

RESEARCH ARTICLE

Five Complete Chloroplast Genome Sequences from *Diospyros*: Genome Organization and Comparative Analysis

Jianmin Fu^{1,2}*, Huimin Liu²*, Jingjing Hu³, Yuqin Liang², Jinjun Liang², Tana Wuyun^{2*}, Xiaofeng Tan^{1*}

1 Key Laboratory of Cultivation and Protection for Non-Wood Forest Trees, Ministry of Education, Central South University of Forestry and Technology, Changsha, Hunan, China, **2** Non-Timber Forestry Research and Development Center, Chinese Academy of Forestry, Zhengzhou, Henan, China, **3** Department of Bioinformatics, Haplox Biotechnology Co., Ltd., Shenzhen, China

* These authors contributed equally to this work.

* tanatanan@163.com (TNW); tanxiaofengcn@126.com (XFT)



CrossMark
click for updates

OPEN ACCESS

Citation: Fu J, Liu H, Hu J, Liang Y, Liang J, Wuyun T, et al. (2016) Five Complete Chloroplast Genome Sequences from *Diospyros*: Genome Organization and Comparative Analysis. PLoS ONE 11(7): e0159566. doi:10.1371/journal.pone.0159566

Editor: Berthold Heinze, Austrian Federal Research Centre for Forests BFW, AUSTRIA

Received: March 2, 2016

Accepted: July 4, 2016

Published: July 21, 2016

Copyright: © 2016 Fu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Complete chloroplast genome sequences have been deposited in GenBank (accession numbers:KT223565, KM522849, KM522848, KM522850, KM504956). The ITS data were from an article of Jinjun Liang. The SSR data were from a journal article of Yuqin Liang. In this article, we used the ITS and SSR data of five *Diospyros* species to make a comparison with our results. We have cited these two references in our paper.

Funding: This research was supported by the National Key Technology R&D Program in the 12th Five-year Plan of China (No. 2013BAD14B0502).

Abstract

Diospyros is the largest genus in Ebenaceae, comprising more than 500 species with remarkable economic value, especially *Diospyros kaki* Thunb., which has traditionally been an important food resource in China, Korea, and Japan. Complete chloroplast (cp) genomes from *D. kaki*, *D. lotus* L., *D. oleifera* Cheng., *D. glaucifolia* Metc., and *Diospyros* ‘Jinzaoshi’ were sequenced using Illumina sequencing technology. This is the first cp genome reported in Ebenaceae. The cp genome sequences of *Diospyros* ranged from 157,300 to 157,784 bp in length, presenting a typical quadripartite structure with two inverted repeats each separated by one large and one small single-copy region. For each cp genome, 134 genes were annotated, including 80 protein-coding, 31 tRNA, and 4 rRNA unique genes. In all, 179 repeats and 283 single sequence repeats were identified. Four hypervariable regions, namely, intergenic region of *trnQ_rps16*, *trnV_ndhC*, and *psbD_trnT*, and intron of *ndhA*, were identified in the *Diospyros* genomes. Phylogenetic analyses based on the whole cp genome, protein-coding, and intergenic and intron sequences indicated that *D. oleifera* is closely related to *D. kaki* and could be used as a model plant for future research on *D. kaki*; to our knowledge, this is proposed for the first time. Further, these analyses together with two large deletions (301 and 140 bp) in the cp genome of *D. ‘Jinzaoshi’*, support its placement as a new species in *Diospyros*. Both maximum parsimony and likelihood analyses for 19 taxa indicated the basal position of Ericales in asterids and suggested that Ebenaceae is monophyletic in Ericales.

Introduction

Diospyros, belonging to Ebenaceae, is a large genus with more than 500 species that are distributed worldwide [1]. *D. kaki* is the most important economic crop and the most widely

Haplox Biotechnology Co., Ltd. provided support in the form of salaries for author JH, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of this author are articulated in the "author contributions" section.

Competing Interests: The authors have the following interests: Jingjing Hu is employed by Haplox Biotechnology Co., Ltd. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

cultivated species of *Diospyros*. It is believed to have originated in China and has been an important food source in China, Korea, and Japan from prehistoric times [2]. The fruit of *D. kaki* is delicious and has an extensive popularity globally. In 2013, the global production of persimmon (*D. kaki*) was 4,637,357 tons, of which 78.0% was from China [3]. In addition, the fruit is used as a source of persimmon lacquer and tannin [4]. The leaves can be used as tea and are known to have phytochemical and pharmacological properties [5, 6]. At present, about 1000 cultivars exist in China [7], most of which are hexaploid, while some are nonaploid [8]. The progenitor, origin, and polyploidization mechanisms of *D. kaki* are still ambiguous; thus, identifying a closely related diploid species to be used as reference for future research is necessary. Previous studies indicated that the diploid species—*D. oleifera*, *D. lotus*, *D. glaucifolia*, and *D. 'Jinzaoshi'*—are the related species of *D. kaki* [9, 10]. They are also widely used species of *Diospyros*. *D. glaucifolia* is used as timber wood; *D. oleifera* is used as a source of tannin, whereas *D. lotus* and *D. 'Jinzaoshi'* are cultivated for their fruits. *D. 'Jinzaoshi'*, known as Jinzaoshi in China, is a controversial species. It has been accepted as a cultivar of *D. kaki*, but recent studies based on morphological as well as internal transcribed sequence (ITS) and *matK* sequence analyses proposed that *D. 'Jinzaoshi'* might be a new species [11].

In addition to these factors, the classification of *Diospyros* is very difficult because of the natural or artificial interspecific hybrids, indistinguishable morphological features across species, and the complex chromosome numbers ($2n = 2X, 4X, 6X, 9X = 30, 60, 90, 135$) [8]. The identification of the phylogenetic relationship of *Diospyros* has been attempted using various methods based on morphological characteristics [12] and molecular markers [13, 14]. Different markers yield inconsistent results, probably because of the discrepant sequence divergence ratios and tree-generating methods used. Additional markers should be detected to reveal the accurate relationship within *Diospyros* and to elucidate phylogeny within the asterids.

The chloroplast (cp) genome of higher plants has a conserved quadripartite structure with one large single-copy region (LSC: 80–90 kb) and one small single-copy region (SSC: 16–27 kb) separated by two identical inverted repeat regions (IR: 20–28 kb in length) [15, 16]. The gene content and gene order in angiosperm cp genomes are usually highly conserved, containing 110–130 distinct genes that encode 4 rRNAs, 30 tRNAs, and 80 protein-coding genes [17]. However, the angiosperm cp genome has also undergone several large mutations such as genome rearrangement and gene loss and gain in both monocots [18] and dicots [19].

Cp genomes are useful in taxonomy and evolutionary studies [20, 21] for their small size, conserved gene content and arrangement, and maternally inherited characteristics [22, 23]. The basal asterids Ericales are a large order containing more than 20 families [24]. However, complete cp genomes have been sequenced from only four families (Ericaceae, Theaceae, Actinidiaceae, and Primulaceae) [25–28]. Analysis of more cp genomes is needed for an accurate phylogeny of angiosperms. The cp genome can also be used in genetic transformation [29], agricultural trait improvement [30], and DNA barcoding [31]. Cp genome transformation is superior to nuclear transformation because of its high level of transgene expression and gene containment [32]. Complete cp genome of *Diospyros* or Ebenaceae has not yet been sequenced despite their remarkable economic value.

In this study, we sequenced complete cp genomes from five species of *Diospyros* and conducted comparative analyses within both *Diospyros* and Ericales. The comparative analyses of the cp genomes of Ebenaceae and four other families with published cp genomes were conducted to elucidate the phylogeny and genomic structures of Ericales.

Materials and Methods

Plant Materials

Healthy and young leaves were collected from adult plants of five species, *D. kaki*, *D.* ‘Jinzaoshi’, *D. glaucifolia*, *D. lotus*, and *D. oleifera*, grown in a field nursery in Yuanyang County, China. This nursery is a germplasm collection center of *Diospyros* species owned by Non-timber Forestry Research and Development Center, Chinese Academy of Forestry. Our study was permitted and approved by this authority. No endangered or protected species were sampled.

DNA Sequencing, Genome Assembly, and Validation

Total DNA was extracted from 50 g of fresh leaves using a DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). After purification, the DNA sample was randomly fragmented to construct paired-end (PE) libraries according to the Illumina preparation manual (San Diego, CA, USA). This sequencing technology was chosen because of its high accuracy in homopolymer sequencing [33] and its wide application to other plastomes [34, 35]. Accurate sequencing of mononucleotide repeats is important since they have variable lengths in different haplotypes [36].

The cp DNA was assembled as follows: all reads were filtered by trimming 20 bp from the PE reads and reads with quality score of less than 20. The clean PE reads were overlapped using FLASH ver. 1.2.6 [37] and then aligned to the cp database by using Burrows–Wheeler Aligner (BWA) software [38]. Celera Assembler [39] was used to assemble the reads into contigs, which were then scaffolded using SSPACE [40]. Mapping assembly was generated using LASTZ [41] and *Camellia yunnanensis* (NC_013707) as a reference sequence.

The gaps were filled using GapFiller [42] to obtain the complete genomes. The complete cp genome sequences were validated by designing 101 pairs of primers to obtain PCR products. Five of these primers covered the four junctions between single-copy (SC) and inverted-repeat (IR) regions. The PCR products were sequenced using Sanger sequencing and aligned to *Diospyros* cp genomes. These complete cp genomes were deposited in GenBank (S1 Table).

