



Comparative analysis of the complete plastid genomes of *Mangifera* species and gene transfer between plastid and mitochondrial genomes

Yingfeng Niu¹, Chengwen Gao² and Jin Liu¹

¹Yunnan Institute of Tropical Crops, Xishuangbanna, China

²The Affiliated Hospital of Qingdao University, Qingdao, China

ABSTRACT

Mango is an important commercial fruit crop belonging to the genus *Mangifera*. In this study, we reported and compared four newly sequenced plastid genomes of the genus *Mangifera*, which showed high similarities in overall size (157,780–157,853 bp), genome structure, gene order, and gene content. Three mutation hotspots (*trnG-psbZ*, *psbD-trnT*, and *ycf4-cemA*) were identified as candidate DNA barcodes for *Mangifera*. These three DNA barcode candidate sequences have high species identification ability. We also identified 12 large fragments that were transferred from the plastid genome to the mitochondrial genome, and found that the similarity was more than 99%. The total size of the transferred fragment was 35,652 bp, accounting for 22.6% of the plastid genome. Fifteen intact chloroplast genes, four tRNAs and numerous partial genes and intergenic spacer regions were identified. There are many of these genes transferred from mitochondria to the chloroplast in other species genomes. Phylogenetic analysis based on whole plastid genome data provided a high support value, and the interspecies relationships within *Mangifera* were resolved well.

Submitted 7 August 2020
Accepted 22 December 2020
Published 10 February 2021

Corresponding authors
Chengwen Gao,
gaochengwen6@126.com
Jin Liu, liujin06@126.com

Academic editor
Rogerio Sotelo-Mundo

Additional Information and
Declarations can be found on
page 10

DOI 10.7717/peerj.10774

© Copyright
2021 Niu et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Agricultural Science, Genetics, Molecular Biology, Plant Science

Keywords *Mangifera*, Chloroplast genome, DNA barcodes, Gene transfer, Phylogenetic analysis

INTRODUCTION

Mango is a tall, evergreen tree belonging to the genus *Mangifera* of the Anacardiaceae family. It is an important tropical fruit (Iquebal et al., 2017; Lora & Hormaza, 2018) that originates in tropical and subtropical regions in Southeast Asia (Dutta et al., 2013; Sherman et al., 2015). Owing to its wide range of cultivation (Bajpai et al., 2016), high nutrient value, pleasing appearance, and unique flavor (Surapaneni et al., 2013), it is widely loved by consumers and has the reputation of being known as the “King of Tropical Fruits” (Khan, Ali & Khan, 2015). Southeast Asian countries have a history of mango cultivation that spans thousands of years (Ravishankar et al., 2013). Mangoes were introduced to Africa, South America, and other continents hundreds of years ago, and several varieties suitable for local cultivation have been developed (Mansour, Mekki & Hussein, 2014; Sennhenn et al., 2014). There are 69 species of mango in the world that are mainly distributed in tropical and subtropical countries including India, Indonesia, the Malay Peninsula, Thailand, and

South China, of which, five species are grown in China, namely *M. indica*, *M. persiciformis*, *M. longipes*, *M. hiemalis*, and *M. sylvatica*; however, the varieties cultivated in production belong to *M. indica*. Phylogenetic analysis of *Mangifera* species has been a hot topic of research (Nishiyama et al., 2006; Sankaran et al., 2018), while the whole chloroplast genome sequences can provide more genetic information and higher species resolution ability than other molecular data. However, the chloroplast genomes of most *Mangifera* plants remain unknown (Azim, Khan & Zhang, 2014).

Chloroplasts are special organelles that are involved in photosynthesis and consist of layers of thylakoids. They have their own DNA and can split. The chloroplast genome is conserved and consists of four parts. Two inverted repeat (IR) regions separate the small copy region (SSC) and large copy region (LSC). Currently, with the rapid development of next-generation sequencing (NGS) technology, the entire chloroplast genome has been widely used for phylogenetic analysis. They can provide a large number of variable sites for phylogenetic analysis (Gitzendanner et al., 2018). Thus, the entire chloroplast genome shows the potential to resolve evolutionary relationships and produce highly resolved phylogenetic and genetic diversity, particularly in some complex taxa or at low taxonomic levels, which have unresolved relationships (Hu et al., 2016; Huang et al., 2020; Xu et al., 2019).

In this study, the chloroplast genomes of four *Mangifera* species were sequenced and compared with *M. Indica* and 21 Sapindales plastids. The objectives of this study were as follows: (1) to comparatively analyze the chloroplast genome structure of five species of *Mangifera*; (2) to identify highly divergent regions of the chloroplast genomes of *Mangifera*; (3) to determine the insertion of chloroplast genes into mitochondria; (4) to explore the evolutionary relationship between the genus, *Mangifera*, and Sapindales. Overall, this study would be helpful to further understand plastid evolution and phylogeny of the genus, *Mangifera*.

MATERIALS AND METHODS

Plant material, DNA extraction, and sequencing

Fresh leaves of four *Mangifera* species (*M. hiemalis*, *M. persiciformis*, *M. longipes*, and *M. sylvatica*) were collected from Xishuangbanna Tropical Flowers and Plants Garden, South Yunnan, China, and frozen in liquid nitrogen. Total genomic DNA was extracted from all samples according to CTAB method (Li et al., 2013). DNA quality was detected using 1% agarose gel electrophoresis and samples were stored at -80°C until further use.

About 5–10 μg of total DNA were extracted from each of the *Mangifera* samples to construct a shotgun library with an average insertion size of 300 bp. Paired-end libraries were constructed with NEBNext[®] DNA Library Prep Master Mix Set for Illumina according to the manufacturer's recommendation. Illumina HiSeq 2500 system (Illumina, San Diego, CA, USA) was used to sequence DNA samples in the paired-end sequencing mode by Novogene Bioinformatics Technology Co. Ltd (Beijing, China), generating approximately 8.0 Gb of raw data per sample. The plastome depth of coverage was more than 2000 \times .

