

RESEARCH ARTICLE

The First Chloroplast Genome Sequence of *Boswellia sacra*, a Resin-Producing Plant in Oman

Abdul Latif Khan¹, Ahmed Al-Harrasi^{1*}, Sajjad Asaf², Chang Eon Park², Gun-Seok Park², Abdur Rahim Khan², In-Jung Lee², Ahmed Al-Rawahi¹, Jae-Ho Shin^{2*}

1 UoN Chair of Oman's Medicinal Plants & Marine Natural Products, University of Nizwa, Nizwa, Oman, **2** School of Applied Biosciences, Kyungpook National University, Daegu, Republic of Korea

* aharrasi@unizwa.edu.om (AAH); jhshin@knu.ac.kr (JHS)



Abstract

Boswellia sacra (Burseraceae), a keystone endemic species, is famous for the production of fragrant oleo-gum resin. However, the genetic make-up especially the genomic information about chloroplast is still unknown. Here, we described for the first time the chloroplast (cp) genome of *B. sacra*. The complete cp sequence revealed a circular genome of 160,543 bp size with 37.61% GC content. The cp genome is a typical quadripartite chloroplast structure with inverted repeats (IRs 26,763 bp) separated by small single copy (SSC; 18,962 bp) and large single copy (LSC; 88,055 bp) regions. *De novo* assembly and annotation showed the presence of 114 unique genes with 83 protein-coding regions. The phylogenetic analysis revealed that the *B. sacra* cp genome is closely related to the cp genome of *Azadirachta indica* and *Citrus sinensis*, while most of the syntenic differences were found in the non-coding regions. The pairwise distance among 76 shared genes of *B. sacra* and *A. indica* was highest for *atpA*, *rpl2*, *rps12* and *ycf1*. The cp genome of *B. sacra* reveals a novel genome, which could be used for further studied to understand its diversity, taxonomy and phylogeny.

OPEN ACCESS

Citation: Khan AL, Al-Harrasi A, Asaf S, Park CE, Park G-S, Khan AR, et al. (2017) The First Chloroplast Genome Sequence of *Boswellia sacra*, a Resin-Producing Plant in Oman. PLoS ONE 12 (1): e0169794. doi:10.1371/journal.pone.0169794

Editor: Xiu-Qing Li, Agriculture and Agri-Food Canada, CANADA

Received: July 7, 2016

Accepted: December 21, 2016

Published: January 13, 2017

Copyright: © 2017 Khan 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: All relevant data are within the paper and its Supporting Information files. Chloroplast genome is deposited at GenBank of NCBI with accession no. KT934315.

Funding: This work was supported by the Oman Research Council (EBR/15/007). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

A major distinguishing organelle of plant cells is the chloroplast, which was suggested to have originated from cyanobacteria through endosymbiosis [1–3]. The main role and function of the chloroplast is to perform photosynthesis and generate the building blocks required for plant growth and development [4,5]. Chloroplasts also participate in the biosynthesis of starch, fatty acids, pigments and amino acids [6]. Until recently, approximately 490 complete chloroplast (cp) genomes have been sequenced, and this information is publicly available (<http://www.ncbi.nlm.nih.gov/> genome). Most of these assembled genomes are associated with economically important crop plants [7]. Chloroplast genome analysis and engineering, either alone or in combination with traditional breeding techniques, might provide information for the future development of novel plant sources to counteract various environmental stress tolerance issues and improve the level of human-derived benefits from the target plant [8]. The techniques might also provide information to improve the current understanding of major

biosynthesis pathways and functions. The chloroplast (cp) genome contains important information in plant systematics and is maternally inherited in most angiosperms [9]. Thus, understanding these highly conserved structures could also reveal vital features for designing genetic markers.

Chloroplasts contain circular DNA with nearly 130 genes ranging in size from 72–217 kb [10–11]. Chloroplast DNA (cpDNA) is remarkably conserved in gene content and structure, envisaging important information for genome-wide studies of plant evolution [12]. Recent studies, using advanced genome sequencing techniques, have improved the prospects for resolving phylogenetic homogeneity at taxonomic levels and have enhanced the current understanding of the structural and functional evolution of economically important plants and their traits [12–15]. In the last decade, numerous studies have reported the results of chloroplast genome sequencing for economically and ecologically important tree species, such as *Eucommia ulmoides* [16], *Populus cathayana* [17], *Quercus spinosa* [18], *Acacia ligulata* [19], *Pinus armandii* [20], *Cocos nucifera* [21], *Citrus aurantiifolia* [22], *Musa acuminata* [23], Norway spruce [24], etc. The elucidation of the chloroplast genomes of important tree species has facilitated to understand the evaluation of gene structure and has targeted conservation and propagation strategies [12–13].

To understand the genome structure of chloroplasts, we have investigated the economically and culturally important frankincense-producing tree (*Boswellia sacra*; Burseraceae) [25]. There are approximately twenty *Boswellia* species, and *B. sacra* is an endemic species growing only in Dhofar region of Oman [25–26]. The trees grow in desert-woodlands (Fig 1) with meager amounts of water and nutrient availability [27], whereas the domesticated trees are supplied with water [28]. In response to the incisions by the local people, the tree activates its defense mechanism by producing resin [28]. The crystalline resin is used to create a fragrant smoke in homes and also sold in the market for income [26–27]. In addition to this, the tree has received much attention because of its medicinal uses in the Arabian region [29]. The essential oil and the boswellic acid or its derivatives possesses potent anticancer activities [30–32]. There is very scarce scientific knowledge exists regarding the genetic make-up and physiological functions of this endemic tree. Previously, Coppi et al. [31] has worked on the genetic diversity of *B. sacra* using ITS and ISSR analysis, where detailed genomic and conservation studies were suggested. Therefore, it is high time to further understand this keystone species. Hence, in the current study it was aimed to elucidate the complete chloroplast genome of *B. sacra* and determine the phylogenetic relationship of this tree with related species.



Fig 1. *Boswellia sacra*—habitat and leaf morphology. This tree grows wildly in the Dhofar region of Oman.

doi:10.1371/journal.pone.0169794.g001

Materials and Methods

Ethics statement

The study site was managed by the Museum of the Frankincense Land (Salalah). The samples were collected from the Natural Park of Frankincense Tree, which is part of the Land of Frankincense sites, inscribed in the UNESCO's World Cultural and Natural Heritage List. A permission was obtained from the Museum of the Frankincense Land to use the leaf part of the tree resources for research purposes. The trees used for sampling were treated ethically, and our study did not harm the local environment.

Site description and plant growth conditions

Wadi Dawkah, Dhofar-Oman (17°25'21"N; 54°00'32"E) is a completely arid desert region with small sandstone hills. The annual mean temperature of the sampling area is ~35°C, and the annual rainfall is approximately 40 mm [22]. The temperature in summer reaches to ~48°C. Leaf samples were collected from healthy trees with the least number of incisions/wounds, immediately frozen in liquid nitrogen during the field trip to Wadi Dawkah and subsequently stored at -80°C until further analysis. The plant samples were collected by the first author in July 2015, which were identified by Plant Taxonomist Mohammed Al-Broumi, University of Nizwa. The specimens were deposited at University of Nizwa Herbarium Center with a voucher number (UC29).

Chloroplast DNA extraction, sequencing and assembly

The fresh leaves from *Boswellia sacra* were washed with sterile distilled water to remove sand particles and soil debris. The chloroplast DNA was extracted by using a modified protocol of Shi et al. [33]. To prepare the sequencing library, 300 ng of extracted DNA was sheared to 400-bp by using the BioRuptor UCD-200 TS Sonication System (Diagenode Co., Belgium and Denville, NJ, USA). The fragmented DNA was used for subsequent library preparation using the Ion Xpress Fragment Library Kit (Life Technologies, Carlsbad, CA, USA). The DNA library was diluted to 4 pM, and emulsion PCR was performed using the Ion OneTouch System 2 (Life Technologies, Gaithersburg, MD, USA). The final sequencing library was loaded onto an Ion 316 v2 chip and sequenced on the Ion Torrent PGM for 850 flows using the Ion PGM Sequencing 400 Kit (Life Technologies, Gaithersburg, MD). The resultant sequence was *de novo* assembled and mapped with the reference genome [34]. The gaps identified were filled through another round of cp DNA extraction and sequencing. Following aforementioned procedures of cp DNA extraction, a second library was also prepared and sequenced separately to increase the raw data. The filtered sequence reads were assembled and aligned to the known cp genome sequences of Sapindales. At first, the filtered reads were *de novo* assembled into contigs using Mimicking Intelligent Read Assembly (MIRA 4.0) [35]. Second, the assembled contigs were mapped against the reference *Azadirachta indica* (accession number: NC 023792) through Map Reads to Reference Tool in CLC Genomics Workbench (version 8.5). The order of contigs was resolute according to the reference genomes. The gaps (7 to 150 bp) between the *de novo* contigs were replaced with consensus sequences of raw reads, which were later, re-mapped to the reference genomes. Any additional gaps were either filled manually by mapping tools in CLC Genomics Workbench (version 8.5) or using GapCloser (<http://soap.genomics.org.cn/index.html>).