Gene Annotation and Repeat Identification

Gene annotation was conducted using the Dual Organellar GenoMe Annotator (DOGMA) [43]. The final annotation was obtained by manual correction based on published cp gene annotations deposited in online databases. The circular gene distribution map was drawn using OGDraw [44].

Four types of repeats—forward, reverse, complement, and palindromic—were assessed using REPuter [45] with the minimal repeat size of approximately ≥ 20 bp. Microsatellites were detected using MISA; they were defined as (unit size/minimum number of repeats) 1/10, 2/6, 3/5, 4/4, 5/3, and 6/3 [46].

Phylogenomic Analyses

Unless otherwise specified, all the multiple sequence alignments in this study were performed using Clustalw v2.0.12 with default parameters. The maximum parsimony (MP) trees were reconstructed using PAUP* v4.0b10 [47] with heuristic search and tree-bisection-reconnection (TBR) for branch-swapping settings. Gaps and multistate taxa were treated as missing and uncertainty, respectively. One tree was held at each step during stepwise addition. The Multi-Trees option was set in effect, and Steepest descent was not in effect. Before maximum likelihood (ML) analyses, the target alignment was uploaded to Cipres to identify the best model by using the Akaike information criterion (AIC) implemented in the jModelTest2 program [48].

The ML trees were reconstructed with RAxML v8.2.6 using the corresponding best model [49]. In both the MP and ML trees, bootstrap analyses were performed with 1000 replicates [50].

Results

Genome Sequencing, Assembly, and Validation

Overall, 477–1,150 million bp short reads were produced by sequencing of the five species on the Illumina HiSeq and Miseq platform. The short reads were aligned against the reference cp genome, and a total of 18.2–58.9 million bp were mapped to the reference genome, with an average of 116–376× read depth (S1 Table).

A total of 101 pairs of primers were designed to validate the genome assemblies, including the junctions between four regions in *Diospyros* cp genome (S2 Table). After PCR and Sanger sequencing, the sequences were aligned directly against the *Diospyros* genomes to correct for nucleotide mismatches or indels.

Genome Features

Diospyros cp genomes consist of two IRs (26,079–26,119 bp) segregated by two SC regions, namely, LSC (86,948–87,059 bp) and SSC (18,076–18,532 bp), thereby presenting a typical quadripartite structure (Fig 1, S3 Table). The genome structure and gene content and order were identical in the five *Diospyros* cp genomes. For each of the five *Diospyros* cp genomes, 134 functional genes were predicted (Table 1), of which 115 were unique genes (including 80 protein-coding genes, 31 transfer RNA genes, and 4 ribosomal RNA genes), and 19 were duplicated genes in the IR regions. Eighteen distinct genes contained one intron, two of which contained two introns. The *rps12* gene, similar to *Actinidia chinensis* [27], is a trans-spliced gene with the 5' end located in the LSC region and the duplicated 3' end in the IR region. As has been reported previously in other plants [51–53], we also detected several non-canonical start codons, e.g., ACG and GTG, in *ndhD* and *rps19*, respectively.

The expansion of *ycf1* into the IRa region is attributed to the formation of the *ycf1* pseudo-gene at the corresponding border of IRb and SSC (Fig 1). Such expansion has been detected in other angiosperm plastid genomes [51].

In total, 58% of the *Diospyros* cp genomes represented coding regions, whereas the remaining 42% were non-coding regions.

Repetitive Sequence

Four repeat types—forward, reverse, palindromic, and complement—were detected using REPuter [45]. The length and similarity of these sequences were more than 20 bp and 90%, respectively (S4 Table). We identified 179 repeats in the five *Diospyros* cp genomes, 100 of which were shared by all the genomes, and four, five, seven, and two repeats were specifically detected in *D. kaki*, *D. oleifera*, *D. 'Jinzaoshi'*, and *D. glaucifolia*, respectively. Palindromic repeats were the most common, accounting for 49%, followed by forward repeats (40%) and reverse repeats (10%). Only one complement repeat (20 bp) was specifically identified in the LSC region in the *D. 'Jinzaoshi'* genome. Except for a few repeats in the coding regions of *ycf2*, *ndhH*, *ndhC*, *trnS-GCU*, *trnS-UGA*, *trnfM-CAU*, *trnV-UAC*, *trnS-GGA*, *trnP-GGG*, and *trnA-UGC*, the majority were located in the noncoding regions.

In total, 53, 52, 61, 55, and 62 single sequence repeat (SSR) loci were identified in *D. kaki*, *D. 'Jinzaoshi'*, *D. lotus*, *D. oleifera*, and *D. glaucifolia* cp genomes, respectively (S5 Table). Among all mononucleotide repeats, 278 were A/T stretches, whereas only one C stretch was found in the *D. locus* and one G stretch was found in the *D. glaucifolia* cp genome. Three tetranucleotide

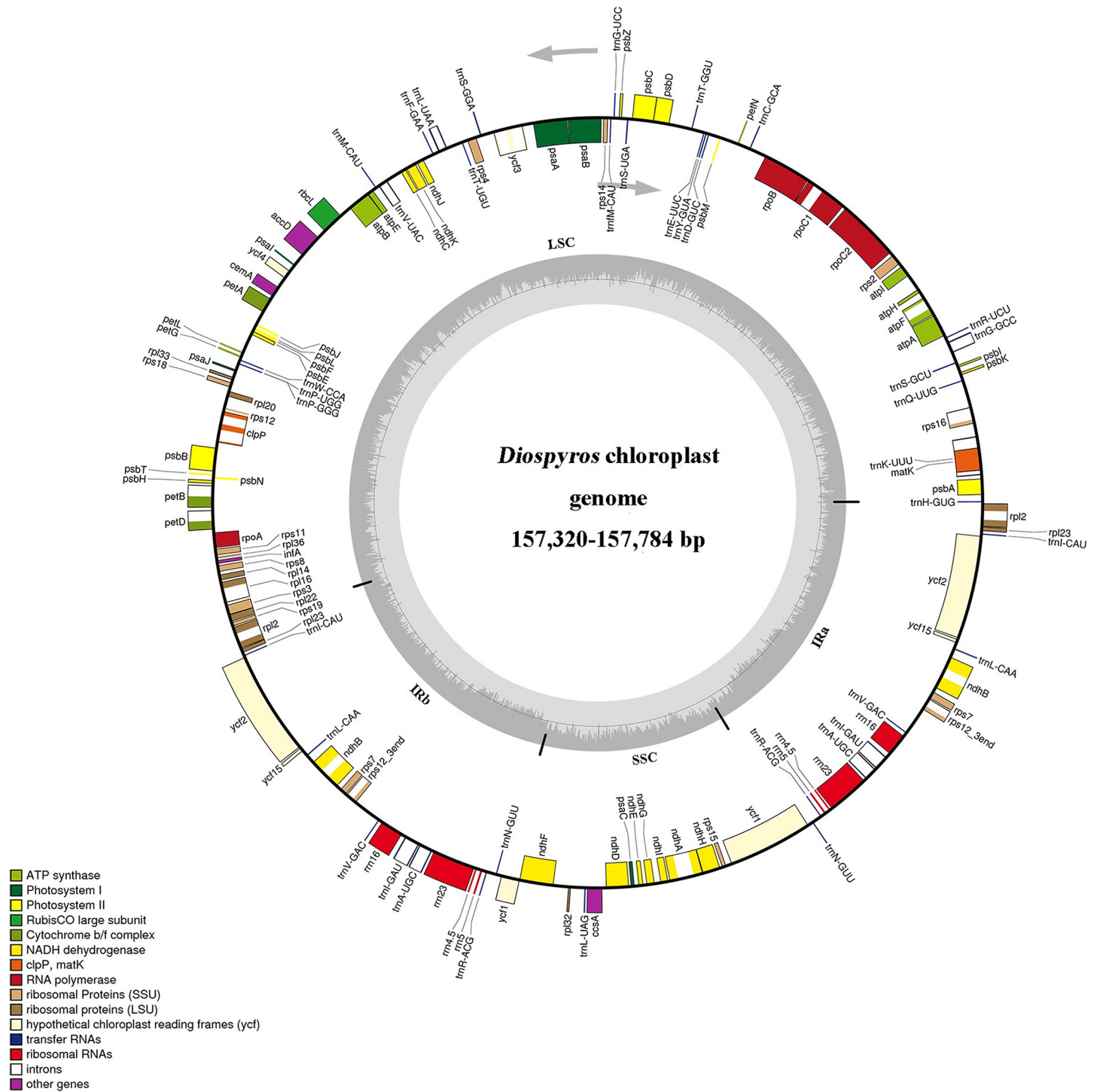


Fig 1. Gene maps of *Diospyros* chloroplast genomes. Genes on the outside of the large circle are transcribed clockwise and those on the inside are transcribed counterclockwise. The genes are color-coded based on their function. Dashed area represents the GC composition of the chloroplast genome.

doi:10.1371/journal.pone.0159566.g001

Table 1. Genes located on *Diospyros* chloroplast genomes.