Chloroplast genome assembly and annotation

The Trimmomatic v0.38 was used to filter raw sequencing data (Bolger, Lohse & Usadel, 2014), and the obtained clean data were de novo assembled using SPAdes v3.61 under different K-mer parameters (Bankevich et al., 2012). The scaffolds that were positively associated with chloroplasts were arranged on the reference chloroplast genome of *M. indica* (NC_035239). Paired-end reads were remapped to consensus assembly and multiple iterations were performed to fill in the gaps in the final consensus sequence using Geneious software v2020.0.4 (Kearse et al., 2012).

Chloroplast genome annotation was performed using GeSeq (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>) to predict genes encoding proteins, transfer RNA (tRNA), and ribosomal RNA (rRNA), and was adjusted manually as needed (Tillich et al., 2017). We also manually examined the IR junctions of all *Mangifera* species. A circular diagram of the chloroplast genomes of *Mangifera* was subsequently drawn using OGDRAW v1.3.1 (Greiner, Lehwark & Bock, 2019).

Genome comparative analysis and divergent hotspot identification

MAFFT v7.221 was used to align the chloroplast genome sequences of five *Mangifera* plants (Katoh & Standley, 2013). Next, DnaSP v6.12 was used to perform a sliding window analysis with the step size of 200 bp and window length of 600 bp, to detect the rapidly evolving molecular markers for performing phylogenetic analysis (Librado & Rozas, 2009).

Identification of chloroplast gene insertion in mitochondria

First, we removed the BLAST hits of genes transferred between chloroplast and mitochondrial genomes by mapping the mitochondrial genome of *M. indica* (GenBank: CM021857) to the plastid genomes. Circos v0.69-9 (Krzywinski et al., 2009) software was used to map the mitochondrial and chloroplast genomes of the *Mangifera* species as well as gene-transfer fragments.

Phylogenetic analysis

Phylogenetic analyses were performed for five *Mangifera* (4 species sequenced here) and 21 Sapindales species, using *Arabidopsis thaliana* as outgroups. MAFFT 7.221 (Katoh & Standley, 2013) was used to align the chloroplast genome sequences of Sapindales species. We used the following three methods to perform phylogenetic analyses of *Mangifera* species: Bayesian Inference (BI) with a GTR + I + G model using MrBayes v3.2 (Ronquist et al., 2012), the Markov chain Monte Carlo (MCMC) algorithm was run for 1 million generations and sampled every 100 generations. Maximum Likelihood (ML) using MEGA v7.0 with 1000 bootstrap replicates (Kumar, Stecher & Tamura, 2016), and Maximum Parsimony (MP) with a heuristic search in PAUP v4.0 with 1,000 random taxon stepwise addition sequences (Swofford, 1993). A 50% majority-rule consensus phylogeny was constructed using 1,000 bootstrap replications.

RESULTS AND DISCUSSION

Basic characteristics of the *Mangifera* chloroplast genomes

Raw data (approximately from 7.1×10^9 to 8.3×10^9 bp) were obtained from *M. hiemalis* (MN917208), *M. persiciformis* (MN917209), *M. longipes* (MN917210), and *M. sylvatica* (MN917211). The four newly sequenced *Mangifera* chloroplast genomes have been presented to the GenBank database.

Characteristics of four newly sequenced and one reported *Mangifera* chloroplast genomes were investigated. *Mangifera* chloroplast genome sequence sizes were 157,780~157,853 bp (Fig. 1), with the largest and smallest being those of *M. longipes* and *M. indica*, respectively. *Mangifera* chloroplast genomes are characterized by a typical four-part structure, two IR copies (26,354–26,379 bp) separating the LSC (86,673–86,726 bp) and SSC (18,347–18,369 bp) regions. In addition, the GC content of *Mangifera* genomes was similar, ranging from 37.88–37.89%. Five *Mangifera* chloroplast genomes contained 113 predicted functional genes, including 79 protein-coding genes, four ribosomal RNA (rRNA) genes, and 30 transfer RNA (tRNA) genes (Tables 1 and 2). Furthermore, 15 functional genes, including four protein-coding genes, four ribosomal RNA genes, and seven transfer RNA gene replicate in the IR regions of the chloroplast genome. The number, type, and order of genes were found to be very similar among the five *Mangifera* chloroplast genomes (Jo et al., 2017; Rabah et al., 2017; Zhang et al., 2020). The whole chloroplast genome sequences of four *Mangifera* species were submitted to GenBank with the accession numbers of MN917208 to MN917211.

The IR/SC connected regions were found nearly identical relative positions in the five *Mangifera* chloroplast genomes (Fig. 2). All LSC-IRb connections were found to be located within the *rps19* gene, resulting in a partial expansion of the IRb region to the *rps19* gene (80–104 bp). The IRb-SSC boundary was located in the *ndhF* gene, while the SSC-IRA boundary in the five chloroplast genomes was located in the *ycf1* gene.

Comparative *Mangifera* chloroplast genomes and Divergence Hotspot Regions

Using the comparative sequence analysis of the five species of *Mangifera*, we found that the plastid genome was quite conservative in the five taxa, although there were a few regions with variations. In general, sequences are conserved in the coding region, and most of the detected variations are in the non-coding region. The results agree with previous reports that non-coding regions showed greater divergence than coding regions, this is possibly caused by coding regions affected by stronger selective pressure (Li et al., 2018). Consistent with similar studies involving other plants, the IR regions appear to be more conservative than the LSC and SSC regions (Fig. 1) (Liang et al., 2019; Song et al., 2019). A search for nucleotide substitutions identified 638 variable sites (0.40%) in the five chloroplast genomes, including 489 parsimony-informative sites (0.31%), this number is smaller than other genus species (Gao et al., 2020; Nguyen et al., 2020).