Chloroplast gene annotation and sequence analyses

Genome annotation was performed by using Dual Organellar GenoMe Annotator (DOGMA) [36], tRNAscan-SE v.1.21 [37] and Basic Local Alignment Search Tool (BLAST) [38] from the National Center for Biotechnology Information (NCBI). The using MAUVE v2.4.0 software

[39] was used to compare the genes and sequences of the *de novo* assembled *B. sacra* cp genome sequence and *Azadirachta indica* cp genome sequence, as a reference. In addition, full alignments with annotations were visualized using the VISTA viewer [40].

Repeat sequence analysis

MISA (<http://pgrc.ipk-gatersleben.de/misa/>) was used to analyze the types and positions of various simple sequence repeats (SSRs) in *B. sacra*. The results of this analysis were also correlated with the reference cp genome of *A. indica* according to the methods of Su et al. [22], where a minimum number of repeats were set at 10, 5, 4, 3, 3, and 3 for di-, tri-, tetra-, penta-, and hexanucleotides. For long repeats, the REPuter program [41] was used to identify the number and location of direct and inverted repeats. Minimum repeat size of 30 bp and sequence identity greater than 90% were used. Tandem repeats in the *B. sacra* cp genome were identified using Tandem Repeats Finder version 4.07 b [42] with the default settings.

Complete cp genomes as well as a separate partition using only 76 shared genes were employed to analyze the average pairwise sequence divergence for *B. sacra* and *Azadirachta indica*. Missing and ambiguous gene annotations were confirmed by comparative sequence analysis after a multiple sequence alignment and gene order comparison [43]. These regions were aligned using MAFFT (version 7.222) [44] with the default parameters. Kimura's two-parameter (K2P) model was selected to calculate pairwise sequence divergences [45]. The *indel* (insertion and deletion) polymorphism of shared genes were calculated using DnaSP 5.10.01 [46]. A custom Python script (<https://www.biostars.org/p/119214/>) based on single-nucleotide polymorphism (SNP) definition (a variation in a single nucleotide that occurs at a specific position in the genome) was employed to call SNPs. In addition, the SNPs in coding regions were classified in two ways: synonymous and nonsynonymous as shown by Guo et al. [47]. The number of synonymous (S) and nonsynonymous (N) substitutions for the *B. sacra* and *Azadirachta indica* was performed on MEGA 7.0 [48].

Phylogenetic analysis

Phylogenetic analysis of the *Boswellia sacra* cp genome was performed using Create Alignment and Create Tree tools in CLC genomics workbench version 8.5 (Gap open cost = 10.0, Gap extension cost = 1.0, Tree construction method = Neighbor Joining, Nucleotide distance measure = Jukes-Cantor) [49]. A total of 32 Malvaceae species were selected as in-groups, whereas *Vitis vinifera* was used as an outgroup. The accession numbers are provided in S1 Table. Complete cp genomes were used, and the bootstrap supports were estimated from 100 re-sampled alignments. Four methods were employed to construct phylogenetic trees from 32 cp genomes, including Bayesian inference (BI) implemented with MrBayes 3.12 [50], maximum parsimony (MP) with PAUP 4.0 [51], and maximum likelihood (ML), maximum parsimony (MP), posterior probabilities (PP) and neighbor-joining (NJ) with MEGA 7 [48].

Results and Discussion

General features of the *Boswellia sacra* chloroplast genome

Genome sequencing and assembly of Ion Torrent PGM sequencing produced 286.8 Mb of data for chloroplast genome of *Boswellia sacra*. A total of 373,248 reads (or 112,936,195 bp) were *de novo* assembled showing a total of 703x cp genome coverage. The chloroplast genome sequence quality was re-confirmed through repeating library preparation and sequencing of the extracted cpDNA from *B. sacra* tree. This result was sufficient to produce a single consensus sequence representing the complete chloroplast genome of *Boswellia sacra*. The *de novo*

assembled genome was referenced with already reported cp genomes of *Sapindales* such as *Citrus sinensis* and *Azadirachta indica*. The gaps between the *de novo* contigs were replaced with consensus sequences of the raw reads and re-mapped to the reference genome. The sequences of the chloroplast genomes were deposited in GenBank (accession numbers: KT934315).

The complete chloroplast genome was 160,543 bp in size, comprising a large single copy (LSC) region of 88,055 bp and a small single copy (SSC) region of 18,962 bp. In addition, a pair of inverted repeats (IRa and IRb) of 26,763 bp each (Fig 2) was also revealed. The

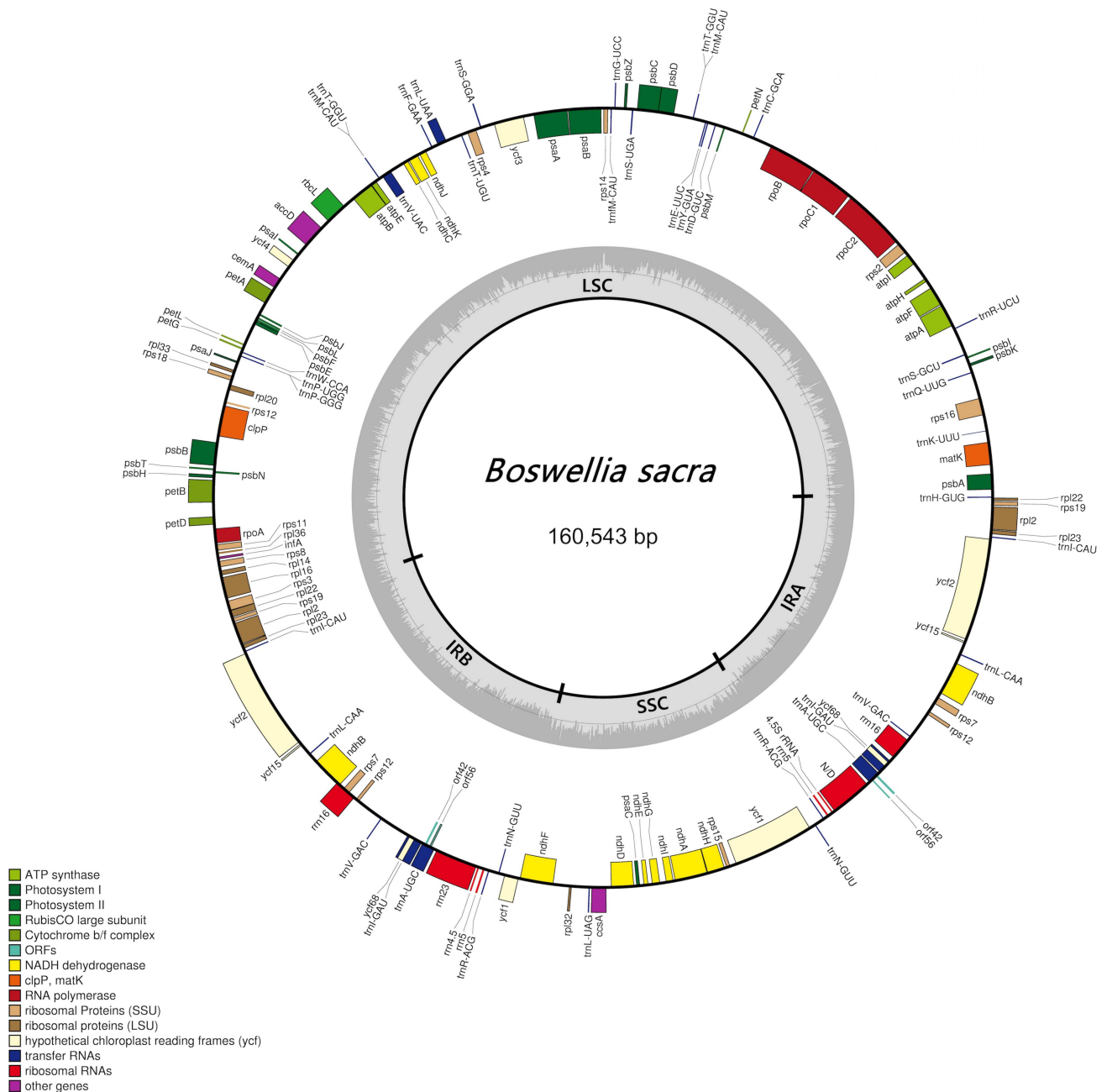


Fig 2. Gene map of the *Boswellia sacra* chloroplast genome. A pair of thick lines in the inside circle represents the inverted repeats (IRa and IRb; 26,763 bp each), separating the large single copy region (LSC; 88,055 bp) from the small single copy region (SSC; 18,962 bp). The genes drawn inside the circle are transcribed clockwise, while those drawn outside the circle are transcribed counterclockwise.

doi:10.1371/journal.pone.0169794.g002

Table 1. Summary of the chloroplast genome characteristics in *Boswellia sacra* and *Azadirachta indica*.

