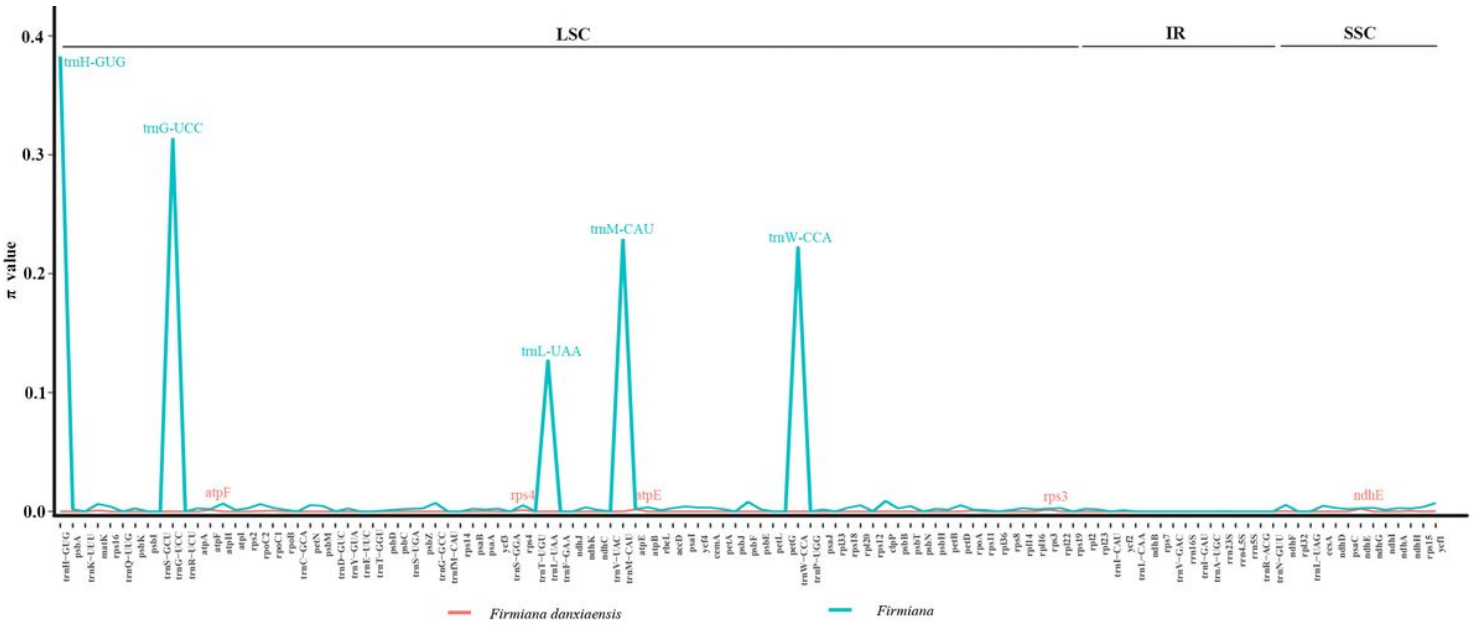

Figure 6

Nucleotide diversity (π) calculated from 111 loci. **Pink line**: π values of four *Firmiana danxiaensis* chloroplast genomes from four populations. **Blue line**: π values of 10 *Firmiana* chloroplastgenomes. **x-axis**: gene name. **y-axis**: π value