To identify hotspots of sequence divergence, the nucleotide diversity (P_i) values within the 600 bp window of the *Mangifera* chloroplast genomes were calculated (Fig. 3). We found that P_i values varied from 0–0.033, and the three hypervariable regions ($P_i > 0.02$)

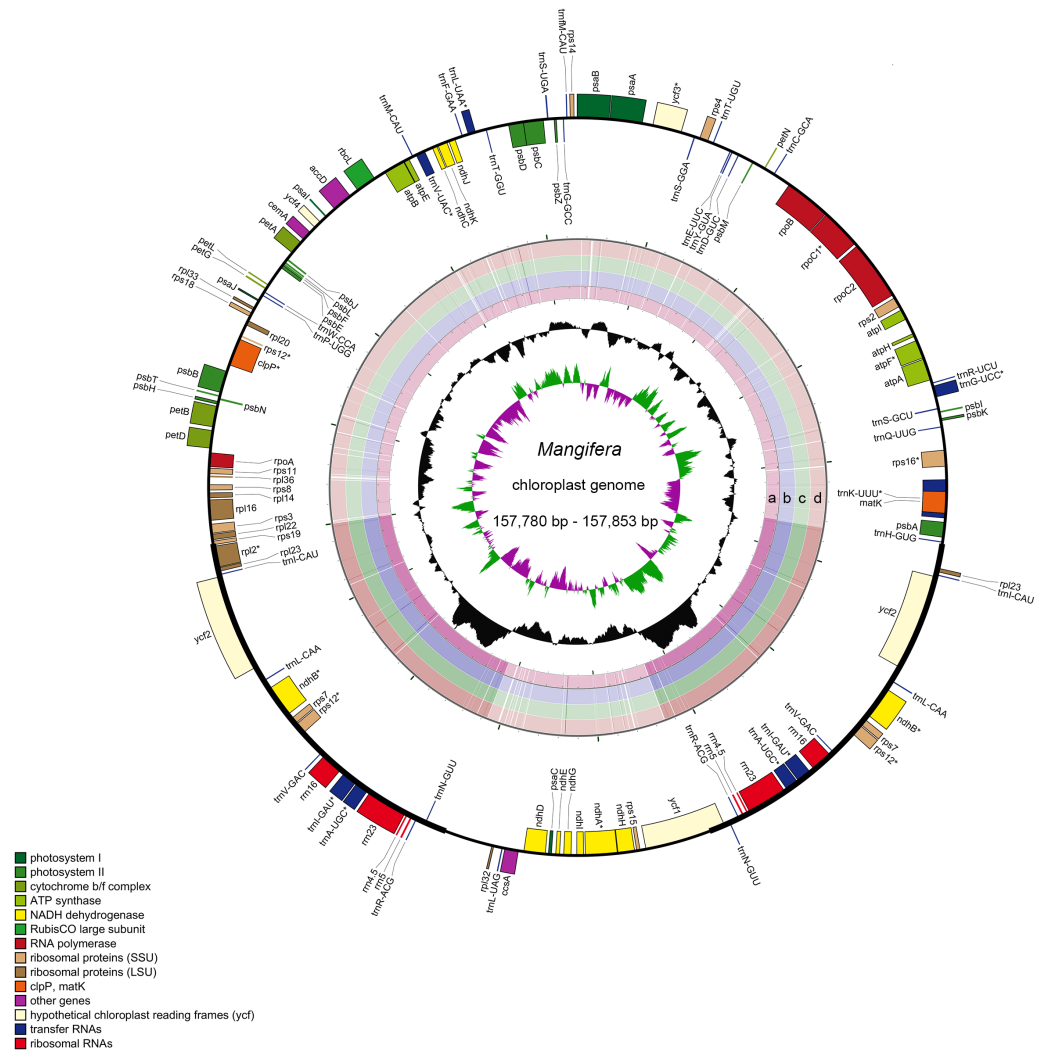


Figure 1 Sequence diagram of *Mangifera* chloroplast genomes. Gene map of *Mangifera* chloroplast genomes, sequence alignment of *Mangifera* species chloroplast genome ((A) *M. Sylvatica*, (B) *M. hiemalis*, (C) *M. longipes*, (D) *M. persiciformis* with reference to *M. indica*), GC content, and GC skew from the outside to inside.

Full-size DOI: 10.7717/peerj.10774/fig-1

of the five *Mangifera* chloroplast genomes were *trnG-psbZ*, *psbD-trnT*, and *ycf4-cemA*. The *trnG-psbZ* region exhibited the highest variability (7.44%).

Here, we found an increase in the number of variable sites in the following three specific regions based on the results of pairwise plastid genomic alignment and SNP analysis: *trnG-psbZ*, *psbD-trnT*, and *ycf4-cemA*. Thus, *Mangifera* species may be detected using these regions as novel candidate fragments. Figure S1 presents the graphical representation of these results using the ML method. These three DNA barcode candidate sequences have high species identification ability. However, further experiments are required to support this *Mangifera* plastid sequence data.

Table 1 Summary of chloroplast genome features of five *Mangifera* species.

Genome feature	<i>M. indica</i>	<i>M. longipes</i>	<i>M. persiciformis</i>	<i>M. hiemalis</i>	<i>M. sylvatica</i>
Total size (bp)	157,780	157,853	157,799	157,796	157,824
LSC Length (bp)	86,673	86,726	86,724	86,718	86,719
SSC Length (bp)	18,349	18,369	18,367	18,368	18,347
IR Length (bp)	26,379	26,379	26,354	26,355	26,379
Total Genes	113	113	113	113	113
Protein coding Genes	79	79	79	79	79
Structure RNAs	34	34	34	34	34
GC Content (%)	37.89%	37.88%	37.88%	37.89%	37.89%
GenBank Accessions	NC035239	MN917210	MN917209	MN917208	MN917211

Table 2 Genes contained in *Mangifera* chloroplast genome.