Category	Gene name
Ribosomal RNAs	<i>rrn16(2), 23(2), 4.5(2), 5(2)</i>
Transfer RNAs	<i>*trnA-UGC(2), C-GCA, D-GUC, E-UUC, F-GAA, *G-GCC, G-UCC, H-GUG, I-CAU(2), *I-GAU(2), *K-UUU, L-CAA(2), *L-UAA, L-UAG, M-CAU, N-GUU(2), P-UGG, P-GGG, Q-UUG, R-UCU, R-ACG(2), S-UGA, S-GCU, S-GGA, T-GGU, T-UGU, V-GAC(2), *V-UAC, W-CCA, Y-GUA, tm-CAU</i>
Proteins of the small ribosomal subunit	<i>rps2, 3, 4, 7(2), 8, 11, *12(2), 14, 15, *16, 18, 19</i>
Proteins of the large ribosomal subunit	<i>*rpl2(2), 14, *16, 20, 22, 23(2), 32, 33, 36</i>
Subunits of RNA polymerase	<i>rpoA, B, *C1, C2</i>
Subunits of NADH- dehydrogenase	<i>*ndhA, *B(2), C, D, E, F, G, H, I, J, K</i>
Subunits of Photosystem I	<i>psaA, B, C, I, J</i>
Subunits of Photosystem II	<i>psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z</i>
Large subunit of Rubisco	<i>rbcL</i>
Subunits of cytochrome b/f complex	<i>petA, *B, *D, G, L, N</i>
Subunits of ATP synthase	<i>atpA, B, E, *F, H, I</i>
Acetyl-CoA carboxylase	<i>accD</i>
Cytochrome c biogenesis	<i>ccsA</i>
Maturase	<i>matK</i>
Protease	<i>**clpP</i>
Envelope membrane protein	<i>cemA</i>
Conserved hypothetical chloroplast reading frames	<i>ycf1(2), 2(2), **3, 4, 15(2)</i>
Translation initiation factor	<i>infA</i>

(2) indicates genes that have undergone duplication. * and ** indicate genes containing one and two introns, respectively.

doi:10.1371/journal.pone.0159566.t001

repeats (AAAT) were found only in the *D. kaki*, *D. 'Jinzaoshi'*, and *D. oleifera* cp genomes. Di-, tri-, penta-, or hexanucleotide repeats were not found. Most of the SSRs were located in the LSC (209) region, followed by those in the SSC (49) and IR (25) regions, and 67% were intergenic sequences.

Comparison of the Whole Chloroplast Genomes among Ericales

The global alignments between Ebenaceae and other published families in Ericales were performed using mVISTA [54] (Fig 2). The cp genome of *Vaccinium macrocarpon* in Ericaceae was remarkably different from that of Ebenaceae. IRs were more conserved than SCs. Unlike coding sequences, non-coding sequences exhibit a higher divergence across different species. The intergenic regions of *trnQ_rps16*, *atpI_atpH*, *psbJ_petA*, *ndhF_rpl32*, *rpl32_trnL*, *trnV_ndhC*, and *psbD_trnT* were highly variable.

Indel Identification and Relationship of the Five *Diospyros* cp Genomes

All the ML trees reconstructed based on the whole cp genome sequences, protein-coding sequences, and intergenic and intron sequences of *Diospyros* indicated that *D. kaki* was closer to *D. oleifera*, whereas *D. lotus* had a closer relationship with *D. glaucifolia* (Fig 3a, S1a and S2a

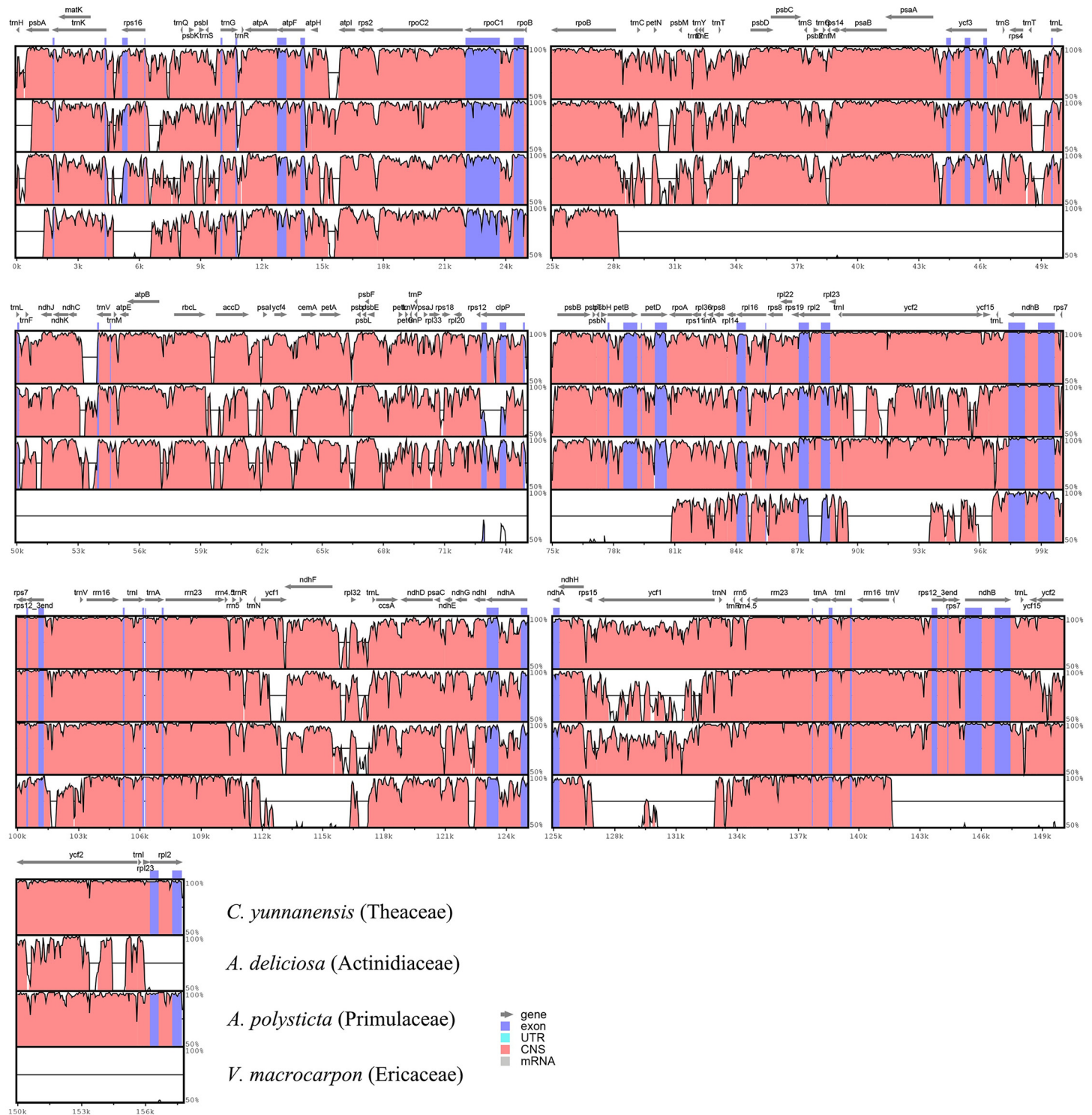


Fig 2. Global alignment of Ebenaceae genome and other published chloroplast genomes in Ericales using VISTA. Y-axis indicates the range of identity (50%–100%). Alignment was performed using *D. kaki* as a reference.

doi:10.1371/journal.pone.0159566.g002

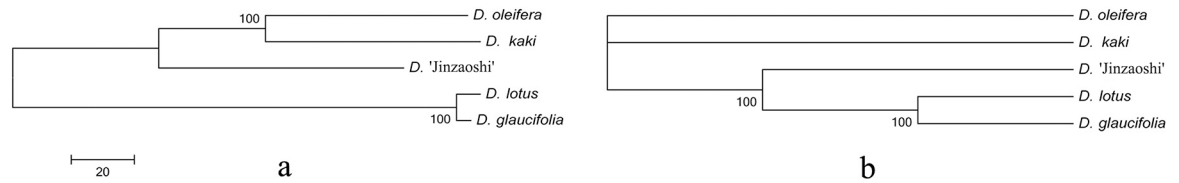


Fig 3. Phylogenetic trees based on whole genome sequences of *Diospyros*. (a) Maximum likelihood tree, (b) Maximum parsimony tree.

doi:10.1371/journal.pone.0159566.g003

Figs). MP trees reconstructed using corresponding sequences were consistent with the ML tree topology (Fig 3b, S1b and S2b Figs).