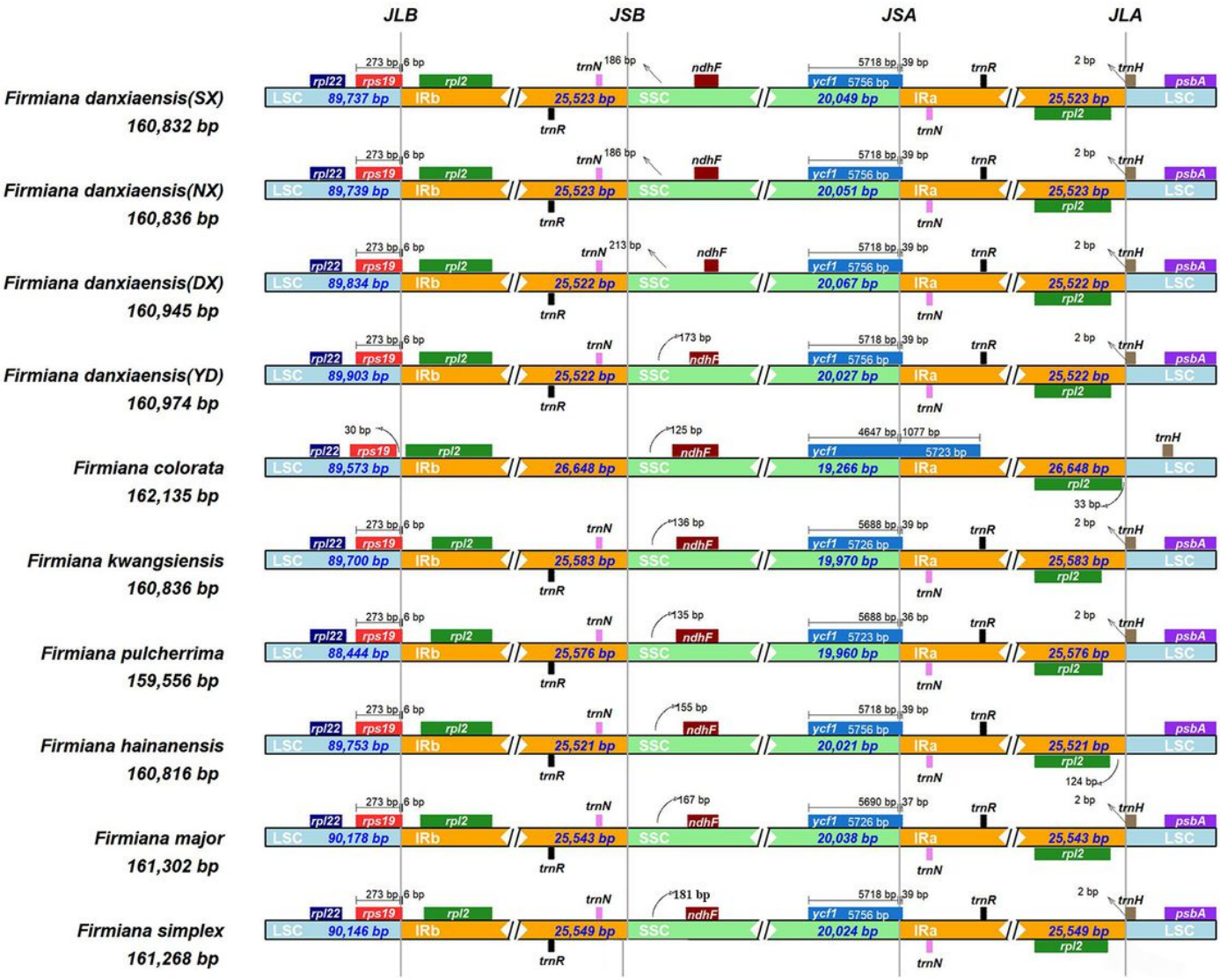

Figure 7

Comparison of the boundaries of the LSC, SSC, and IR regions among ten *Firmiana* chloroplast genomes. **SX**:Shixing population. **NX**:Nanxiong population. **DX**:Danxia Mountain population. **YD**:Yingde population