Category	Group of genes	Name of genes
Self replication	Ribosomal RNA genes	<i>rrn4.5, rrn5, rrn16, rrn23</i>
	Small subunit of ribosome	<i>rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps16, rps18, rps19</i>
	Transfer RNA genes	<i>trnR-UCU, trnS-GCU, trnA-UGC, trnC-GCA, trnF-GAA, trnG-GCC, trnG-UCC, trnD-GUC, trnE-UUC, trnH-GUG, trnN-GUU, trnP-UGG, trnQ-UUG, trnR-ACG, trnI-GAU, trnY-GUA, trnK-UUU, trnL-CAA, trnL-UAA, trnI-CAU, trnV-GAC, trnV-UAC, trnW-CCA, trnL-UAG, trnFM-CAU, trnM-CAU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU</i>
	DNA dependent RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>
Photosynthesis	Large subunit of ribosome	<i>rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>
	Subunits of photosystem I	<i>psaA, psaB, psaC, psaI, psaJ, ycf3, ycf4</i>
	Subunits of NADH-dehydrogenase	<i>ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Subunits of cytochrome complex	<i>petA, petB, petD, petG, petL, petN</i>
Other genes	Protease	<i>clpP</i>
	Maturase	<i>matK</i>
	Acetyl-CoA-carboxylase c-type Cytochrome synthesis gene	<i>ccsA</i>
	Large subunit of rubisco	<i>rbcl</i>
	Envelop membrane protein	<i>cemA</i>
	Subunit of Acetyl-CoA-carboxylase	<i>accD</i>
Hypothetical chloroplast	<i>ycf1, ycf2, ycf15</i>	

Characterization of gene transfer of *Mangifera* chloroplast genome to mitochondrial genome

The mitochondrial genome of *M. indica* was obtained from GenBank and was 87,1458 bp in size, approximately 5.5 times that of the chloroplast genome consisting of 94 functional genes. We identified 12 large chloroplast genome fragments in the mitochondrial genome,

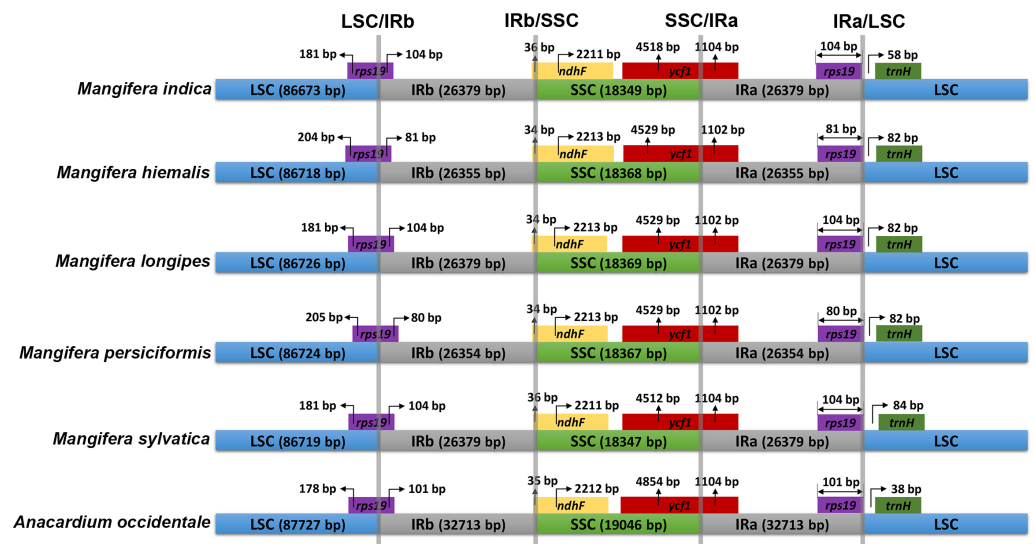


Figure 2 Comparison of inverted repeat (IR) boundary among *Mangifera* species, where genes and gene fragments across IRa/b junctions are represented in color boxes above the horizontal line. Genes and IR segments are not mapped to scale.

Full-size [DOI: 10.7717/peerj.10774/fig-2](https://doi.org/10.7717/peerj.10774/fig-2)

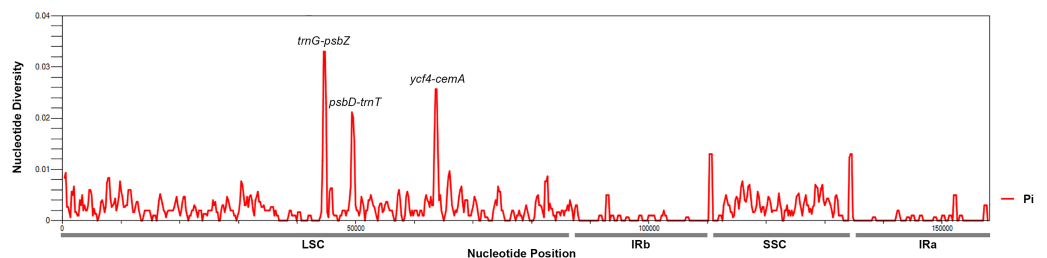


Figure 3 *Mangifera* Chloroplast genomes sliding window analysis (window length: 600 bp; step size: 200 bp). X-axis, Position of a window; Y-axis, Genetic diversity per window.