Attribute	<i>Boswellia sacra</i> (KT934315)	<i>Azadirachta indica</i> (NC023792)
Size (bp)	160,543	160,737
Overall GC content (%)	37.6	37.5
LSC size in bp (% total)	88,055 (54.8%)	88,136 (54.8%)
SSC size in bp (% total)	18,962 (11.8%)	18,635 (11.6%)
IR size in bp (% total) ^a	26,763 (16.7%)	26,983 (16.8%)
Protein-coding regions size in bp (% total)	80,289 (50.0%)	79,685 (49.6%)
rRNA and tRNA size in bp (% total)	11,978 (7.5%)	11,863 (7.38%)
Introns size in bp (% total)	17,621 (11.0%)	18,993 (11.8%)
Intergenic spacer size in bp (% total)	57,025 (35.5%)	52,699 (32.8%)
Number of different genes	114	112 ^b
Number of different protein-coding genes	83	78
Number of different rRNA genes	4	4
Number of different tRNA genes	27	30
Number of different genes duplicated by IR	24	19
Number of different genes with introns	16	17

^aEach cp genome contains two copies of inverted repeats (IRs).

^bAccording to the original annotation, *infA*, *ycf15*, *ycf68*, *orf42*, *orf56* not including.

doi:10.1371/journal.pone.0169794.t001

annotation analysis revealed 114 different genes, with 83 protein-coding genes. The number of tRNA and rRNA genes was 27 and 4, respectively (Table 1). Among these, 24 genes were duplicated in the IR regions. The cp genome size of *B. sacra* was almost similar to *Azadirachta indica* (160,737 bp) and *Citrus sinensis* (160,129 bp). However, the size was almost 380 bp higher than closer relative in Order Sapindales i.e. *Citrus aurantifolia* (159,893 bp). In case of other arid-land trees, *B. sacra* size was higher than *Acacia ligulata* (158,724 bp) as shown in the previous studies of cp genomes of these plants [19, 22, 52]. This structure is consistent with the results of previous restriction mapping study [53], although the total lengths slightly differed.

Most of the essential photosynthesis-related genes were identified in the cp genome (Table 2; S2 Table). Twenty genes related to photosystem I and II, seventeen genes related to ATP synthase NADH dehydrogenase, six genes associated with the cytochrome b/f complex, a large subunit *rbcL*, two ORF, and other important genes related to ATP-dependent protease/ATP binding subunit (*clpP*), translational initiation factor (*infA*), Maturase K (*matK*), envelope membrane protein (*cemA*), subunit of acetyl-Co-Acarboxylase (*accD*), and c-type cytochrome synthesis gene (*ccsA*) were identified during annotation. The existence of chloroplast-encoded *clpP* in *B. sacra* suggested that this gene is indispensable for cell survival. Though this needs further verifications, but the presence of *clpP* indicates a better system to combat severe heat and drought stress in *B. sacra*, whereas *Cicer arietinum*, *Lathyrus sativus*, and *Pisum sativum* have lost an intron in both *clpP1* and *rps12* [54]. The *clpP* gene has also been identified in *Acacia ligulata* [19], *Trifolium subterraneum* [55], *Camellia* sp. [56], Actinidiaceae [57] and *Wollemia nobilis* [58].

General comparisons of *Boswellia sacra* and *Azadirachta indica*

The general characteristics of the *B. sacra* and *A. indica* cp genomes are presented in Table 1. The results showed a high level of similarity in the overall composition of the two cp genomes.

Table 2. List of Genes found in cp genome of *Boswellia sacra*.

Group of gene	Name of gene
Photosystem I	<i>psaA,psaB,psaC,psaI,psaJ</i>
Photosystem II	<i>psbA,psbB,psbC,psbD,psbE,psbF,psbH,psbI,psbJ,psbK,psbL,psbM,psbN,psbT,psbZ</i>
Cytochrome b/f complex	<i>petA,petB,petD,petG,petL,petN</i>
ATP synthase	<i>atpA,atpB,atpE,atpF,atpH,atpI</i>
NADH dehydrogenase	<i>ndhA,ndhB,ndhC,ndhD,ndhE,ndhF,ndhG,ndhH,ndhI,ndhJ,ndhK</i>
RubisCO	<i>large subunit rbcL</i>
RNA polymerase	<i>rpoA,rpoB,rpoC1,rpoC2</i>
Ribosomal proteins (SSU)	<i>rps2,rps3,rps4,rps7,rps8,rps11,rps12,rps12_3end,rps14,rps15,rps16,rps18,rps19</i>
Ribosomal proteins (LSU)	<i>rpl2,rpl14,rpl16,rpl20,rpl22,rpl23,rpl32,rpl33,rpl36</i>
Other genes	<i>clpP,matK,accD,ccsA,infA,cemA</i>
hypothetical chloroplast reading frames	<i>ycf1,ycf2,ycf3,ycf4,ycf15,ycf68</i>
ORFs	<i>orf42,orf56</i>
Transfer RNAs	<i>trnA-UUC, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-UCC, trnH-GUG, trnI-CAU, trnI-GAU, trnK-UUU, trnL-CAA, trnL-UAG, trnM-CAU, trnN-GUU, trnP-GGG, trnP-UGG, trnQ-UUG, trnR-ACG, trnR-UCU, trnS-GCU, trnS-GGA, trnT-GGU, trnV-GAC, trnV-UAC, trnW-CCA, trnY-GUA</i>
Ribosomal RNAs	<i>rrn4.5,rrn5,rrn16,rrn23</i>

doi:10.1371/journal.pone.0169794.t002

The *B. sacra* cp genome had approximately 37.61% GC content, which is higher than that of *A. indica* (37.5%). In the *B. sacra* cp genome, the intergenic regions, introns, and genic regions account for ca. 33.5%, 9.6%, and 57.5%, respectively, as shown in Tables 1 & 2. The pairwise alignment of these two cp genomes revealed approximately 9.52% sequence divergence (Table 3; S3 Table; S4 Table; S5 Table) comprising 8,929 substitutions (5.56%) and 3,971 Indels (3.96%). The sizes of the LSC and SSC regions were somehow different between these two cp genomes, primarily reflecting the large index in each region. The 10 most divergent intergenic regions included the spacer between *trnH*—*GUG-psbA* (457 bp, 40.54% divergence), *accD*—*psaI* (746 bp, 37.27% divergence), *atpH*—*atpI* (1,167 bp, 33.94% divergence), *psbZ*—*trnG-UCC* (546 bp, 28.14% divergence), *trnK*—*UUU-rps16* (1,034 bp, 26.40% divergence), *petN*—*psbM* (1,139 bp, 36.21% divergence), *ycf4*—*cemA* (942 bp, 26.69% divergence), *ccsA*—*ndhD* (307 bp, 25.24% divergence), *trnR-UCU*—*atpA* (198 bp, 28.97% divergence) and *trnL-UAG*—*ccsA* (111 bp, 33.05% divergence). These findings suggest that a deletion might have occurred in this region in *A. indica* after the divergence of *A. indica* and *B. sacra*.

In case of indels, the cp genome of *B. sacra* was highly conserved, however, in total 25 loci were identified with most variable regions. *ycf1* was highly variable region showing a 4354 indel sites, followed by *atpA* with 124 indel sites. The average indel length for *ycf1* was 174 bp, followed by *rpl14* with deletion of 25.5 bp. On the other hand, the indel diversity per site was highest for *rpl2* (S6 Table). We also compared the cp genomes of *B. sacra* and *A. indica* and calculated the average pairwise sequence divergence among 76 shared genes (S7 Table). Of these, the *B. sacra* genome has 0.066 average sequence divergence. Furthermore, among the genes identified to show highest level of divergence were *atpA*, *rpl2*, *rps12*, *ycf1*, *infA*, *rpl22*, *rpl32*, *rpl20*, *ndhF* and *rps8* (S3 Fig). The highest average sequence distance was found for *atpA* (0.852), followed by *rps11* (0.301).

The *trnH/psbA* intergenic region has been studied as a barcode of differentiation at the species level in the plastids of inland plants, such as *Araucaria* [57–59]. This region, highly

Table 3. Differences between the *B. sacra* and *A. indica* cp genomes.

Indel				
		Length (bp)	Count	
		1	2,743	
		2–10	1,216	
		11–21	12	
	Sum	6,353	3,971	Percentage ^a : 3.96%
Substitution				
		Type	Count	
		A<->T	1,571	
		C<->G	837	
		A<->C	1,416	
		T<->C	1,828	
		A<->G	1,777	
		T<->G	1,500	
	Sum		8,929	Percentage ^a : 5.56%
10 most divergent intergenic regions				
		Region	length ^b (bp)	Pairwise distance
		trnH—GUG-psbA	457	0.47
		accD—psal	746	0.32
		atpH—atpl	1,167	0.31
		psbZ—trnG-UCC	546	0.30
		trnK—UUU-rps16	1,034	0.25
		petN—psbM	1,139	0.25
		ycf4—cemA	942	0.25
		ccsA—ndhD	307	0.24
		trnR-UCU—atpA	198	0.23
		trnL-UAG—ccsA	111	0.21

^aRelative to the length of *B. sacra*.

^bLength in *B. sacra*.

doi:10.1371/journal.pone.0169794.t003

variable in length and sequence, is an on-coding region flanked by two conserved coding regions, i.e., *psbA*, which encodes photosystem II protein D1, and *trnH-GUG*, which functions in histidine translation [12–13]. The results revealed a 457 bp sequence *trnH/psbA* intergenic region in *B. sacra*. Yap et al. [57] suggested that this sequence was not present in *Agathis dammara* compared with *Wollemia nobili*. Previous detailed BLAST analyses have shown that this *Indel* is present in nineteen *Araucaria* species, including *W. nobili* [57]. In addition, *psbA* protects photosystem II from oxidative stress developed through drought. A previous report of Huo et al. [60] demonstrated that the overexpression of the *psbA* gene improves drought tolerance in tobacco plant. The occurrence of *psbA* in *B. sacra* suggests a strong ability to counteract oxidative stress, as this tree frequently encounters low-water conditions, however, this still needs further in-depth investigations.