Multiple sequence alignment was performed, and indels more than 5 bp long were detected to reveal the variations within the five *Diospyros* cp genomes (Fig 4). Although the five *Diospyros* cp genomes were highly conserved, the existing differences might reveal species variation and differentiation. In total, 66 loci were identified, and the intergenic region *trnQ_rps16* with five loci was the most variable region, followed by *trnV_ndhC*(4), *ndhA*intron (4), and

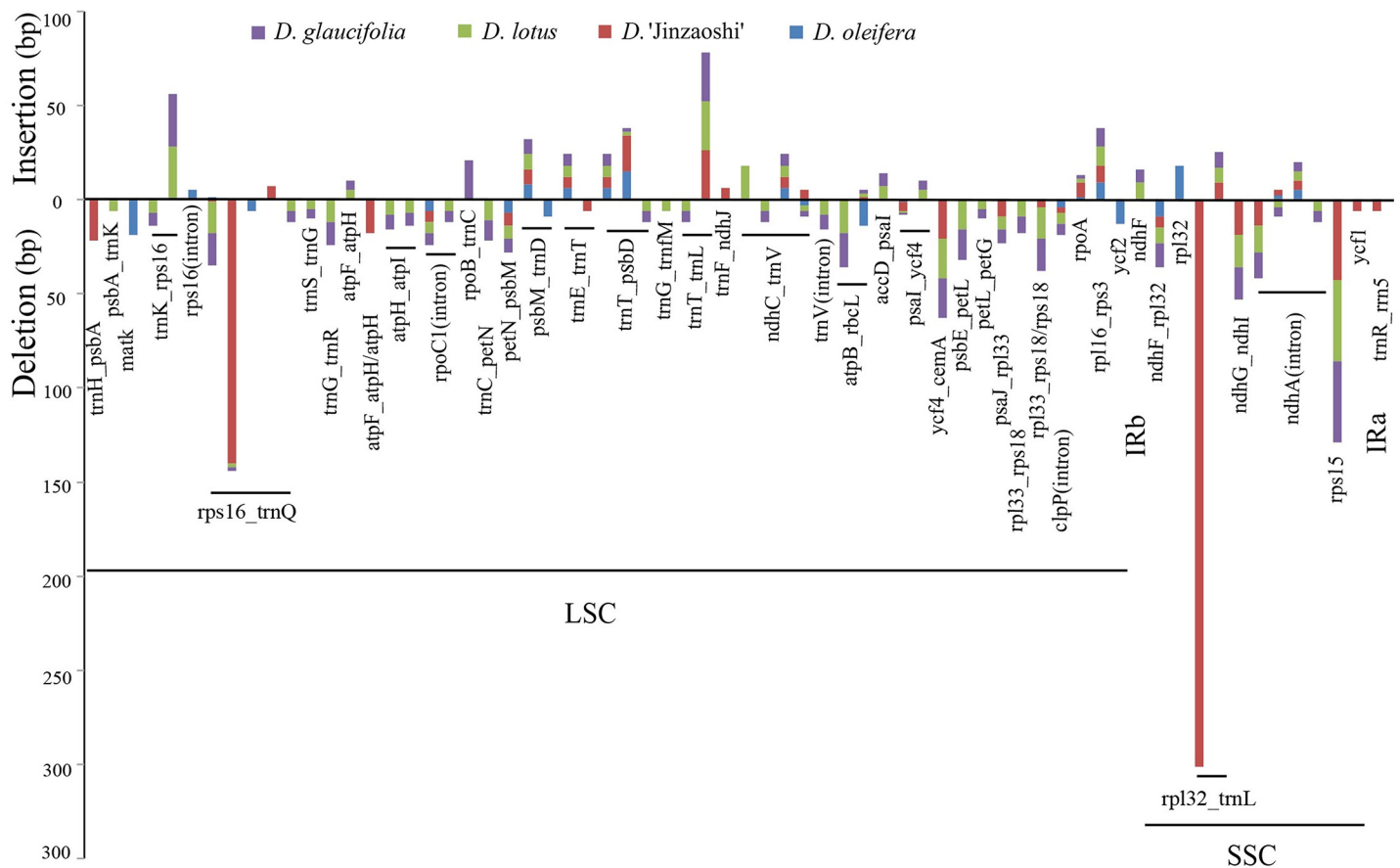


Fig 4. Indels (≥ 5 bp) identified based on multiple sequence alignment of five *Diospyros* cp genomes. Insertions are shown above and deletions below the horizontal axis. Indel distribution was positioned using *D. kaki* as a reference.

doi:10.1371/journal.pone.0159566.g004

psbD_trnT(3). The two largest indels were the deletions of 140 bp and 301 bp located in *trnQ_rps16* and *rpl32_trnL* in the cp genome of *D. 'Jinzaoshi'*, respectively. Both MP and ML trees based on the sequences of these four hypervariable regions corroborated the results based on whole cp genome sequences ([S3a and S3b Fig](#)).

Analysis of IRs

In Ebenaceae, the IRa/SSC borders were located in the 3' region of the *ycf1* gene creating the *ycf1* pseudogene at the IRb/SSC border ([Fig 5](#)). This finding is similar with those in Actinidiaceae, Theaceae, and Primulaceae but remarkably different from that in Ericaceae. The IRb/SSC borders were located upstream of the *ndhF* gene, except in Primulaceae whose IRb/SSC junction was located in the 5' region of *ndhF*. In Ebenaceae, the IRa/LSC junctions were located in the upstream region of *trnH-GUG*, similar to that in Theaceae. However, this gene was found in the IRs in Actinidiaceae and Ericaceae, as well as in most monocot cp genomes [[55](#)]. In Ebenaceae and Primulaceae, the IRb/LSC junctions were located within *rps19*, but no copy was generated in the corresponding region.

Phylogenetic Analysis

The phylogenetic relationship between *Diospyros* and other asterids was determined by collecting 18 published cp genome sequences from the GenBank of the NCBI database ([S6 Table](#)). Two cp genome sequences from *Spinacia* and *Silene* belonging to Caryophyllales were included as outgroup taxa. Sixty-one protein-coding sequences shared by these cp genomes were aligned in a single data matrix with a total of 52,294 characters included. Of all the characters, 35,097, 8414, and 8783 were constant, variable, and parsimony-informative, respectively. All the nodes in the phylogenetic tree received high bootstrap (83%–100%). The MP tree strongly indicated that Ericales is a basal sister order to the subdivision of euasterids (euasterids I and II; [Fig 6](#)) and suggested the monophyletic placement of Ebenaceae in Ericales. Lamiales, Solanales, and Gentianales were clustered into the subdivision of euasterids I, whereas Apiales and Asterales were included in euasterids II. The tree topology reconstructed using the ML method was consistent with the MP tree topology ([S4 Fig](#)).

Discussion

In this study, five sequences of *Diospyros* cp genomes were sequenced and validated using PCR-based Sanger sequencing. The complete cp genomes ranged from 157,300 to 157,784 bp, which is within the range of the cp genomes of other angiosperms [[51](#)]. Despite the occurrence of frequent large-scale genome rearrangements and gene loss-and-gain events in several lineages of land plants [[56, 57](#)], the cp genomes of *Diospyros* were highly conserved with identical gene content and gene order and genome structure comprising four parts, as noted in other angiosperms [[58](#)]. Similar to previously published asterid plastid genomes [[59, 60](#)], the *Diospyros* cp genome contained more AT and had a GC content of 37%.

SSRs are widely used markers in population genetics [[61, 62](#)] and in phylogenetic investigations [[63, 64](#)] because of their high polymorphism even within species. A total of 283 SSR loci were identified in the five *Diospyros* cp genomes; most of them were intergenic sequences, indicating numerous variations in these regions. Most of the mononucleotide repeats were A/T stretches, contributing to the rich A/T content in the cp genomes of *Diospyros* and suggesting that most of the cp SSRs are short polyadenine (polyA) or polythymine (polyT) repeats [[34](#)]. Thus, *Diospyros* cp microsatellites might be useful tools in ecological and evolutionary studies, which warrants further research.