Full-size [DOI: 10.7717/peerj.10774/fig-3](https://doi.org/10.7717/peerj.10774/fig-3)

including genes and intergenomic regions. These fragments ranged from 1,522–5,400 bp and the sequences were over 99% consistent. The total length of these fragments was 35,652 bp, accounting for 22.6% of the chloroplast genome (Fig. 4 and Table S1). Fifteen intact chloroplast genes (*rps19*, *rpl2*, *rpl23*, *petN*, *rbcL*, *accD*, *psbJ*, *psbL*, *psbF*, *psbE*, *petL*, *petG*, *psaA*, *atpA*, *cemA*), four tRNAs (*trnI*-CAU, *trnC*-GCA, *trnW*-CCA, *trnP*-UGG) and numerous partial genes and intergenic spacer regions were identified. There are many of these genes transferred from mitochondria to the chloroplast in other species genomes, such as *rps12*, *rpl23*, *rbcL*, *petL*, *petG*, *trnW*-CCA and *trnP*-UGG (Gao et al., 2020; Gui et al., 2016).

Intracellular gene transfer exists between different genomes, including those of the chloroplasts, mitochondria, and nuclei (Nguyen et al., 2020; Timmis et al., 2004). Research shows that the frequency of nuclear DNA transfer from organelles in angiosperms is very high (Hazkani-Covo, Zeller & Martin, 2010; Park et al., 2014; Smith, 2011). Gene transfer

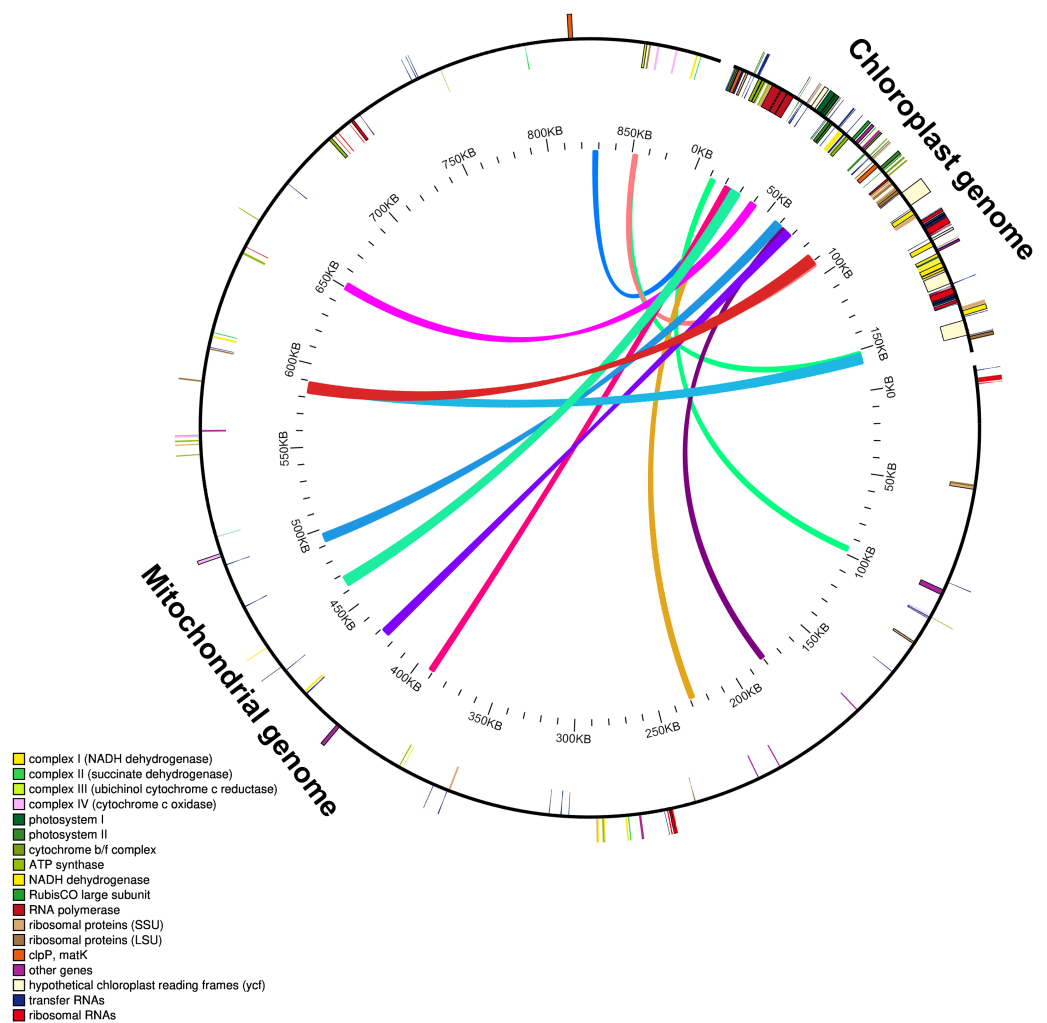


Figure 4 Schematic diagram of gene transfer between chloroplast and mitochondria in *Mangifera* species. Colored lines within the circle show where the chloroplast genome is inserted into the mitochondrial genome. Genes within a circle are transcribed clockwise, while those outside the circle are transcribed counterclockwise.

Full-size DOI: 10.7717/peerj.10774/fig-4

from chloroplast to mitochondrial genomes is a common phenomenon during long-term evolution (*Gui et al., 2016; Nguyen et al., 2020*). Due to high sequence identity between the transferred chloroplast genome fragments in the mitochondrial and original chloroplast genomes, gene transfer can lead to assembly errors in these genomes.