Synteny between *B. sacra* and *A. indica*

A syntenic structure was also visualized in the MAUVE alignment for both cp genomes as shown in S1 Fig. The MAUVE alignment showed that there are two locally collinear blocks (LCBs) in the chloroplast genomes of *B. sacra* and *A. indica*. These regions showed homology in genome organization between the two species, although the sequences are inverted relative

to each other. The LCBs of protein coding, tRNA, and rRNA genes showed a similar synteny for both cp genomes. Similarly, the two cp genomes were run global alignment through VISTA to understand the comparative genomic organization [40]. The results of the alignment showed that both are having high sequence similarity. However, the marked differences were observed between the non-coding regions (*petN* and *psbM*; *accD* and *psaI*; *ycf4* and *cemA*; *rpl32* and *trnL*; *atpH* and *atpI*). The most divergent in coding regions were *trnI*, *matK*, *clpP*, *rpl33*, *accD*, *rps3*, *ndh* (A, H, D and G), *ycf3* and *ycf1* (Fig 3). Leliaert and Lopez-Bautista [61] revealed a similar homogenous matrix in cp genome sequencing between *Bryopsis plumosa* and *Tydemania expeditions*. However, dissimilarity in the synteny of some fragments of both cp genomes was observed, and this high dissimilarity is not unexpected, as high erraticism in the architecture of cpDNA has been observed among various species of congeneric taxa, such as picoplanktonic species, and *Trebouxiophyceae* and *Chlorophyceae* [62–63].

Analysis of repetitive sequences

Repetitive sequencing in cp genomes plays an important function and enhances the current understanding of the evolutionary aspects of plant species [62–63]. Mononucleotide microsatellite length polymorphisms have been used as markers to understand the evolutionary history of cp genomes, reflecting the high variability rates of these anomalies [56]. In the present study, a total of 109 simple sequence repeat (SSR) loci were observed in the *B. sacra* chloroplast genome, sharing 1,256 bp (ca. 0.78%) of the total sequence (Table 4). Most of the SSRs were identified in intergenic regions, and some SSRs were identified in coding areas, such as *rpoC2* and *ycf1*, which could also be used as molecular markers for population diversity studies [63]. A total 39 large repeats of more than 30 bp in length were identified in the *B. sacra* cp genome (Table 5).

In case of tandem repeat, some of the conifers have been noted for the presence of large number of such repeats [64]. These repeats have been associated with gene duplication and chloroplast rearrangement [65]. The results of tandem repeat analysis showed a total 30 repeats (Table 6). The highest size of repeats was 76 bp in length *accD* in the coding region whereas the lowest was 12 bp *trnK/rps16* in the intergenic spacer region, with average copy number 2.4. The percent of *indels* were ranged from 0 to 18 (Table 6). A majority of these repeats were located in intergenic spacers, while five repeat sequences were identified in the coding regions of *trnS-GCU*, *trnS-GGA*, *ycf2*, *psaB* and *psaA*. Eight long repeats were identified in *A. indica*, suggesting that these repeats might be well distributed among Sapindales. In a recent study, Addisalem et al. [65] developed the first set of 46 SSR markers and estimated the genome size for *B. papyrifera* as 705 Mb, representing the first genome sequence analysis for any species of this genus. However, the current knowledge of the genomics and transcriptomics of this key-stone species remains limited.

In case of single-nucleotide polymorphism (SNP) (a variation in a single nucleotide that occurs at a specific position in the genome), maximum likelihood analysis of natural selection codon-by-codon, was performed to estimate the number of inferred synonymous (s) and non-synonymous (n) substitutions which are presented along with the numbers of sites that are estimated to be synonymous (S) and non-synonymous (N) [66]. These estimates are produced using the joint Maximum Likelihood reconstructions of ancestral states under a Muse-Gaut Model [67] of nucleotide substitution. A total to 8,301 positions were noted in the chloroplast genomes of *B. sacra*, *A. indica* and *Citrus sinensis*. In case of codons, TCT, GTT, GGA and CGT showed high n, S, N and *dN-dS* (S8 Table). However, the *P* value was less than 0.05 showing a significant substitution among cp genomes. The normalized *dN/dS* ratio was calculated for the three cp genomes which showed a positive selection ranging from 0~10.17. The

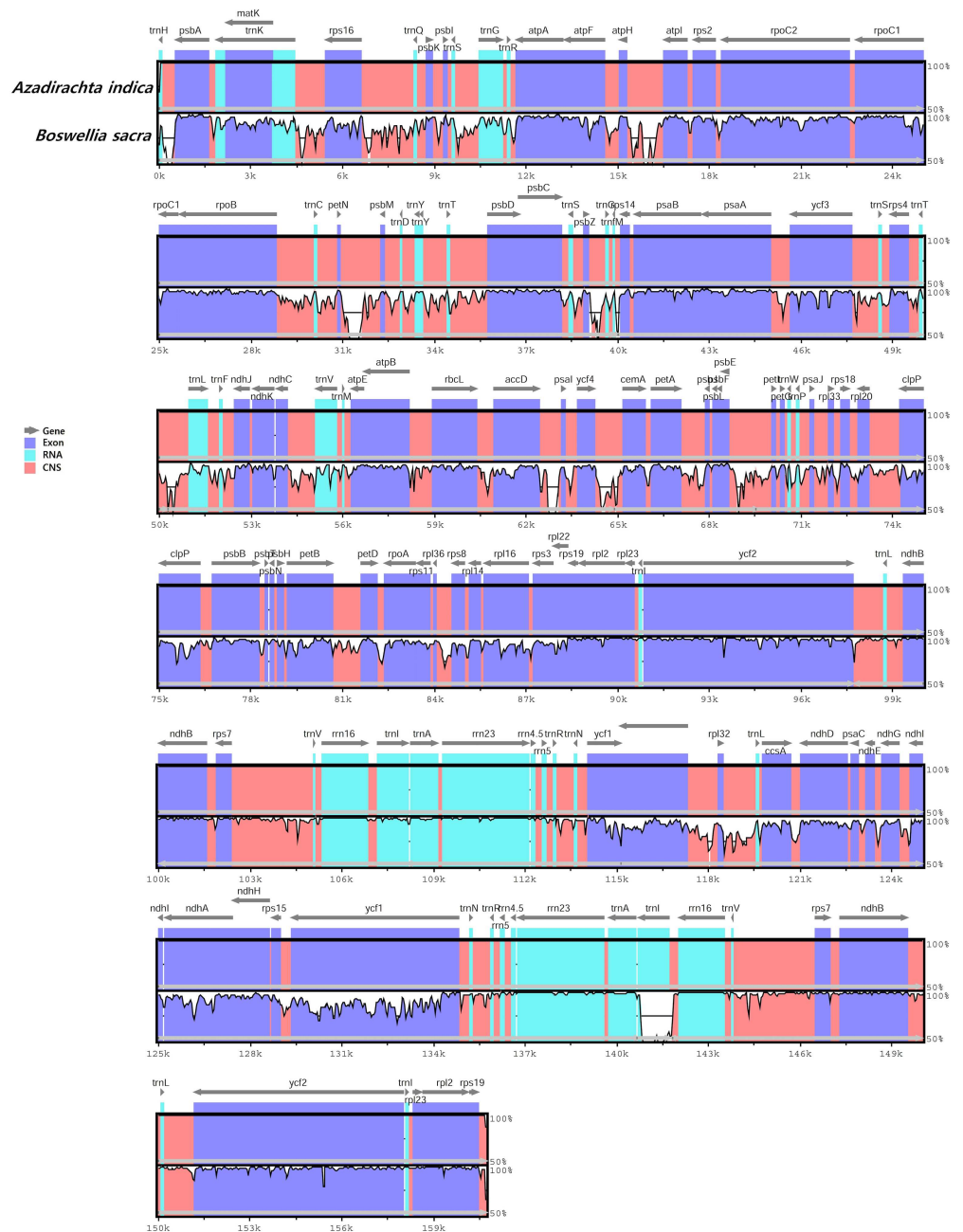


Fig 3. Visualization alignment of chloroplast genome sequences of *B. sacra* and *A. indica*. VISTA based similarity graphical information portraying sequence identity of *B. sacra* with reference *A. indica* cp genomes. Thick black lines show the inverted repeats (IRs) in the chloroplast genomes. Genome regions are color-coded as protein coding, rRNA coding, tRNA coding or conserved noncoding sequences (CNS) whereas arrows show the gene presence.

doi:10.1371/journal.pone.0169794.g003

number of segregating sites were 14,249 in three cp genomes with a nucleotide diversity was 0.063. The nucleotide frequencies are 30.59% (A), 31.21% (T/U), 19.49% (C), and 18.71% (G) in the cp genomes. The transition/transversion rate ratios are $k1 = 2.149$ (purines) and $k2 = 2.072$ (pyrimidines). The overall transition/transversion bias is $R = 0.996$. These shows that *B. sacra*, *A. indica* and *Citrus sinensis* species with long generation times, and both exhibit

Table 4. List of simple sequence repeats.