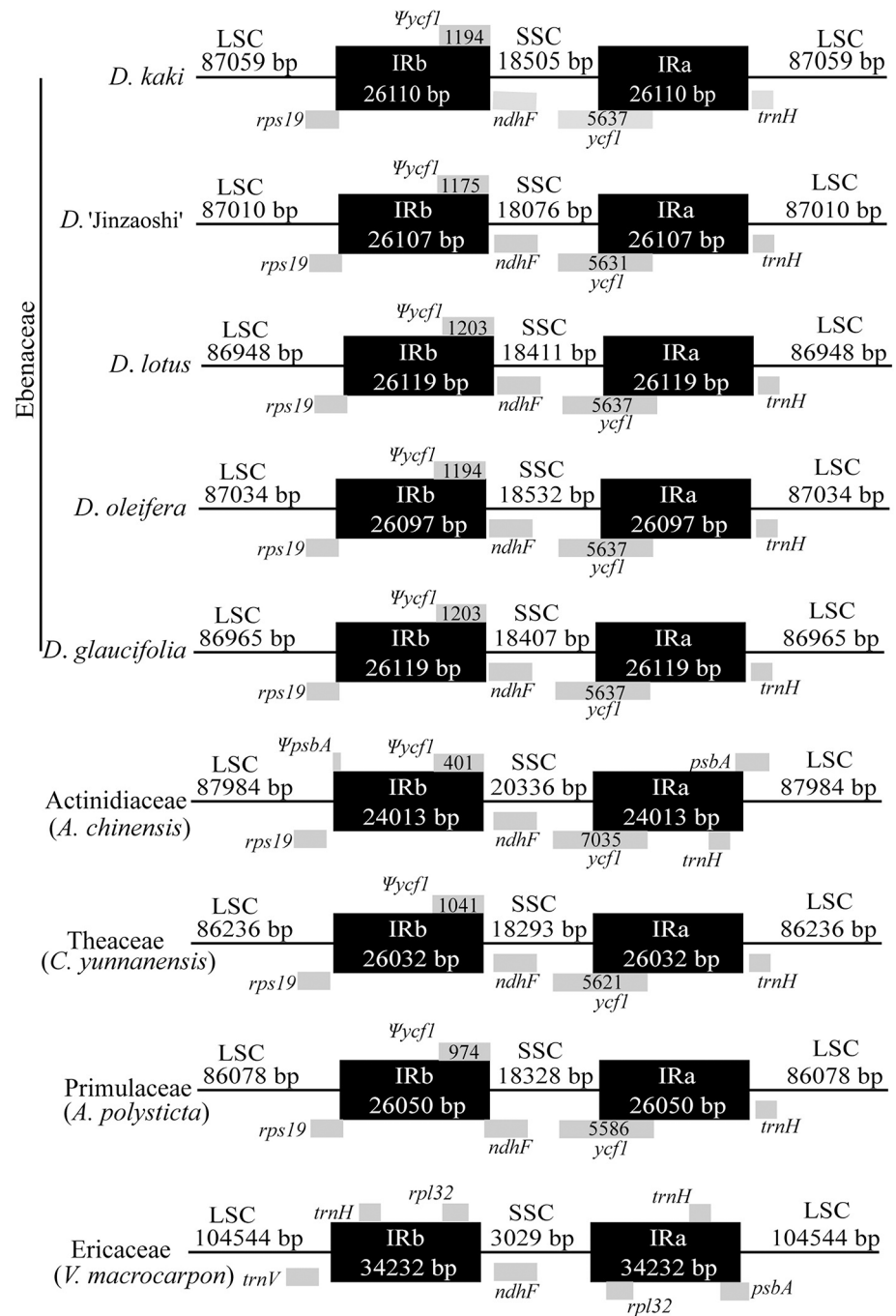


Fig 5. The comparison of inverted-repeat (IR) and single-copy (SC) borders among nine chloroplast genomes. Gene annotation or portions are represented by gray boxes above or below.

doi:10.1371/journal.pone.0159566.g005

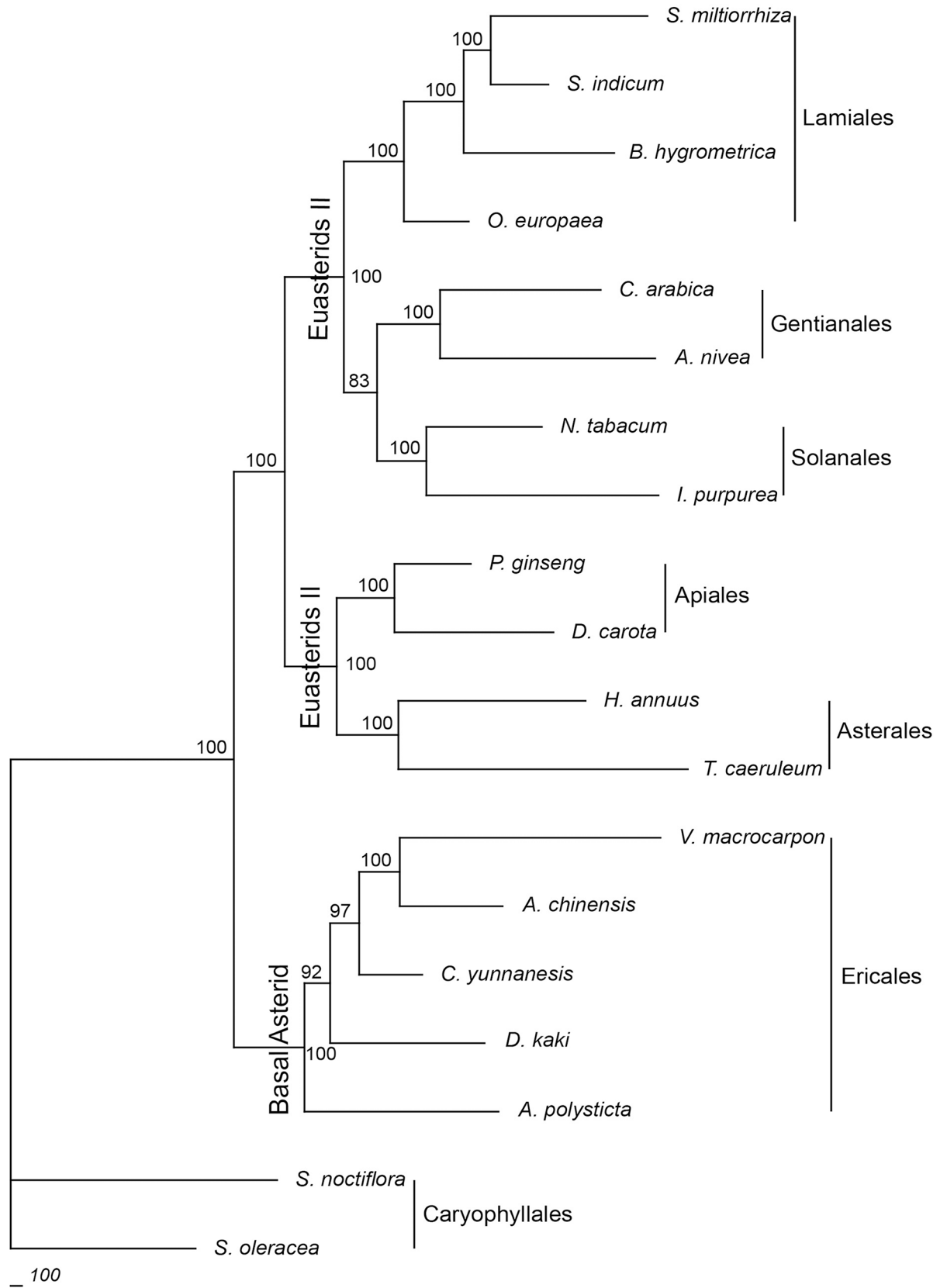


Fig 6. Phylogenetic tree of the asterid clade. The tree was reconstructed based on 61 protein-coding sequences shared by 19 angiosperm species. The numbers at the nodes indicate bootstrap values (1000 replications).

doi:10.1371/journal.pone.0159566.g006

Global alignment between Ebenaceae and other published cp genomes in Ericales indicated that the IR regions were more conserved, probably because of copy correction by gene conversion when mutations are introduced into IRs [65]. The significant difference between the cp genome of *V. macrocarpon* and that of other species might have been caused by multiple structural rearrangements in its cp genome [25]. Seven intergenic regions with rich variation were included in the 13 hotspots reported in the plastid genomes of several plants, including asterids [66]. These regions could be developed as interspecific DNA markers for the phylogenetic analysis in Ericales.

The IR regions play an important role in stabilizing plastid genome structure [67]. Although IRs are highly conserved, IR contraction and expansion events are common in the evolutionary history and are mainly responsible for length mutations of plastid genomes [51, 68]. In this study, we compared the IR/SC junctions within Ericales. The IR/SC junctions of *Diospyros* were similar and showed little difference with those of Actinidiaceae, Theaceae, and Primulaceae. The cp genome of Ericaceae was significantly different from those of others, further confirming the rearrangements during its evolution [25]. Our results indicated that the cp genomes might be conserved in closely related species, whereas species belonging to different families might have greater diversity, such as the large inversions in the cp genome of *Eucommia ulmoides* [69] and one inverted repeat loss in *Astragalus membranaceus* [70].

Phylogenetic trees reconstructed using different sequences indicated the closer relationship between *D. kaki* and *D. oleifera*. This finding is consistent with that of our previous study based on SSR and ITS regions (S5 and S6 Figs) [71, 72] and with that of a study investigating taxonomy based on morphology [73]. The morphological characteristics of *D. oleifera* are similar to those of *D. kaki*: both have pistillate flowers, styles that are parted, and branches without pellicle. However, the branches of *D. lotus* and *D. glaucifolia* are covered with pellicle and pistillate flower styles are parted halfway. In *D. 'Jinzaoshi'*, the branches are covered with pellicle, but the pistillate flower style is joined (for more details, see [11, 73]). Multiple sequence alignment among five *Diospyros* genomes indicated that most of the indels were intergenic sequences located in the LSC and SSC regions, which is consistent with the findings of previous studies suggesting that SC regions are less conserved than IR regions [58, 74, 75]. The large deletions identified in the cp genome of *D. 'Jinzaoshi'* might have been caused by slipped-strand mispairing [76] or illegitimate recombination events [77–79]. The indels identified in the *Diospyros* cp genomes might have numerous important applications in systematics and evolutionary biology, such as elucidating the origin of domesticated species [80], tracing biogeographic movements [81–83], and clarifying complex relationships among species [84]. Furthermore, these hotspot regions could be used to determine the molecular phylogeny of other *Diospyros* species. Previous studies based on morphological as well as ITS and *matK* sequence analyses proposed that “Jinzaoshi” does not belong to *D. kaki* and other related *Diospyros* species and might be a new species of *Diospyros* [11]. The two large deletions in the cp genome of *D. 'Jinzaoshi'* and the phylogenetic trees inferred from the five *Diospyros* cp genomes indicated that *D. 'Jinzaoshi'* is a new species and should be named in the future.