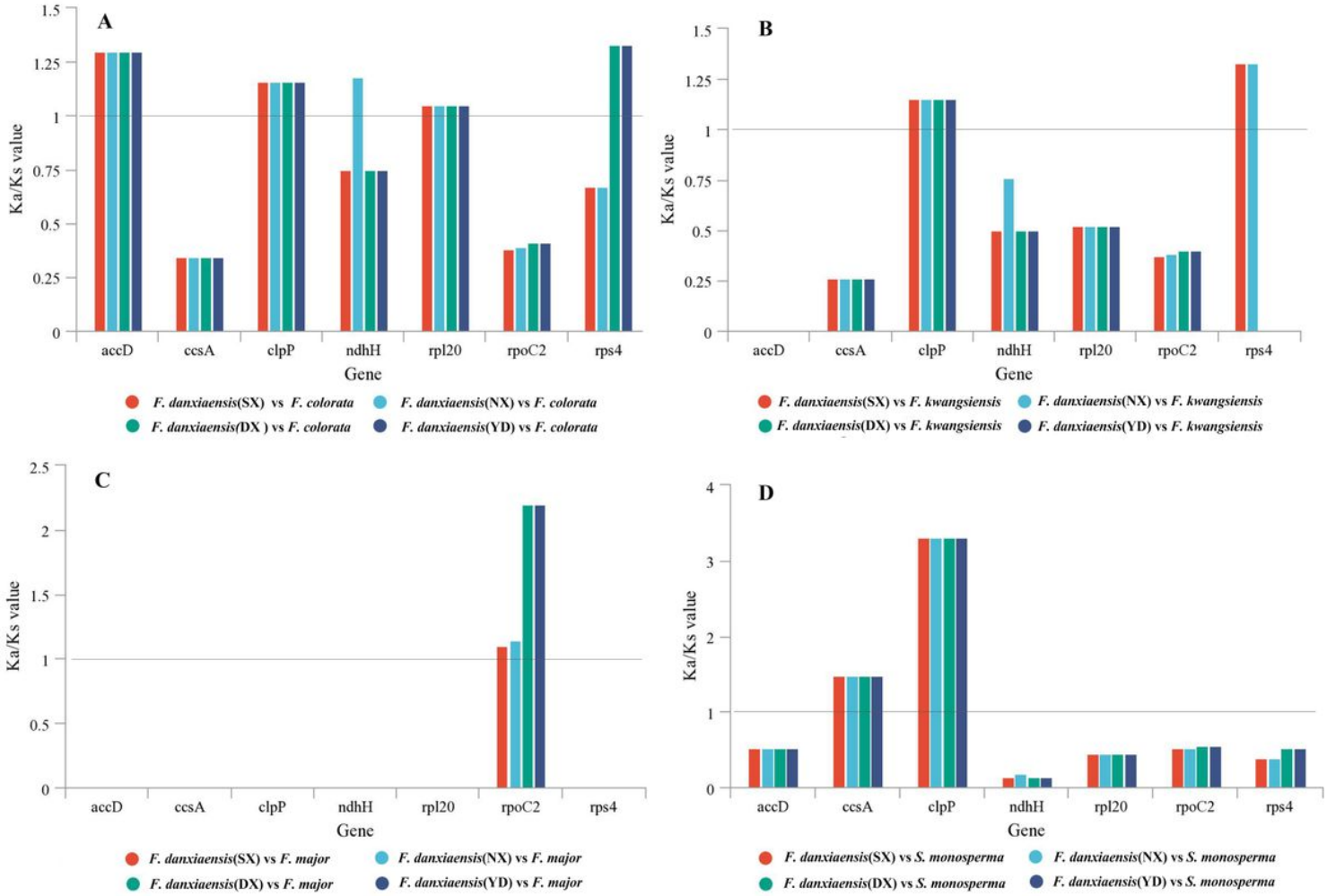

Figure 8

Comparison of Ka/Ks value of seven genes when *Firmiana danxiaensis* from four populations were pairwise compared to other four species. **SX**:Shixing population. **NX**:Nanxiong population. **DX**:Danxia Mountain population. **YD**:Yingde population

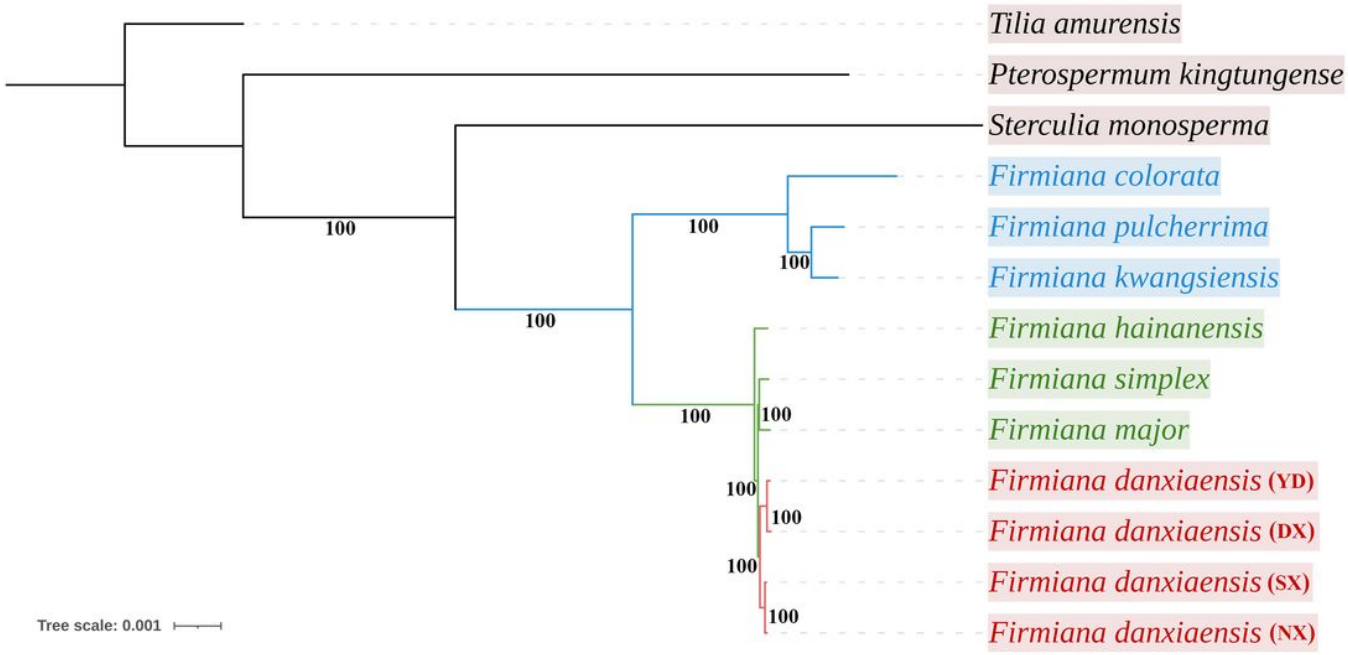

Figure 9

Phylogenetic relationships of *Firmiana*species. *Tilia amurensis* was set as outgroup. Phylogenetic tree was constructed by MrBayes (BI). The bootstrap values are represented at nodes

Supplementary Files

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

- supplementaryfile.zip