Repeat unit	Length (bp)	Number of SSRs	Start position ^a
AT	10	5	21733 (<i>rpoC2</i>); 50519; 63260; 120682; 122312 (<i>ndhD</i>)
	12	2	72021; 118991
TA	10	1	50086
	12	1	49380
TC	10	1	65118 (<i>cemA</i>)
AAG	12	1	98114
ATA	12	3	52097 ; 57925 (<i>atpB</i>); 70908
CTT	12	1	150474
TAT	12	1	50027
TTA	12	2	54843 ; 131240 (<i>ycf1</i>)
AAAT	12	1	6383
AATT	12	1	117168
ATAG	12	1	34089
CATT	12	1	131003 (<i>ycf1</i>)
CTTT	12	1	51864
TATC	12	1	38353
TTTC	12	1	125090
ATATGA	18	1	118904

^aThe SSR containing coding regions are indicated in parentheses. SSRs that are identical in the *A. indica* chloroplast genome are highlighted in bold.

doi:10.1371/journal.pone.0169794.t004

Table 5. List of long repeat sequence.

Repeat Size	Type ^a	Start position of 1st repeat	Start position the repeat found in other region	Location ^b	Region
30	P	9578	48299	<i>trnS-GCU, trnS-GGA</i>	LSC
30	D	30965	31304	IGS (<i>petN-psbM</i>)	LSC
30	D, P	95755	95773, 152797, 152815	<i>ycf2</i>	IR
31	D	50042	50183	IGS (<i>trnT-UGU-trnL-UAA</i>)	LSC
32	D	15474	15696, 16029	IGS (<i>atpH-atpI</i>)	LSC
34	P	5863	5863	intron (<i>rps16</i>)	LSC
36	D, P	103143	125725, 145421	IGS (<i>rps12-trnV-GAC</i>), intron (<i>ndhA</i>), IGS (<i>trnV-GAC-rps12</i>)	IR, SSC
40	D	64428	64797	IGS (<i>ycf4-cemA</i>)	LSC
42	D	49991	50135	IGS (<i>trnT-UGU-trnL-UAA</i>)	LSC
42	D	64386	64759	IGS (<i>ycf4-cemA</i>)	LSC
52	P	444	444	IGS (<i>trnH-GUG-psbA</i>)	LSC
56	P	31815	31815	IGS (<i>petN-psbM</i>)	LSC
65	D	15507	^{16061*}	IGS (<i>atpH-atpI</i>)	LSC
69	D	31107	31522	IGS (<i>petN-psbM</i>)	LSC
76	D	41575	43799	<i>psaB, psaA</i>	LSC
109	D	64274	64648	IGS (<i>ycf4-cemA</i>)	LSC
155	D	15417	15639	IGS (<i>atpH-atpI</i>)	LSC
204	D	15729	^{16061*}	IGS (<i>atpH-atpI</i>)	LSC
251	D	62442	62769	IGS (<i>accD-psaI</i>), <i>psaI</i>	LSC

^aD: direct repeat; P: palindrome inverted repeat.

^bIGS: intergenic spacer region. Sequences conserved in the *A. indica* chloroplast genome are highlighted in bold.

*The 65 bp sequence is truncated copy of 204 bp sequence.

doi:10.1371/journal.pone.0169794.t005

Table 6. Distribution of tandem repeats in the *B. sacra* chloroplast genome.

Indices	Repeat Length	Copy Number	Percent Indels	Percent Matches	Location
347–387	16	2.6	11	88	<i>trnH/psbA</i> (IGS)
4566–4594	12	2.4	0	100	<i>trnK/rps16</i> (IGS)
4819–4874	19	2.9	5	78	<i>trnK/rps16</i> (IGS)
7784–7814	16	1.9	0	100	<i>rps16/trnQ</i> (IGS)
8690–8726	18	2.1	0	94	<i>trnQ/psbK</i> (IGS)
9763–9828	27	2.4	5	84	<i>trnS/trnR</i> (IGS)
9783–9831	14	3.6	5	83	<i>trnS/trnR</i> (IGS)
9779–9867	27	3.5	13	85	<i>trnS/trnR</i> (IGS)
9863–9892	14	2.1	0	100	<i>trnS/trnR</i> (IGS)
11501–11572	31	2.5	18	81	<i>trnR/atpA</i> (IGS)
11504–11573	25	2.7	17	76	<i>trnR/atpA</i> (IGS)
11535–11583	22	2.4	12	80	<i>trnR/atpA</i> (IGS)
16355–16385	14	2.2	0	94	<i>atpH/atpI</i> (IGS)
29582–29606	12	2.1	0	100	<i>rpoB/trnC</i> (IGS)
31751–31776	13	2	0	100	<i>petN/psbM</i> (IGS)
49997–50046	18	2.9	11	82	<i>trnT/trnL</i> (IGS)
60254–60278	12	2.1	0	100	<i>rbcL/accD</i> (IGS)
62281–62429	76	2	5	88	<i>accD</i> (CDS)
67916–67954	19	2.1	4	90	<i>psbJ/psbL</i> (IGS)
69384–69454	12	6.1	18	72	<i>psbE/petL</i> (IGS)
72176–72214	21	1.9	5	94	<i>rps18</i> (CDS)
73287–73312	13	2	0	100	<i>rpl20/rps12</i> (IGS)
95746–95802	18	3.2	0	97	<i>ycf2</i> (CDS)
112196–112261	32	2.1	0	97	<i>rrn 4.5/rrn5</i> (IGS)
113767–113809	20	2.2	0	100	<i>trnN/ycf1</i> (IGS)
119576–119605	15	2	0	93	<i>ccsA</i> (CDS)
129059–129085	13	2.1	0	100	<i>rps15/ycf1</i> (IGS)
134790–134832	20	2.2	0	100	<i>ycf1</i> (CDS)
136338–136403	32	2.1	0	97	<i>rrn5/rrn4.5</i> (IGS)
152797–152853	18	3.2	0	97	<i>ycf2</i> (CDS)

doi:10.1371/journal.pone.0169794.t006

little changes in the structural organization of the chloroplast genomes. This could further reveal how these species evolved at low rates. Similar perspectives were shown for the chloroplast genomes of five *Quercus* species [66]. The estimation of synonymous and non-synonymous substitution rates may play an important role in understanding the dynamics of molecular evolution, and non-synonymous substitutions could be subject to natural selection during the evolutionary process [66–68].

Evolution of *orf56* and *ycf68* within *sapindales*

We compared the *orf56* and *ycf68* genes of three tree species viz. *B. sacra*, *A. indica* and *C. sinensis* (Fig 4), revealing the presence of these genes in the three species. The *orf56* gene is present in an intron in the *trnA-UGC* gene, which contains one homologous sequence, known as mitochondrial ACR-toxin sensitivity gene in *Citrus* trees [22, 69]. Interestingly, the full-length *orf56* gene sequence was different but was intact in all three species. Goremykin et al. [70] previously reported that the ACR-toxin sensitivity gene was transferred between plastid and mitochondrial genomes [22].

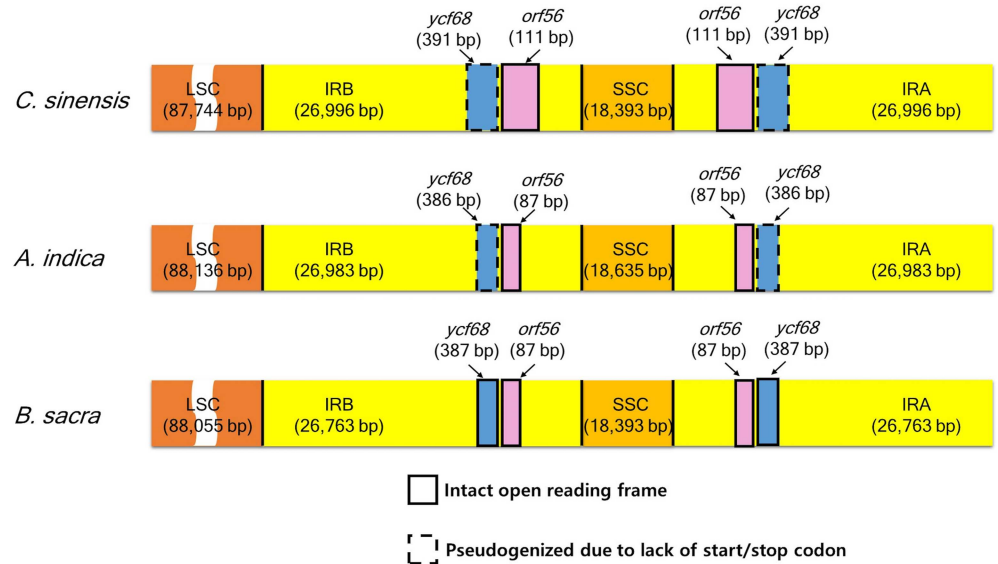


Fig 4. Comparison of the border positions of the LSC, SSC, and IR regions in three Sapindales species.

doi:10.1371/journal.pone.0169794.g004

The *ycf68* gene, in the *trnI*-GAU intron of the IR region, showed high similarity between *C. sinensis* and *B. sacra*. However, an additional T-insertion near the C-terminus was abolished at the stop codon [22]. *A. indica* has a T-deletion near the C-terminus abolished at the stop codon in the same position. Comparatively, an intact *ycf68* gene was detected in *B. sacra* (Fig 4). Raubeson et al. [71] suggested that *ycf68* is not a protein-coding gene based on the lack of intron folding patterns. The high levels of sequence conservation among the ORF of identified homologs indicates true functionality [72]. However, the function of this putative gene still remains unclear and needs further investigation.