Both tree topologies reconstructed using the MP and ML methods confirmed the basal position of Ericales in asterids and the subdivision of this clade. This is consistent with the findings of a previous phylogenetic analysis based on the complete cp genomes of 15 asterid species and

one outgroup [27]. Thirteen out of 16 nodes in the MP tree received a bootstrap support of 100%, suggesting that proper settings were used during the reconstruction. Ebenaceae was resolved monophyletic, which corroborated the findings of a previous study based on five genes from the plastid and mitochondrial genomes [85]. Numerous studies use DNA sequences from complete cp genomes to estimate phylogenetic classification of angiosperms [86, 87]. Completely sequenced cp genomes comprise abundant phylogenetic information, and several complete cp genome sequences have been successfully applied to study the phylogenetic relationships among angiosperms [21, 87]. Better understanding of the evolutionary history of asterids requires expanded range of sampling.

Conclusion

To our knowledge, this is the first report of the complete cp genome sequence of Ebenaceae. The sequences of the complete cp genomes of *Diospyros* and sequencing and assembly strategies can be used as a reference for future cp genome sequencing within Ebenaceae, or even Ericales. The available plastid genomes contain sufficient phylogenetic information to resolve interspecific relationships, conduct phylogenetic and classification analyses, and trace the origin of *Diospyros*, in particular, of economically important plants. Since the majority of *D. kaki* are hexaploid, with a few being nonaploid [8], further investigation of its genetic background is challenging, especially the whole-genome sequencing. *D. oleifera* could be considered as a model plant to study *D. kaki* and its cultivars. Furthermore, our study findings confirmed that *D. 'Jinzaoshi'* is a new species and indicated that the complete cp sequences might provide a practical and efficient approach to clarify the phylogenetic relationships among *Diospyros* species.

Supporting Information

S1 Table. Statistical analysis of the sequencing information.

(XLSX)

S2 Table. Primers used for assembly and junction verification.

(XLSX)

S3 Table. Genomic features of the five *Diospyros* species.

(XLSX)

S4 Table. Results of the repeated statistical analysis.

(XLSX)

S5 Table. Single sequence repeats identified in *Diospyros* genomes.

(XLSX)

S6 Table. Accession numbers of the chloroplast genome sequences used in this study.

(XLSX)

S1 Fig. Phylogenetic trees reconstructed based on 80 protein-coding sequences of *Diospyros*. (a) Maximum likelihood tree (b) Maximum parsimony tree.

(TIF)

S2 Fig. Phylogenetic trees reconstructed based on intergenic and intron sequences of *Diospyros*. (a) Maximum likelihood tree (b) Maximum parsimony tree.

(TIF)

S3 Fig. Phylogenetic trees reconstructed based on 4 hypervariable sequences of *Diospyros*.

(a) Maximum likelihood tree (b) Maximum parsimony tree.

(TIF)

S4 Fig. Maximum likelihood tree reconstructed based on 61 protein-coding sequences shared by 19 angiosperm species.

(TIF)

S5 Fig. Phylogenetic tree constructed based on the single sequence repeat sequences.

(TIF)

S6 Fig. Phylogenetic tree constructed based on the internal transcribed spacer region sequences.

(TIF)

Acknowledgments

This research was supported by the National Key Technology R&D Program in the 12th Five-year Plan of China (No., 2013BAD14B0502).

Author Contributions

Conceived and designed the experiments: T-NW X-FT. Performed the experiments: Y-QL J-JL. Analyzed the data: J-MF H-ML. Contributed reagents/materials/analysis tools: J-MF Y-QL J-JL J-JH. Wrote the paper: J-MF H-ML.

References

1. Duangjai S, Samuel R, Munzinger J, Forest F, Wallnöfer B, Barfuss MH, et al. A multi-locus plastid phylogenetic analysis of the pantropical genus *Diospyros* (Ebenaceae), with an emphasis on the radiation and biogeographic origins of the New Caledonian endemic species. *Molecular Phylogenetics and Evolution*. 2009; 52(3):602–620. doi: [10.1016/j.ympev.2009.04.021](https://doi.org/10.1016/j.ympev.2009.04.021) PMID: [19427384](https://pubmed.ncbi.nlm.nih.gov/19427384/)
2. Yonemori K, Sugiura A, Yamada M. Persimmon genetics and breeding. *Plant Breeding Reviews*, Vol 19. 2000; p. 191–225.
3. Available from: <http://faostat3.fao.org/download/Q/QV/E>
4. Luo Z, Wang R. Persimmon in China: domestication and traditional utilizations of genetic resources. *Advances in Horticultural Science*. 2008; 22(4):239–243.
5. Kawakami K, Aketa S, Sakai H, Watanabe Y, Nishida H, Hirayama M. Antihypertensive and vasorelaxant effects of water-soluble proanthocyanidins from persimmon leaf tea in spontaneously hypertensive rats. *Bioscience, biotechnology, and biochemistry*. 2011; 75(8):1435–1439. doi: [10.1271/bbb.100926](https://doi.org/10.1271/bbb.100926) PMID: [21821959](https://pubmed.ncbi.nlm.nih.gov/21821959/)
6. Xie C, Xie Z, Xu X, Yang D. Persimmon (*Diospyros kaki* L.) leaves: A review on traditional uses, phytochemistry and pharmacological properties. *Journal of ethnopharmacology*. 2015; 163:229–240. doi: [10.1016/j.jep.2015.01.007](https://doi.org/10.1016/j.jep.2015.01.007) PMID: [25637828](https://pubmed.ncbi.nlm.nih.gov/25637828/)
7. Renzi W, Yong Y, Gaochao L. CHINESE PERSIMMON GERMPLASM RESOURCES. In: I International Persimmon Symposium 436; 1996. p. 43–50.
8. Zhuang DH, Kitajima A, Ishida M, Sobajima Y. Chromosome numbers of *Diospyros kaki* cultivars. *Journal of the Japanese Society for Horticultural Science*. 1990; 59(2):289–297. doi: [10.2503/jjshs.59.289](https://doi.org/10.2503/jjshs.59.289)
9. Du X, Zhang Q, Luo ZR. Comparison of four molecular markers for genetic analysis in *Diospyros* L. (Ebenaceae). *Plant systematics and evolution*. 2009; 281(1-4):171–181. doi: [10.1007/s00606-009-0199-z](https://doi.org/10.1007/s00606-009-0199-z)
10. Guo D, Luo Z. Genetic relationships of the Japanese persimmon *Diospyros kaki* (Ebenaceae) and related species revealed by SSR analysis. *Genet Mol Res*. 2011; 10(2):1060–1068. doi: [10.4238/vol10-2gmr1100](https://doi.org/10.4238/vol10-2gmr1100) PMID: [21710456](https://pubmed.ncbi.nlm.nih.gov/21710456/)
11. Tang D, Hu Y, Zhang Q, Yang Y, Luo Z. Discriminant analysis of “Jinzaoshi” from persimmon (*Diospyros kaki* Thunb.; Ebenaceae): A comparative study conducted based on morphological as well as