Phylogenetic relationship of chloroplast genomes

In this study, the chloroplast genome was used for infer the phylogenetic location of *Mangifera* in Sapindales (*Fig. 5*) and performed a phylogenetic analysis of the chloroplast genome using three different methods, namely, ML, MP, and BI. BI and ML analyses revealed almost the same topology, and most branches had very high support (*Fig. S2*). However, MP trees differed slightly from BI and ML trees in some taxa (*Fig. S3*). Despite

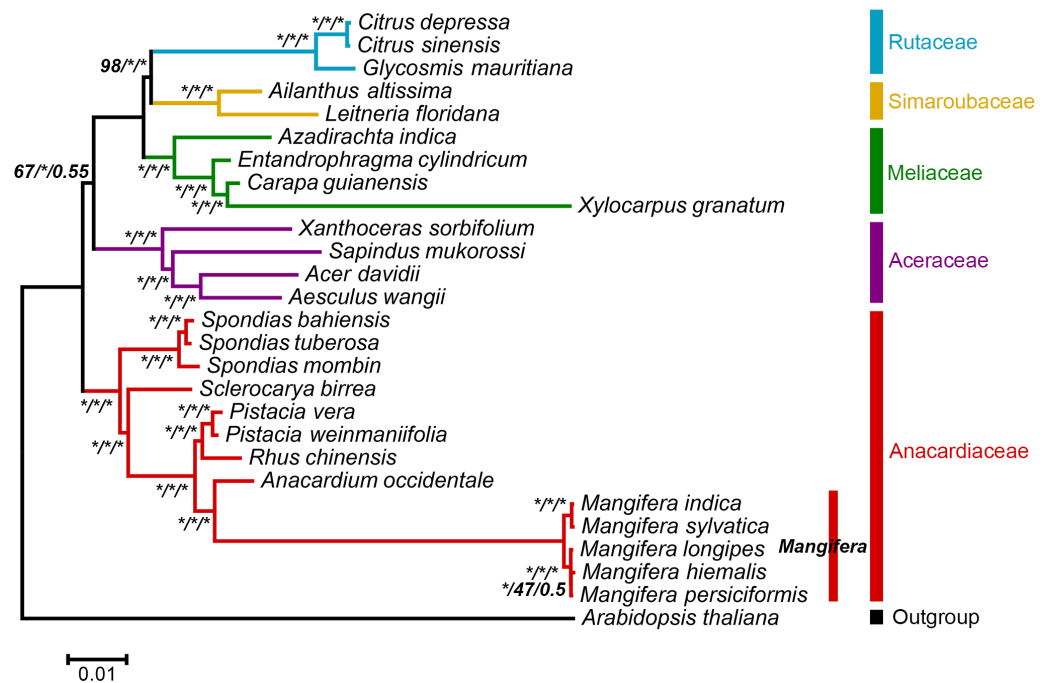


Figure 5 ML phylogenetic tree of five *Mangifera* species with 21 related species in the Sapindales based on whole chloroplast genome sequence. Numbers related to the branches are ML bootstrap value, MP bootstrap value, and Bayesian posterior probability, respectively. Asterisk denotes 100% bootstrap support or 1.0 posterior probability.

Full-size [DOI: 10.7717/peerj.10774/fig-5](https://doi.org/10.7717/peerj.10774/fig-5)

differences between these three approaches, the relationships between most groups were well resolved and highly supported, suggesting that the use of chloroplast genome data does significantly improve the resolution of phylogenetic analysis. Previous studies have revealed the genetic relationship of *Mangifera* through morphological, nuclear, amplified fragment length polymorphism, ribosomal internal transcribed spacer (ITS), and partial chloroplast gene analysis (Eiadthong et al., 2000; Nishiyama et al., 2006; Sankaran et al., 2018; Yonemori et al., 2002). The whole chloroplast genome sequence-based phylogenetic tree was built to explore the evolutionary similarities/differences between *Mangifera* species and between genera in the Sapindales. Phylogenetic analysis based on complete genome sequences, rather than a few genes, has been carried out in a large number of higher plant species, significantly improving the resolution of phylogenetic analysis (Zhai et al., 2019).

CONCLUSIONS

In this study, the chloroplast genomes of four *Mangifera* species were sequenced and compared. It was found that the size, structure, and gene content of the *Mangifera* chloroplast genomes were conserved. Comparative analysis showed a low degree of sequence variation. We identified 13 large fragments that were transferred from the chloroplast genome to the mitochondrial genome. In addition, we identified three mutation hotspots as DNA barcodes for the identification of *Mangifera* species. These

complete chloroplast genome sequences and highly variable markers provide sufficient genetic information for the phylogenetic reconstruction and species identification of the genus *Mangifera*.

ACKNOWLEDGEMENTS

We are grateful to thank Zhangguang Ni for the collection of experiment material.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Youth Talent Growth Fund of YITC (QNCZ2020-3), Technology Innovation Talents Project of Yunnan Province (2018HB086), and Sci-tech Innovation System Construction for Tropical Crops Grant of Yunnan Province (No. RF2021). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

The Youth Talent Growth Fund of YITC: QNCZ2020-3.

Technology Innovation Talents Project of Yunnan Province: 2018HB086.

Sci-tech Innovation System Construction for Tropical Crops Grant of Yunnan Province: No. RF2021.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Yingfeng Niu conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Chengwen Gao conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Jin Liu conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

Data are available at GenBank: [MN917208–MN917211](#).

Data Availability

The following information was supplied regarding data availability:

Data are available at NCBI: [PRJNA655379](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.10774#supplemental-information>.