Phylogenetic analysis of *B. sacra*

To obtain insight into the position of *B. sacra* within the Malvaceae, we generated datasets of 32 completely sequenced chloroplast genomes, using *Vitis vinifera* as an outgroup (S1 Table). We conducted a phylogenetic analysis of the *B. sacra* cp genome with the cp genomes from 8 species of Myrtaceae (*Corymbia eximia*, *Angophora costata*, *Eucalyptus umbra*, *E. deglupta*, *E. globulus*, *Stockwellia quadrifida*, *Syzygium cumini* and *Allosyncarpia ternata*), 2 species from Melianthaceae (*Francoa sonchifolia* and *Theorbroma cacao*), 3 species from Malvaceae (*Gossypium incanum*, *G. tomentosum*, and *G. hirsutum*), 15 species from Brassicaceae (*Barbarea verna*, *Nasturtium officinale*, *Lepidium virginicum*, *Pachycladon ensyisii*, *P. cheesemani*, *Crucihimalaya wallichii*, *Olimarabidopsis pumila*, *Capsella bursa-pastoris*, *Arabidopsis thaliana*, *Lobularia maritima*, *Brassica napus*, *Arabis hirsuta*, *Draba nemorosa*, *Aethionema cordifolium* and *Aethionema grandiflorum*), one species from Caricaceae (*Carica papaya*) and one species from Rutaceae (*Citrus sinensis*).

The phylogenetic analysis included some commercially important Sapindales plants and the more relevant *A. indica* species from Malvaceae. The neighbor-joining phylogenetic tree resolved 33 nodes, with strong bootstrap support of 55–100 (Fig 5). These results strongly support the position of the *B. sacra* within Sapindales. The phylogenetic trees obtained in the present study also indicate a close relationship between *B. sacra* and *A. indica* with high bootstrap support (100%). Krishnan et al. [73] suggested that both *A. indica* and *Citrus sinensis* form a homogenous group (100%), validating the conventional taxonomic classification of *A. indica*.

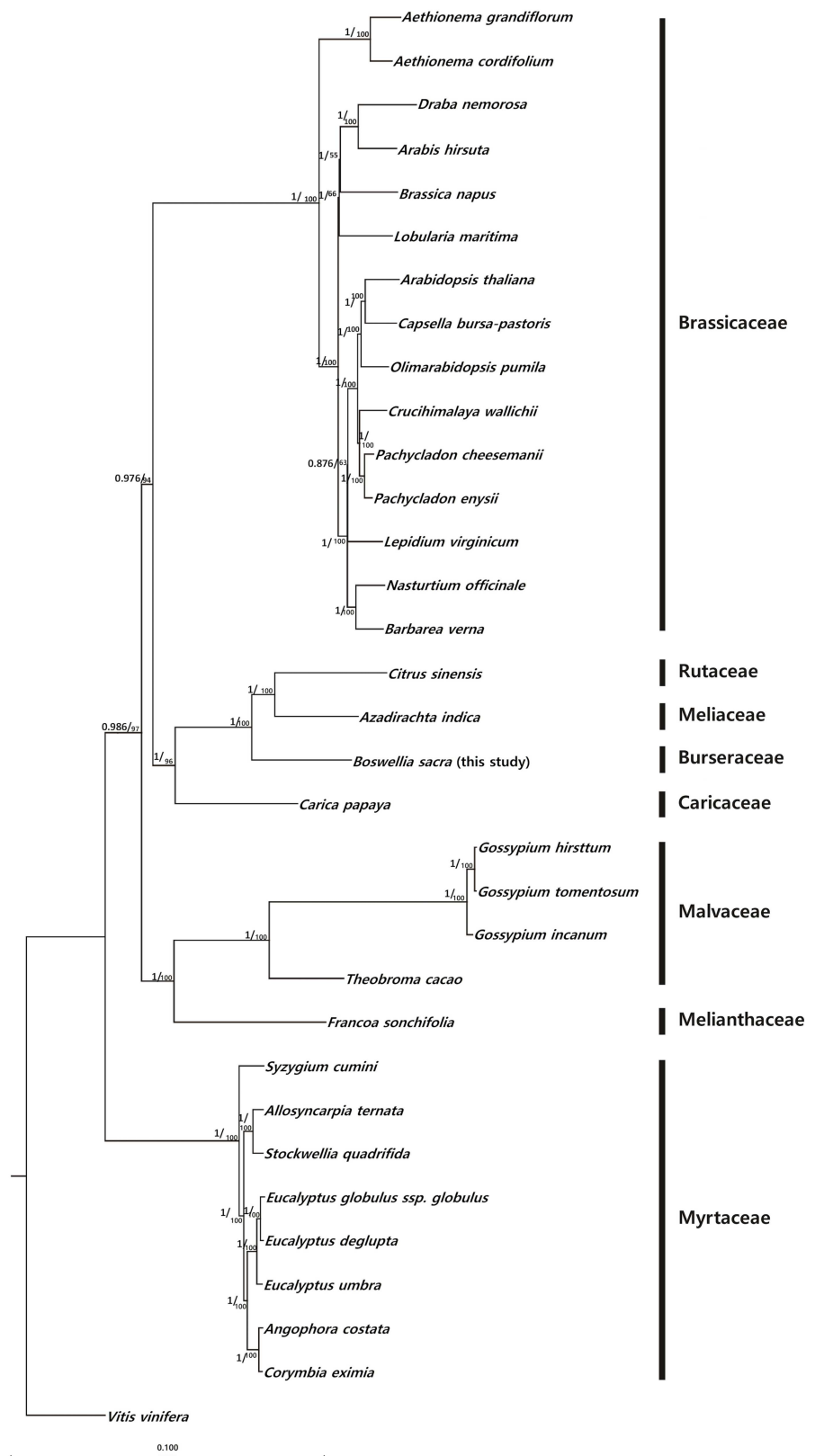


Fig 5. Neighbor-Joining phylogeny of the representative Malvidae lineages. The common grape vine (*Vitis vinifera*) is included as the outgroup to root the tree. A total of 33 complete chloroplast genomes were aligned. In total, 33 nodes were resolved. The position of *B. sacra* is shown in bold.

doi:10.1371/journal.pone.0169794.g005

Thus, the results of the phylogenetic analysis of the cp genome of both species showed the formation of a distinct cladogram with *B. sacra*, revealing new insights to understand the gene structure of these three plants.

The habitats of *B. sacra* and *A. indica* are different, however, the fact that both plants produce oleogum and belong to the same order suggests interesting prospects for broader evolutionary studies of Sapindales [74]. This phylogenetic relationship was further confirmed by constructing the maximum parsimony tree [48], suggesting a close homology with *C. sinensis* and *A. indica* (S2 Fig). The Bayesian Inference (BI) also highly supported as monophyletic (100/1.00) behavior *C. sinensis* and *A. indica*. Whereas a sister relationship of *B. sacra* was well supported (100/1.00). The posterior probabilities (PP) were also 1.0 for *B. sacra*, *C. sinensis* and *A. indica* (S4 Fig), suggesting a branching order as ((*C. sinensis*, *A. indica*), *B. sacra*). A similar phylogenomic reconstruction was also noticed by Wang et al. [16,75], where Actinidiaceae formed a monophyly relationship in Ericales.

Conclusions

Using a combination of *de novo* assembly and reference genome analysis, we provided the first completely sequenced chloroplast genome of endemic *Boswellia sacra*. The genome organization and gene structure is almost typical to that of angiosperms plants. The main difference in the gene composition between the *B. sacra* and *A. indica* cp genomes was the protein-coding gene in the *trnH-GUG-psbA* region. Notably, eight shared SSR regions were identified in the *B. sacra* and *A. indica* cp genome comparison. Elucidating these regions could provide additional phylogenetic insights at the taxonomic level, and these results could be used to develop molecular markers [72]. Previously, Coppi et al. [31] suggested that different stands of tree populations growing in different areas showed significant variations in genetic distances. The findings of the present study are essential to envisage fresh in-depth intuitions into various aspects of *Boswellia* chloroplast genome evolution. The current new genomic datasets will enable further exploration of the genetic diversity of *Boswellia sacra* and the other 18 species of this genus.

Supporting Information

S1 Table. List of the complete chloroplast genome sequences included in the phylogenetic analysis.

(XLSX)

S2 Table. Functions and products of the chloroplast genes identified in *B. sacra*.

(DOCX)

S3 Table. Sequences of important gene fragments in *B. sacra* and comparison with *A. indica*.

(XLSX)

S4 Table. Simple Sequence Repeat analysis of *B. sacra*.

(XLSX)

S5 Table. Pairwise distances of the important regions identified in *B. sacra* and *A. indica*.