- ITS and matK sequence analyses. *Scientia Horticulturae*. 2014; 168:168–174. doi: [10.1016/j.scienta.2014.01.033](https://doi.org/10.1016/j.scienta.2014.01.033)
12. Venkatasamy S, Khittoo G, Nowbuth P, Vencatasamy DR. Phylogenetic relationships based on morphology among the *Diospyros* (Ebenaceae) species endemic to the Mascarene Islands. *Botanical Journal of the Linnean Society*. 2006; 150(3):307–313. doi: [10.1111/j.1095-8339.2006.00474.x](https://doi.org/10.1111/j.1095-8339.2006.00474.x)
 13. Hu D, Zhang Q, Luo Z. Phylogenetic analysis in some *Diospyros* spp. (Ebenaceae) and Japanese persimmon using chloroplast DNA PCR-RFLP markers. *Scientia horticulturae*. 2008; 117(1):32–38. doi: [10.1016/j.scienta.2008.02.001](https://doi.org/10.1016/j.scienta.2008.02.001)
 14. Yonemori K, Honsho C, Kanzaki S, Ino H, Ikegami A, Kitajima A, et al. Sequence analyses of the ITS regions and the matK gene for determining phylogenetic relationships of *Diospyros kaki* (persimmon) with other wild *Diospyros* (Ebenaceae) species. *Tree Genetics & Genomes*. 2008; 4(2):149–158. doi: [10.1007/s11295-007-0096-y](https://doi.org/10.1007/s11295-007-0096-y)
 15. Jansen RK, Raubeson LA, Boore JL, Depamphilis CW, Chumley TW, Haberle RC, et al. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in enzymology*. 2005; 395:348–384. doi: [10.1016/S0076-6879\(05\)95020-9](https://doi.org/10.1016/S0076-6879(05)95020-9) PMID: [15865976](https://pubmed.ncbi.nlm.nih.gov/15865976/)
 16. Palmer JD, Stein DB. Conservation of chloroplast genome structure among vascular plants. *Current genetics*. 1986; 10(11):823–833. doi: [10.1007/BF00418529](https://doi.org/10.1007/BF00418529)
 17. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, et al. The complete chloroplast genome sequence of *Pelargonium × hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular biology and evolution*. 2006; 23(11):2175–2190. doi: [10.1093/molbev/msl089](https://doi.org/10.1093/molbev/msl089) PMID: [16916942](https://pubmed.ncbi.nlm.nih.gov/16916942/)
 18. Mardanov AV, Ravin NV, Kuznetsov BB, Samigullin TH, Antonov AS, Kolganova TV, et al. Complete sequence of the Duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. *Journal of molecular evolution*. 2008; 66(6):555–564. doi: [10.1007/s00239-008-9091-7](https://doi.org/10.1007/s00239-008-9091-7) PMID: [18463914](https://pubmed.ncbi.nlm.nih.gov/18463914/)
 19. Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Molecular biology and evolution*. 2011; 28(1):583–600. doi: [10.1093/molbev/msq229](https://doi.org/10.1093/molbev/msq229) PMID: [20805190](https://pubmed.ncbi.nlm.nih.gov/20805190/)
 20. Matsuoka Y, Yamazaki Y, Ogihara Y, Tsunewaki K. Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. *Molecular Biology and Evolution*. 2002; 19(12):2084–2091. doi: [10.1093/oxfordjournals.molbev.a004033](https://doi.org/10.1093/oxfordjournals.molbev.a004033) PMID: [12446800](https://pubmed.ncbi.nlm.nih.gov/12446800/)
 21. Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC biology*. 2009; 7(1):84. doi: [10.1186/1741-7007-7-84](https://doi.org/10.1186/1741-7007-7-84) PMID: [19954512](https://pubmed.ncbi.nlm.nih.gov/19954512/)
 22. Bock R. Structure, function, and inheritance of plastid genomes. In: *Cell and molecular biology of plastids*. Berlin and Heidelberg: Springer; 2007. p. 29–63.
 23. Zhang Q, Sodmergen. Why does biparental plastid inheritance revive in angiosperms? *Journal of plant research*. 2010; 123(2):201–206. doi: [10.1007/s10265-009-0291-z](https://doi.org/10.1007/s10265-009-0291-z) PMID: [20052516](https://pubmed.ncbi.nlm.nih.gov/20052516/)
 24. Bremer B, Bremer K, Chase M, Fay M, Reveal J, Soltis D, et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*. 2009.
 25. Fajardo D, Senalik D, Ames M, Zhu H, Steffan SA, Harbut R, et al. Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. *Tree Genetics & Genomes*. 2013; 9(2):489–498. doi: [10.1007/s11295-012-0573-9](https://doi.org/10.1007/s11295-012-0573-9)
 26. Yang JB, Yang SX, Li HT, Yang J, Li DZ. Comparative chloroplast genomes of *Camellia* species. *PLoS One*. 2013; 8(8):e73053. doi: [10.1371/journal.pone.0073053](https://doi.org/10.1371/journal.pone.0073053) PMID: [24009730](https://pubmed.ncbi.nlm.nih.gov/24009730/)
 27. Yao X, Tang P, Li Z, Li D, Liu Y, Huang H. The First Complete Chloroplast Genome Sequences in Actinidiaceae: Genome Structure and Comparative Analysis. *PloS one*. 2015; 10(6):e0129347. doi: [10.1371/journal.pone.0129347](https://doi.org/10.1371/journal.pone.0129347) PMID: [26046631](https://pubmed.ncbi.nlm.nih.gov/26046631/)
 28. Ku C, Hu JM, Kuo CH. Complete plastid genome sequence of the basal asterid *Ardisia polysticta* Miq. and comparative analyses of asterid plastid genomes. *PLoS One*. 2013; 8(4):e62548. doi: [10.1371/journal.pone.0062548](https://doi.org/10.1371/journal.pone.0062548) PMID: [23638113](https://pubmed.ncbi.nlm.nih.gov/23638113/)
 29. Maliga P. Engineering the plastid genome of higher plants. *Current opinion in plant biology*. 2002; 5(2):164–172. doi: [10.1016/S1369-5266\(02\)00248-0](https://doi.org/10.1016/S1369-5266(02)00248-0) PMID: [11856614](https://pubmed.ncbi.nlm.nih.gov/11856614/)
 30. Daniell H, Ruiz ON, Dhingra A. Chloroplast genetic engineering to improve agronomic traits. In: *Transgenic plants: methods and protocols*. vol 286. 2004; p. 111–137. doi: [10.1385/1-59259-827-7:111](https://doi.org/10.1385/1-59259-827-7:111)

31. Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, et al. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One*. 2008; 3(7):e2802. doi: [10.1371/journal.pone.0002802](https://doi.org/10.1371/journal.pone.0002802) PMID: [18665273](https://pubmed.ncbi.nlm.nih.gov/18665273/)
32. Daniell H, Cohill PR, Kumar S, Dufourmantel N. Chloroplast Genetic Engineering. In: Daniell H CC, editor. *Molecular Biology and Biotechnology of Plant Organelles*. Netherlands: Springer; 2004. p. 443–490.
33. Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS one*. 2012; 7(2): e30087. doi: [10.1371/journal.pone.0030087](https://doi.org/10.1371/journal.pone.0030087) PMID: [22347999](https://pubmed.ncbi.nlm.nih.gov/22347999/)
34. Kuang DY, Wu H, Wang YL, Gao LM, Zhang SZ, Lu L, et al. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome*. 2011; 54(8):663–673. doi: [10.1139/g11-026](https://doi.org/10.1139/g11-026) PMID: [21793699](https://pubmed.ncbi.nlm.nih.gov/21793699/)
35. Lin CP, Wu CS, Huang YY, Chaw SM. The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. *Genome biology and evolution*. 2012; 4(3):374–381. doi: [10.1093/gbe/evs021](https://doi.org/10.1093/gbe/evs021) PMID: [22403032](https://pubmed.ncbi.nlm.nih.gov/22403032/)
36. Provan J, Powell W, Hollingsworth PM. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution*. 2001; 16(3):142–147. doi: [10.1016/S0169-5347\(00\)02097-8](https://doi.org/10.1016/S0169-5347(00)02097-8)
37. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011; 27(21):2957–2963. doi: [10.1093/bioinformatics/btr507](https://doi.org/10.1093/bioinformatics/btr507) PMID: [21903629](https://pubmed.ncbi.nlm.nih.gov/21903629/)
38. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–1760. doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/)
39. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, et al. A whole-genome assembly of *Drosophila*. *Science*. 2000; 287(5461):2196–2204. doi: [10.1126/science.287.5461.2196](https://doi.org/10.1126/science.287.5461.2196) PMID: [10731133](https://pubmed.ncbi.nlm.nih.gov/10731133/)
40. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*. 2011; 27(4):578–579. doi: [10.1093/bioinformatics/btq683](https://doi.org/10.1093/bioinformatics/btq683) PMID: [21149342](https://pubmed.ncbi.nlm.nih.gov/21149342/)
41. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, et al. Human–mouse alignments with BLASTZ. *Genome research*. 2003; 13(1):103–107. doi: [10.1101/gr.809403](https://doi.org/10.1101/gr.809403) PMID: [12529312](https://pubmed.ncbi.nlm.nih.gov/12529312/)
42. Nadalin F, Vezzi F, Policriti A. GapFiller: a *de novo* assembly approach to fill the gap within paired reads. *BMC bioinformatics*. 2012; 13(Suppl 14):S8. doi: [10.1186/1471-2105-13-S14-S8](https://doi.org/10.1186/1471-2105-13-S14-S8) PMID: [23095524](https://pubmed.ncbi.nlm.nih.gov/23095524/)
43. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004; 20(17):3252–3255. doi: [10.1093/bioinformatics/bth352](https://doi.org/10.1093/bioinformatics/bth352) PMID: [15180927](https://pubmed.ncbi.nlm.nih.gov/15180927/)
44. Lohse M, Drechsel O, Bock R. Organellar Genome DRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current genetics*. 2007; 52(5-6):267–274. doi: [10.1007/s00294-007-0161-y](https://doi.org/10.1007/s00294-007-0161-y) PMID: [17957369](https://pubmed.ncbi.nlm.nih.gov/17957369/)
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 research*. 2001; 29(22):4633–4642. doi: [10.1093/nar/29.22.4633](https://doi.org/10.1093/nar/29.22.4633) PMID: [11713313](https://pubmed.ncbi.nlm.nih.gov/11713313/)
46. Available from: <http://pgrc.ipk-gatersleben.de/misa/misa.html>
47. PAUP DS. *Phylogenetic analysis using parsimony (* and other methods)*. Sunderland, MA: Sinauer Associates. 2003.
48. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Gateway Computing Environments Workshop (GCE)*, 2010. IEEE; 2010. p. 1–8.
49. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014; 30(9):1312–1313. doi: [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033) PMID: [24451623](https://pubmed.ncbi.nlm.nih.gov/24451623/)
50. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985; p. 783–791. doi: [10.2307/2408678](https://doi.org/10.2307/2408678)
51. Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, et al. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS one*. 2010; 5(9):e12762. doi: [10.1371/journal.pone.0012762](https://doi.org/10.1371/journal.pone.0012762) PMID: [20856810](https://pubmed.ncbi.nlm.nih.gov/20856810/)
52. 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 genomics*. 2007; 8(1):174. doi: [10.1186/1471-2164-8-174](https://doi.org/10.1186/1471-2164-8-174) PMID: [17573971](https://pubmed.ncbi.nlm.nih.gov/17573971/)
53. Gao L, Yi X, Yang YX, Su YJ, Wang T. Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. *BMC evolutionary biology*. 2009; 9(1):130. doi: [10.1186/1471-2148-9-130](https://doi.org/10.1186/1471-2148-9-130) PMID: [19519899](https://pubmed.ncbi.nlm.nih.gov/19519899/)

54. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic acids research*. 2004; 32(suppl 2):W273–W279. doi: [10.1093/nar/gkh458](https://doi.org/10.1093/nar/gkh458) PMID: [15215394](https://pubmed.ncbi.nlm.nih.gov/15215394/)
55. Huotari T, Korpelainen H. Complete chloroplast genome sequence of *Elodea canadensis* and comparative analyses with other monocot plastid genomes. *Gene*. 2012; 508(1):96–105. doi: [10.1016/j.gene.2012.07.020](https://doi.org/10.1016/j.gene.2012.07.020) PMID: [22841789](https://pubmed.ncbi.nlm.nih.gov/22841789/)
56. Palmer JD. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *American Naturalist*. 1987; p. S6–S29. doi: [10.1086/284689](https://doi.org/10.1086/284689)
57. Knox EB, Palmer JD. Chloroplast DNA evidence on the origin and radiation of the giant lobelias in eastern Africa. *Systematic Botany*. 1998; p. 109–149. doi: [10.2307/2419583](https://doi.org/10.2307/2419583)
58. Palmer JD. Plastid chromosomes: structure and evolution. *The molecular biology of plastids*. 1991; 7:5–53. doi: [10.1016/B978-0-12-715007-9.50009-8](https://doi.org/10.1016/B978-0-12-715007-9.50009-8)
59. Yi DK, Kim KJ. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PloS one*. 2012; 7(5):e35872. doi: [10.1371/journal.pone.0035872](https://doi.org/10.1371/journal.pone.0035872) PMID: [22606240](https://pubmed.ncbi.nlm.nih.gov/22606240/)
60. Mariotti R, Cultrera NG, Díez CM, Baldoni L, Rubini A. Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. *BMC plant biology*. 2010; 10(1):211. doi: [10.1186/1471-2229-10-211](https://doi.org/10.1186/1471-2229-10-211) PMID: [20868482](https://pubmed.ncbi.nlm.nih.gov/20868482/)
61. Grassi F, Labra M, Scienza A, Imazio S. Chloroplast SSR markers to assess DNA diversity in wild and cultivated grapevines. *VITIS—Journal of Grapevine Research*. 2015; 41(3):157.
62. Powell W, Morgante M, McDevitt R, Vendramin G, Rafalski J. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proceedings of the National Academy of Sciences*. 1995; 92(17):7759–7763. doi: [10.1073/pnas.92.17.7759](https://doi.org/10.1073/pnas.92.17.7759)
63. Pauwels M, Vekemans X, Godé C, Frérot H, Castric V, Saumitou-Laprade P. Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae). *New Phytologist*. 2012; 193(4):916–928. doi: [10.1111/j.1469-8137.2011.04003.x](https://doi.org/10.1111/j.1469-8137.2011.04003.x) PMID: [22225532](https://pubmed.ncbi.nlm.nih.gov/22225532/)
64. Xue J, Wang S, Zhou SL. Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). *American journal of botany*. 2012. doi: [10.3732/ajb.1100547](https://doi.org/10.3732/ajb.1100547)
65. Khakhlova O, Bock R. Elimination of deleterious mutations in plastid genomes by gene conversion. *The Plant Journal*. 2006; 46(1):85–94. doi: [10.1111/j.1365-313X.2006.02673.x](https://doi.org/10.1111/j.1365-313X.2006.02673.x) PMID: [16553897](https://pubmed.ncbi.nlm.nih.gov/16553897/)
66. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany*. 2007; 94(3):275–288. doi: [10.3732/ajb.94.3.275](https://doi.org/10.3732/ajb.94.3.275) PMID: [21636401](https://pubmed.ncbi.nlm.nih.gov/21636401/)
67. Maréchal A, Brisson N. Recombination and the maintenance of plant organelle genome stability. *New Phytologist*. 2010; 186(2):299–317. doi: [10.1111/j.1469-8137.2010.03195.x](https://doi.org/10.1111/j.1469-8137.2010.03195.x) PMID: [20180912](https://pubmed.ncbi.nlm.nih.gov/20180912/)
68. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC evolutionary biology*. 2008; 8(1):36. doi: [10.1186/1471-2148-8-36](https://doi.org/10.1186/1471-2148-8-36) PMID: [18237435](https://pubmed.ncbi.nlm.nih.gov/18237435/)
69. Wang L, Wuyun Tn, Du H, Wang D, Cao D. Complete chloroplast genome sequences of *Eucommia ulmoides*: genome structure and evolution. *Tree Genetics & Genomes*. 2016; 12(1):1–15.
70. Lei W, Ni D, Wang Y, Shao J, Wang X, Yang D, et al. Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of *Astragalus membranaceus*. *Scientific reports*. 2016; 6. doi: [10.1038/srep21669](https://doi.org/10.1038/srep21669)
71. Liang Y, Han W, Sun P, Liang J, Wuyun T, Li F, et al. Genetic diversity among germplasms of *Diospyros kaki* based on SSR markers. *Scientia Horticulturae*. 2015; 186:180–189. doi: [10.1016/j.scienta.2015.02.015](https://doi.org/10.1016/j.scienta.2015.02.015)
72. Fu J, Liang J, Wuyun T, Liang Y, Sun P, Li F. Sequence Analysis of ITS Regions and *ndhA* Gene for Determining Phylogenetic Relationship of *Diospyros kaki* (persimmon) with Other Related Wild *Diospyros* (Ebenaceae) Species. *Bulletin of Botanical Research*. 2015; 35(4):515–520.
73. Peng Z, Zhuang X, Li S. *Flora Peipublicae Popularis Sinicae*. vol. 60. Beijing: Science Press; 1987.
74. Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proceedings of the National Academy of Sciences*. 1989; 86(16):6201–6205. doi: [10.1073/pnas.86.16.6201](https://doi.org/10.1073/pnas.86.16.6201)
75. Kim YK, Park Cw, Kim KJ. Complete chloroplast DNA sequence from a Korean endemic genus, *Megaleranthis saniculifolia*, and its evolutionary implications. *Molecules and cells*. 2009; 27(3):365–381. doi: [10.1007/s10059-009-0047-6](https://doi.org/10.1007/s10059-009-0047-6) PMID: [19326085](https://pubmed.ncbi.nlm.nih.gov/19326085/)
76. Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular biology and evolution*. 1987; 4(3):203–221. PMID: [3328815](https://pubmed.ncbi.nlm.nih.gov/3328815/)

77. Milligan BG, Hampton JN, Palmer JD. Dispersed repeats and structural reorganization in subclover chloroplast DNA. *Molecular biology and evolution*. 1989; 6(4):355–368. PMID: [2615639](#)
78. Ogihara Y, Terachi T, Sasakuma T. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proceedings of the National Academy of Sciences*. 1988; 85(22):8573–8577. doi: [10.1073/pnas.85.22.8573](#)
79. Shimada H, Sugiura M. Pseudogenes and short repeated sequences in the rice chloroplast genome. *Current genetics*. 1989; 16(4):293–301. doi: [10.1007/BF00422116](#) PMID: [2627714](#)
80. Wills DM, Burke JM. Chloroplast DNA variation confirms a single origin of domesticated sunflower (*Helianthus annuus* L.). *Journal of Heredity*. 2006; 97(4):403–408. doi: [10.1093/jhered/esl001](#) PMID: [16740625](#)
81. Ickert-Bond SM, Wen J. Phylogeny and biogeography of Altingiaceae: evidence from combined analysis of five non-coding chloroplast regions. *Molecular phylogenetics and evolution*. 2006; 39(2):512–528. doi: [10.1016/j.ympev.2005.12.003](#) PMID: [16439163](#)
82. Schönswetter P, Popp M, Brochmann C. Central Asian origin of and strong genetic differentiation among populations of the rare and disjunct *Carex atrofusca* (Cyperaceae) in the Alps. *Journal of Biogeography*. 2006; 33(5):948–956. doi: [10.1111/j.1365-2699.2006.01462.x](#)
83. Schönswetter P, Popp M, Brochmann C. Rare arctic-alpine plants of the European Alps have different immigration histories: the snow bed species *Minuartia biflora* and *Ranunculus pygmaeus*. *Molecular Ecology*. 2006; 15(3):709–720. doi: [10.1111/j.1365-294X.2006.02821.x](#) PMID: [16499696](#)
84. Shaw J, Small RL. Chloroplast DNA phylogeny and phylogeography of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*, Rosaceae). *American Journal of Botany*. 2005; 92(12):2011–2030. doi: [10.3732/ajb.92.12.2011](#) PMID: [21646120](#)
85. Anderberg AA, Rydin C, Källersjö M. Phylogenetic relationships in the order Ericales s.l.: analyses of molecular data from five genes from the plastid and mitochondrial genomes. *American Journal of Botany*. 2002; 89(4):677–687. doi: [10.3732/ajb.89.4.677](#) PMID: [21665668](#)
86. Goremykin VV, Holland B, Hirsch-Ernst KI, Hellwig FH. Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. *Molecular biology and evolution*. 2005; 22(9):1813–1822. doi: [10.1093/molbev/msi173](#) PMID: [15930156](#)
87. Jansen RK, Cai Z, Raubeson LA, Daniell H, Leebens-Mack J, Müller KF, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences*. 2007; 104(49):19369–19374. doi: [10.1073/pnas.0709121104](#)