REFERENCES

- Azim MK, Khan IA, Zhang Y. 2014.** Characterization of mango (*Mangifera indica* L.) transcriptome and chloroplast genome. *Plant Molecular Biology* **85**:193–208 DOI [10.1007/s11103-014-0179-8](https://doi.org/10.1007/s11103-014-0179-8).
- Bajpai A, Muthukumar M, Ahmad I, Ravishankar KV, Parthasarthy VA, Sthapit B, Rao R, Verma SP, Rajan S. 2016.** Molecular and morphological diversity in locally grown non-commercial (heirloom) mango varieties of North India. *Journal of Environmental Biology* **37**:221–228.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012.** SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**:455–477 DOI [10.1089/cmb.2012.0021](https://doi.org/10.1089/cmb.2012.0021).
- Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120 DOI [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170).
- Dutta SK, Srivastav M, Rymbai H, Chaudhary R, Singh AK, Dubey AK, Lal K. 2013.** Pollen-pistil interaction studies in mango (*Mangifera indica* L.) cultivars. *Scientia Horticulturae* **160**:213–221 DOI [10.1016/j.scienta.2013.05.012](https://doi.org/10.1016/j.scienta.2013.05.012).
- Eiadthong W, Yonemori K, Kanzaki S, Sugiura A, Utsunomiya N, Subhadrabandhu S. 2000.** Amplified fragment length polymorphism analysis for studying genetic relationships among *Mangifera* species in Thailand. *Journal of the American Society for Horticultural Science* **125**:160–164 DOI [10.21273/jashs.125.2.160](https://doi.org/10.21273/jashs.125.2.160).
- Gao CW, Wu CH, Zhang Q, Zhao X, Wu MX, Chen RR, Zhao YL, Li ZQ. 2020.** Characterization of chloroplast genomes from two *Salvia* medicinal plants and gene transfer among their mitochondrial and chloroplast genomes. *Frontiers in Genetics* **11**:574962 DOI [10.3389/fgene.2020.574962](https://doi.org/10.3389/fgene.2020.574962).
- Gitzendanner MA, Soltis PS, Wong GKS, Ruhfel BR, Soltis DE. 2018.** Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *American Journal of Botany* **105**:291–301 DOI [10.1002/ajb2.1048](https://doi.org/10.1002/ajb2.1048).
- Greiner S, Lehwark P, Bock R. 2019.** OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* **47**:W59–W64 DOI [10.1093/nar/gkz238](https://doi.org/10.1093/nar/gkz238).
- Gui ST, Wu ZH, Zhang HY, Zheng YZ, Zhu ZX, Liang DQ, Ding Y. 2016.** The mitochondrial genome map of *Nelumbo nucifera* reveals ancient evolutionary features. *Scientific Reports* **6**:30158 DOI [10.1038/srep30158](https://doi.org/10.1038/srep30158).
- Hazkani-Covo E, Zeller RM, Martin W. 2010.** Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLOS Genetics* **6**:e1000834 DOI [10.1371/journal.pgen.1000834](https://doi.org/10.1371/journal.pgen.1000834).

- Hu H, Hu QJ, Al-Shehbaz IA, Luo X, Zeng TT, Guo XY, Liu JQ. 2016. Species delimitation and interspecific relationships of the genus *Orychophragmus* (Brassicaceae) inferred from whole chloroplast genomes. *Frontiers in Plant Science* 7:1826 DOI 10.3389/fpls.2016.01826.
- Huang J, Yu Y, Liu YM, Xie DF, He XJ, Zhou SD. 2020. Comparative chloroplast genomics of fritillaria (Liliaceae), inferences for phylogenetic relationships between *Fritillaria* and liliaceae and plastome evolution. *Plants-Basel* 9:133 DOI 10.3390/plants9020133.
- Iqbal MA, Jaiswal S, Mahato AK, Jayaswal PK, Angadi UB, Kumar N, Sharma N, Singh AK, Srivastava M, Prakash J, Singh SK, Khan K, Mishra RK, Rajan S, Bajpai A, Sandhya BS, Nischita P, Ravishankar KV, Dinesh MR, Rai A, Kumar D, Sharma TR, Singh NK. 2017. MiSNPdb: a web-based genomic resources of tropical ecology fruit mango (*Mangifera indica* L.) for phylogeography and varietal differentiation. *Scientific Reports* 7:14968 DOI 10.1038/s41598-017-14998-2.
- Jo S, Kim HW, Kim YK, Sohn JY, Cheon SH, Kim KJ. 2017. The complete plastome sequences of *Mangifera indica* L. (Anacardiaceae). *Mitochondrial DNA Part B-Resources* 2:698–700 DOI 10.1080/23802359.2017.1390407.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772–780 DOI 10.1093/molbev/mst010.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649 DOI 10.1093/bioinformatics/bts199.
- Khan AS, Ali S, Khan IA. 2015. Morphological and molecular characterization and evaluation of mango germplasm: an overview. *Scientia Horticulturae* 194:353–366 DOI 10.1016/j.scienta.2015.08.031.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Research* 19:1639–1645 DOI 10.1101/gr.092759.109.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870–1874 DOI 10.1093/molbev/msw054.
- Li J, Wang S, Jing Y, Ling W, Zhou S. 2013. A modified CTAB protocol for plant DNA extraction. *Chinese Bulletin of Botany* 48:72–78 DOI 10.3724/SP.J.1259.2013.00072.
- Li YT, Zhang J, Li LF, Gao LJ, Xu JT, Yang MS. 2018. Structural and comparative analysis of the complete chloroplast genome of *Pyrus hopeiensis* “Wild Plants with a Tiny Population” and three other *Pyrus* species. *International Journal of Molecular Sciences* 19:3262 DOI 10.3390/ijms19103262.
- Liang CL, Wang L, Lei J, Duan BZ, Ma WS, Xiao SM, Qi HJ, Wang Z, Liu YQ, Shen XF, Guo S, Hu HY, Xu J, Chen SL. 2019. A comparative analysis of the