(XLSX)

S6 Table. Comparative assessment of Indels *B. sacra* and *A. indica*.
(XLSX)

S7 Table. Pairwise distance among the shared genes of *B. sacra* and *A. indica*.
(XLSX)

S8 Table. Comparisons of synonymous (S) and nonsynonymous (N), transitional substitutions are shown in bold and those of transversalsal substitutions in chloroplast genome of *B. sacra*, *A. indica* and *Citrus sinensis*.
(XLS)

S1 Fig. MAUVE alignment of the *Boswellia sacra* chloroplast genome. The *B. sacra* genome is shown on top as the reference genome. Within each of the alignments, local collinear blocks are represented as blocks of the same color connected with lines. Annotations are shown above and below the LCBs, protein-coding genes are indicated as white boxes, tRNA genes are shown in green, and rRNA genes are shown in red. The lowered position of a box indicates an inverted orientation.
(TIFF)

S2 Fig. Maximum parsimony phylogeny of the representative cp genomes. The *Gossypium incanum* was included as an outgroup to root the tree. A total of 21 complete chloroplast genomes were aligned. In total, 21 nodes were resolved. The position of *B. sacra* is shown in green circular symbol.
(TIF)

S3 Fig. Pairwise distances of shared 76 genes *B. sacra* and *A. indica*.
(TIF)

S4 Fig. Posterior probabilities out put of the selected cp genomes in relation with *B. sacra* and *A. indica*.
(PDF)

Author Contributions

Conceptualization: ALK AAH AAR.

Data curation: CEP GSP JHS ALK.

Formal analysis: SA.

Funding acquisition: AAH AAR ALK.

Investigation: ALK.

Methodology: ALK ARK.

Project administration: AAH.

Resources: AAH IJL AAR.

Software: IJL JHS.

Supervision: AAR AAH.

Validation: JHS IJL SA.

Visualization: JHS SA.

Writing – original draft: ALK AAH.

Writing – review & editing: ALK AAH.

References

1. Jin SX, Daniell H. The Engineered Chloroplast Genome Just Got Smarter. *Trends Plant Sci.* 2015; 20(10):622–40. doi: [10.1016/j.tplants.2015.07.004](https://doi.org/10.1016/j.tplants.2015.07.004) PMID: [26440432](https://pubmed.ncbi.nlm.nih.gov/26440432/)
2. Qian J, Song JY, Gao HH, Zhu YJ, Xu J, Pang XH, et al. The Complete Chloroplast Genome Sequence of the Medicinal Plant *Salvia miltiorrhiza*. *Plos One.* 2013; 8(2).
3. Sugiura M. 1992. The chloroplast genome. *Plant Molecular Biology* 19:149e168
4. Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, et al. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature.* 1986; 322(6079):572–4.
5. Xie QJ, Michaeli S, Peled-Zehavi N, Galili G. Chloroplast degradation: one organelle, multiple degradation pathways. *Trends Plant Sci.* 2015; 20(5):264–5. doi: [10.1016/j.tplants.2015.03.013](https://doi.org/10.1016/j.tplants.2015.03.013) PMID: [25865278](https://pubmed.ncbi.nlm.nih.gov/25865278/)
6. Jensen RG. Biochemistry of the chloroplast. *Biochemistry Plants.* 2013; 1:273–313
7. Cosner ME, Raubeson LA, Jansen RK. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *Bmc Evol Biol.* 2004;4.
8. Jansen RK, Ruhlman TA. Plastid genomes of seed plants. In *Genomics of chloroplasts and mitochondria 2012* (pp. 103–126). Springer Netherlands.
9. Saski C, Lee SB, Fjellheim S, Guda C, Jansen RK, Luo H, et al. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theor Appl Genet.* 2007; 115(4):571–90. doi: [10.1007/s00122-007-0567-4](https://doi.org/10.1007/s00122-007-0567-4) PMID: [17534593](https://pubmed.ncbi.nlm.nih.gov/17534593/)
10. Zhang Q, Liu Y, Sodmergen. Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species. *Plant Cell Physiol.* 2003; 44(9):941–51. PMID: [14519776](https://pubmed.ncbi.nlm.nih.gov/14519776/)
11. Corriveau JL, Coleman AW. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Am J Bot.* 1988; 75: 1443–1458.
12. Sugiura M. The chloroplast genome. *Essays Biochem.* 1995; 30:49–57. Epub 1995/01/01. PMID: [8822148](https://pubmed.ncbi.nlm.nih.gov/8822148/)
13. Moore MJ, Bell CD, Soltis PS, Soltis DE. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *P Natl Acad Sci USA.* 2007; 104(49):19363–8.
14. Jansen RK, Kaitanis C, Saski C, Lee SB, Tomkins J, Alverson AJ, et al. Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. *Bmc Evol Biol.* 2006; 6.
15. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences.* 2010; 107(10):4623–8.
16. Wang L, Wuyun TN, Du HY, Wang DP, Cao DM. Complete chloroplast genome sequences of *Eucommia ulmoides*: genome structure and evolution. *Tree Genet Genomes.* 2016; 12(1).
17. Shang C, Du FK, Yin K, Zhang Z. The complete chloroplast genome of *Cathay Poplar*: *Populus cathayana* Rehder. *Mitochondrial DNA Part B.* 2016; 1(1):86–7.
18. Du FK, Lang TG, Lu SH, Wang YY, Li JQ, Yin KQ. An improved method for chloroplast genome sequencing in non-model forest tree species. *Tree Genet Genomes.* 2015; 11(6).
19. Williams AV, Boykin LM, Howell KA, Nevill PG, Small I. The Complete Sequence of the *Acacia ligulata* Chloroplast Genome Reveals a Highly Divergent *clpP1* Gene (vol 10, e0125768, 2015). *Plos One.* 2015; 10(9).
20. Li ZH, Qian ZQ, Liu ZL, Deng TT, Zu YM, Zhao P, et al. The complete chloroplast genome of *Armand pine* *Pinus armandii*, an endemic conifer tree species to China. *Mitochondr DNA.* 2016; 27(4):2635–6.
21. Huang YY, Matzke AJM, Matzke M. Complete Sequence and Comparative Analysis of the Chloroplast Genome of Coconut Palm (*Cocos nucifera*). *Plos One.* 2013; 8(8).
22. Su HJ, Hogenhout SA, Al-Sadi AM, Kuo CH. Complete Chloroplast Genome Sequence of Omani Lime (*Citrus aurantiifolia*) and Comparative Analysis within the Rosids. *Plos One.* 2014; 9(11).

23. Martin G, Baurens FC, Cardi C, Aury JM, D'Hont A. The Complete Chloroplast Genome of Banana (*Musa acuminata*, Zingiberales): Insight into Plastid Monocotyledon Evolution. *Plos One*. 2013; 8(6).
24. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, et al. The Norway spruce genome sequence and conifer genome evolution. *Nature*. 2013; 497(7451):579–84. doi: [10.1038/nature12211](https://doi.org/10.1038/nature12211) PMID: [23698360](https://pubmed.ncbi.nlm.nih.gov/23698360/)
25. Raffaelli M, Mosti S, Tardelli M. The Frankincense Tree (*Boswellia sacra* Flueck., Burseraceae) in Dhofar, southern Oman: field-investigations on the natural populations. *Webbia*. 2003; 58(1):133–49.
26. Farah M. Non-Timber Forest Product (NTFP) Extraction in Arid Environments: Land-use Change, Frankincense Production and the Sustainability of *Boswellia sacra* in Dhofar (Oman). The University of Arizona.
27. Raffaelli M, Mosti S, Tardelli M. *Boswellia sacra* Flueck. (Burseraceae) in the Hasik area (Eastern Dhofar, Oman) and a list of the surrounding flora. *Webbia*. 2006; 61(2):245–51.
28. Khan AL, Al-Harrasi A, Al-Rawahi A, Al-Farsi Z, Al-Mamari A, Waqas M, et al. Endophytic Fungi from Frankincense Tree Improves Host Growth and Produces Extracellular Enzymes and Indole Acetic Acid. *Plos One*. 2016; 11(6).
29. Al-Harrasi A, Al-Saidi S. Phytochemical analysis of the essential oil from botanically certified Oleogum Resin of *Boswellia sacra* (Omani luban). *Molecules*. 2008; 13(9):2181–9. PMID: [18830149](https://pubmed.ncbi.nlm.nih.gov/18830149/)
30. Takahashi M, Sung B, Shen Y, Hur K, Link A, Boland CR, et al. Boswellic acid exerts antitumor effects in colorectal cancer cells by modulating expression of the let-7 and miR-200 microRNA family. *Carcinogenesis*. 2012; 33(12):2441–9. doi: [10.1093/carcin/bgs286](https://doi.org/10.1093/carcin/bgs286) PMID: [22983985](https://pubmed.ncbi.nlm.nih.gov/22983985/)
31. Coppi A, Cecchi L, Selvi F, Raffaelli M. The Frankincense tree (*Boswellia sacra*, Burseraceae) from Oman: ITS and ISSR analyses of genetic diversity and implications for conservation. *Genet Resour Crop Ev*. 2010; 57(7):1041–52.
32. Tolera M, Sass-Klaassen U, Eshete A, Bongers F, Sterck FJ. Frankincense tree recruitment failed over the past half century. *Forest Ecology and Management*. 2013; 304:65–72.
33. Shi C, Hu N, Huang H, Gao J, Zhao YJ, Gao LZ. An Improved Chloroplast DNA Extraction Procedure for Whole Plastid Genome Sequencing. *Plos One*. 2012; 7(2).
34. Xie ZZ, Yu CP, Guo L, Li MM, Yong Z, Liu XC, et al. Ion Torrent next-generation sequencing reveals the complete mitochondrial genome of black and reddish morphs of the Coral Trout *Plectropomus leopardus*. *Mitochondr DNA*. 2016; 27(1):609–12.
35. Tarkka MT, Herrmann S, Wubet T, Feldhahn L, Recht S, Kurth F, et al. OakContigDF159.1, a reference library for studying differential gene expression in *Quercus robur* during controlled biotic interactions: use for quantitative transcriptomic profiling of oak roots in ectomycorrhizal symbiosis. *New Phytol*. 2013; 199(2):529–40. Epub 2013/05/16. doi: [10.1111/nph.12317](https://doi.org/10.1111/nph.12317) PMID: [23672230](https://pubmed.ncbi.nlm.nih.gov/23672230/)
36. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004; 20(17):3252–5. doi: [10.1093/bioinformatics/bth352](https://doi.org/10.1093/bioinformatics/bth352) PMID: [15180927](https://pubmed.ncbi.nlm.nih.gov/15180927/)
37. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997; 25.
38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215(3):403–10. Epub 1990/10/05. doi: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/)
39. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004; 14(7):1394–403. doi: [10.1101/gr.2289704](https://doi.org/10.1101/gr.2289704) PMID: [15231754](https://pubmed.ncbi.nlm.nih.gov/15231754/)
40. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, et al. VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics*. 2000; 16.
41. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001; 29(22):4633–42. PMID: [11713313](https://pubmed.ncbi.nlm.nih.gov/11713313/)
42. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999; 27(2):573–80. Epub 1998/12/24. PubMed Central PMCID: [PMC148217](https://pubmed.ncbi.nlm.nih.gov/PMC148217/). PMID: [9862982](https://pubmed.ncbi.nlm.nih.gov/9862982/)
43. Asaf S, Khan AL, Khan AR, Waqas M, Kang S-M, Khan MA, et al. Complete chloroplast genome of *Nicotiana glauca* and its comparison with related species. *Front Plant Sci*. 2016; 7.
44. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 2005; 33(2):511–8. doi: [10.1093/nar/gki198](https://doi.org/10.1093/nar/gki198) PMID: [15661851](https://pubmed.ncbi.nlm.nih.gov/15661851/)
45. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 2003; 52(5):696–704. PMID: [14530136](https://pubmed.ncbi.nlm.nih.gov/14530136/)
46. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009; 25(11):1451–2. doi: [10.1093/bioinformatics/btp187](https://doi.org/10.1093/bioinformatics/btp187) PMID: [19346325](https://pubmed.ncbi.nlm.nih.gov/19346325/)

47. Guo ZY, Zhang HR, Shrestha N, Zhang XC. Complete Chloroplast Genome of a Valuable Medicinal Plant, *Huperzia Serrata* (Lycopodiaceae), and Comparison with Its Congener. *Appl Plant Sci*. 2016; 4 (11).
48. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016; 33(7):1870–4. doi: [10.1093/molbev/msw054](https://doi.org/10.1093/molbev/msw054) PMID: [27004904](https://pubmed.ncbi.nlm.nih.gov/27004904/)
49. Williams AV, Miller JT, Small I, Nevill PG, Boykin LM. Integration of complete chloroplast genome sequences with small amplicon datasets improves phylogenetic resolution in *Acacia*. *Mol Phylogenet Evol*. 2016; 96:1–8. doi: [10.1016/j.ympev.2015.11.021](https://doi.org/10.1016/j.ympev.2015.11.021) PMID: [26702955](https://pubmed.ncbi.nlm.nih.gov/26702955/)
50. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001; 17(8):754–5. Epub 2001/08/29. PMID: [11524383](https://pubmed.ncbi.nlm.nih.gov/11524383/)
51. Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods), Version 4. Sunderland, MA: Sinauer Associates; 2002.
52. Bausher MG, Singh ND, Lee SB, Jansen RK, Daniell H. The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. *Bmc Plant Biol*. 2006; 6.
53. Hudson KR, Gardner RC. Organisation of the chloroplast genome of kiwifruit (*Actinidia deliciosa*). *Curr Genet*. 1988; 13(4):339–42.
54. Jansen RK, Wojciechowski MF, Sanniyasi E, Lee SB, Daniell H. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of rps12 and clpP intron losses among legumes (Leguminosae). *Mol Phylogenet Evol*. 2008;48.
55. Sveinsson S, Cronk Q. Evolutionary origin of highly repetitive plastid genomes within the clover genus (*Trifolium*). *Bmc Evol Biol*. 2014; 14. doi: [Artn 228 10.1186/S12862-014-0228-6](https://doi.org/Artn22810.1186/S12862-014-0228-6).
56. Yang JB, Yang SX, Li HT, Yang J, Li DZ. Comparative chloroplast genomes of *Camellia* species. *Plos One*. 2013; 8.
57. 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
58. Yap JY, Rohner T, Greenfield A, Merwe M, McPherson H, Glenn W. Complete chloroplast genome of the wollemi pine (*Wollemia nobilis*): structure and evolution. *Plos One*. 2015; 10.
59. 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. *Am J Bot*. 2007; 94(3):275–88. 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/)
60. Huo YJ, Wang MP, Wei YY, Xia ZL. Overexpression of the Maize psbA Gene Enhances Drought Tolerance Through Regulating Antioxidant System, Photosynthetic Capability, and Stress Defense Gene Expression in Tobacco. *Front Plant Sci*. 2016; 6.
61. Leliaert F, Lopez-Bautista JM. The chloroplast genomes of *Bryopsis plumosa* and *Tydemania expeditiones* (Bryopsidales, Chlorophyta): compact genomes and genes of bacterial origin. *Bmc Genomics*. 2015; 16.
62. Letsch MR, Lewis LA. Chloroplast gene arrangement variation within a closely related group of green algae (Trebouxiophyceae, Chlorophyta). *Mol Phylogenet Evol*. 2012; 64(3):524–32. doi: [10.1016/j.ympev.2012.05.027](https://doi.org/10.1016/j.ympev.2012.05.027) PMID: [22659018](https://pubmed.ncbi.nlm.nih.gov/22659018/)
63. Lemieux C, Otis C, Turmel M. Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. *Bmc Genomics*. 2014; 15.
64. Qin Z, Wang YP, Wang QM, Li AX, Hou FY, Zhang LM. Evolution Analysis of Simple Sequence Repeats in Plant Genome. *Plos One*. 2015; 10(12).
65. Addisalem AB, Esselink GD, Bongers F, Smulders MJM. Genomic sequencing and microsatellite marker development for *Boswellia papyrifera*, an economically important but threatened tree native to dry tropical forests. *Aob Plants*. 2015; 7.
66. Yang YC, Zhou T, Duan D, Yang J, Feng L, Zhao GF. Comparative Analysis of the Complete Chloroplast Genomes of Five *Quercus* Species. *Front Plant Sci*. 2016; 7.
67. Muse SV, Gaut BS. A Likelihood Approach for Comparing Synonymous and Nonsynonymous Nucleotide Substitution Rates, with Application to the Chloroplast Genome. *Mol Biol Evol*. 1994; 11(5):715–24. PMID: [7968485](https://pubmed.ncbi.nlm.nih.gov/7968485/)
68. Seo TK, Kishino H. Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. *Syst Biol*. 2008; 57(3):367–77. doi: [10.1080/10635150802158670](https://doi.org/10.1080/10635150802158670) PMID: [18570032](https://pubmed.ncbi.nlm.nih.gov/18570032/)
69. Ohtani K, Yamamoto H, Akimitsu K. Sensitivity to *Alternaria alternata* toxin in citrus because of altered mitochondrial RNA processing. *P Natl Acad Sci USA*. 2002; 99(4):2439–44.

70. Goremykin V, Hirsch-Ernst KI, Wolf S, Hellwig FH. The chloroplast genome of the "basal" angiosperm *Calycanthus fertilis*—structural and phylogenetic analyses. *Plant Syst Evol.* 2003; 242(1–4):119–35.
71. 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.
72. Jansen RK, Cai Z, Raubeson LA, Daniell H, de Pamphilis CW, Leebens-Mack J, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci U S A.* 2007; 104.
73. Krishnan NM, Pattnaik S, Jain P, Gaur P, Choudhary R, Vaidyanathan S, et al. A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica*. *Bmc Genomics.* 2012; 13.
74. Inamdar JA, Subramanian RB, Mohan JSS. Studies on the Resin Glands of *Azadirachta indica* A. Juss. (Meliaceae). *Ann Bot-London.* 1986; 58(3):425–9.
75. Wang WC, Chen SY, Zhang XZ. Chloroplast Genome Evolution in Actinidiaceae: clpP Loss, Heterogeneous Divergence and Phylogenomic Practice. *Plos One.* 2016; 11(9).