- chloroplast genomes of four *Salvia* medicinal plants. *Engineering* 5:907–915 DOI 10.1016/j.eng.2019.01.017.
- Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452 DOI 10.1093/bioinformatics/btp187.
- Lora J, Hormaza JI. 2018.** Pollen wall development in mango (*Mangifera indica* L., Anacardiaceae). *Plant Reproduction* 31:385–397 DOI 10.1007/s00497-018-0342-5.
- Mansour H, Mekki LE, Hussein MA. 2014.** Assessment of genetic diversity and relationships among Egyptian mango (*Mangifera indica* L.) cultivars grown in Suez Canal and Sinai region using RAPD markers. *Pakistan Journal of Biological Sciences* 17:56–61.
- Nguyen VB, Giang VNL, Waminal NE, Park HS, Kim NH, Jang W, Lee J, Yang TJ. 2020.** Comprehensive comparative analysis of chloroplast genomes from seven *Panax* species and development of an authentication system based on species-unique single nucleotide polymorphism markers. *Journal of Ginseng Research* 44:135–144 DOI 10.1016/j.jgr.2018.06.003.
- Nishiyama K, Choi YA, Honsho C, Eiadthong W, Yonemori K. 2006.** Application of genomic in situ hybridization for phylogenetic study between *Mangifera indica* L. and eight wild species of *Mangifera*. *Scientia Horticulturae* 110:114–117 DOI 10.1016/j.scienta.2006.06.005.
- Park S, Ruhlman TA, Sabir JSM, Mutwakil MHZ, Baeshen MN, Sabir MJ, Baeshen NA, Jansen RK. 2014.** Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. *Bmc Genomics* 15:405 DOI 10.1186/1471-2164-15-405.
- Rabah SO, Lee C, Hajrah NH, Makki RM, Alharby HF, Alhebshi AM, Sabir JSM, Jansen RK, Ruhlman TA. 2017.** Plastome sequencing of ten nonmodel crop species uncovers a large insertion of mitochondrial DNA in cashew. *Plant Genome* 10:29293812 DOI 10.3835/plantgenome2017.03.0020.
- Ravishankar KV, Dinesh MR, Mani BH, Padmakar B, Vasugi C. 2013.** Assessment of genetic diversity of mango (*Mangifera indica* L.) cultivars from Indian Peninsula using Sequence Tagged Microsatellite Site (STMS) markers. In: Lu P, ed. *IX international mango symposium*. Leuven 1: Int Soc Horticultural Science, 269–275.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012.** MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61:539–542 DOI 10.1093/sysbio/sys029.
- Sankaran M, Dinesh MR, Chaitra N, Ravishankar KV. 2018.** Morphological, cytological, palynological and molecular characterization of certain *Mangifera* species. *Current Science* 115:1379–1386 DOI 10.18520/cs/v115/i7/1379-1386.
- Sennhenn A, Prinz K, Gebauer J, Whitbread A, Jamnadass R, Kehlenbeck K. 2014.** Identification of mango (*Mangifera indica* L.) landraces from Eastern and Central Kenya using a morphological and molecular approach. *Genetic Resources and Crop Evolution* 61:7–22 DOI 10.1007/s10722-013-0012-2.

- Sherman A, Rubinstein M, Eshed R, Benita M, Ish-Shalom M, Sharabi-Schwager M, Rozen A, Saada D, Cohen Y, Ophir R. 2015.** Mango (*Mangifera indica* L.) germplasm diversity based on single nucleotide polymorphisms derived from the transcriptome. *Bmc Plant Biology* **15**:277 DOI [10.1186/s12870-015-0663-6](https://doi.org/10.1186/s12870-015-0663-6).
- Smith DR. 2011.** Extending the limited transfer window hypothesis to inter-organelle DNA migration. *Genome Biology and Evolution* **3**:743–748 DOI [10.1093/gbe/evr068](https://doi.org/10.1093/gbe/evr068).
- Song Y, Zhang YJ, Xu J, Li WM, Li MF. 2019.** Characterization of the complete chloroplast genome sequence of *Dalbergia* species and its phylogenetic implications. *Scientific Reports* **9**:20401 DOI [10.1038/s41598-019-56727-x](https://doi.org/10.1038/s41598-019-56727-x).
- Surapaneni M, Vemireddy LR, Begum H, Reddy BP, Neetasri C, Nagaraju J, Anwar SY, Siddiq EA. 2013.** Population structure and genetic analysis of different utility types of mango (*Mangifera indica* L.) germplasm of Andhra Pradesh state of India using microsatellite markers. *Plant Systematics and Evolution* **299**:1215–1229 DOI [10.1007/s00606-013-0790-1](https://doi.org/10.1007/s00606-013-0790-1).
- Swofford DL. 1993.** PAUP – a Computer-Program for phylogenetic inference using maximum parsimony. *Journal of General Physiology* **102**:A9–A9.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017.** GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* **45**:W6–W11 DOI [10.1093/nar/gkx391](https://doi.org/10.1093/nar/gkx391).
- Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004.** Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* **5**:123–U116 DOI [10.1038/nrg1271](https://doi.org/10.1038/nrg1271).
- Xu WQ, Losh J, Chen C, Li P, Wang RH, Zhao YP, Qiu YX, Fu CX. 2019.** Comparative genomics of figworts (*Scrophularia*, Scrophulariaceae), with implications for the evolution of Scrophularia and Lamiales. *Journal of Systematics and Evolution* **57**:55–65 DOI [10.1111/jse.12421](https://doi.org/10.1111/jse.12421).
- Yonemori K, Honsho C, Kanzaki S, Eiadthong W, Sugiura A. 2002.** Phylogenetic relationships of *Mangifera* species revealed by ITS sequences of nuclear ribosomal DNA and a possibility of their hybrid origin. *Plant Systematics and Evolution* **231**:59–75 DOI [10.1007/s006060200011](https://doi.org/10.1007/s006060200011).
- Zhai W, Duan XS, Zhang R, Guo CC, Li L, Xu GX, Shan HY, Kong HZ, Ren Y. 2019.** Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. *Molecular Phylogenetics and Evolution* **135**:12–21 DOI [10.1016/j.ympev.2019.02.024](https://doi.org/10.1016/j.ympev.2019.02.024).
- Zhang Y, Ou KW, Huang GD, Lu YF, Yang GQ, Pang XH. 2020.** The complete chloroplast genome sequence of *Mangifera sylvatica* Roxb. (Anacardiaceae) and its phylogenetic analysis. *Mitochondrial DNA Part B-Resources* **5**:738–739 DOI [10.1080/23802359.2020.1715286](https://doi.org/10.1080/23802359.2020.1715286).