

RESEARCH ARTICLE

First complete chloroplast genomics and comparative phylogenetic analysis of *Commiphora gileadensis* and *C. foliacea*: Myrrh producing trees

Arif Khan^{1,2}, Sajjad Asaf¹, Abdul Latif Khan^{1*}, Ahmed Al-Harrasi^{1*}, Omar Al-Sudairy¹, Noor Mazin Abdulkareem¹, Adil Khan^{1,2}, Tariq Shehzad³, Nadiya Alsaady⁴, Ali Al-Lawati⁴, Ahmed Al-Rawahi¹, Zabta Khan Shinwari²

1 Natural and Medical Sciences Research Center, University of Nizwa, Nizwa, Oman, **2** Department of Biotechnology, Quaid-i-Azam University, Islamabad, Pakistan, **3** Plant Genome Mapping Lab, Center for Applied Genetic Technologies, University of Georgia, Georgia, United States of America, **4** Oman Animal & Plant Genetic Resources Center, The Research Council, Muscat, Oman

These authors contributed equally to this work.

* aharrasi@unizwa.edu.om (AAH); abdullatif@unizwa.edu.om (ALK)



OPEN ACCESS

Citation: Khan A, Asaf S, Khan AL, Al-Harrasi A, Al-Sudairy O, Abdulkareem NM, et al. (2019) First complete chloroplast genomics and comparative phylogenetic analysis of *Commiphora gileadensis* and *C. foliacea*: Myrrh producing trees. PLoS ONE 14(1): e0208511. <https://doi.org/10.1371/journal.pone.0208511>

Editor: Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

Received: July 11, 2018

Accepted: November 18, 2018

Published: January 10, 2019

Copyright: © 2019 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 the data is available with the MS in figure and supplementary files in this submission.

Funding: The authors are thankful to The Research Council, Oman for their financial support to ORG/BER/15/007. The funder 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.

Abstract

Commiphora gileadensis and *C. foliacea* (family *Burseraceae*) are pantropical in nature and known for producing fragrant resin (myrrh). Both the tree species are economically and medicinally important however, least genomic understanding is available for this genus. Herein, we report the complete chloroplast genome sequences of *C. gileadensis* and *C. foliacea* and comparative analysis with related species (*C. wightii* and *Boswellia sacra*). A modified chloroplast DNA extraction method was adopted, followed with next generation sequencing, detailed bioinformatics and PCR analyses. The results revealed that the cp genome sizes of *C. gileadensis* and *C. foliacea*, are 160,268 and 160,249 bp, respectively, with classic quadripartite structures that comprises of inverted repeat's pair. Overall, the organization of these cp genomes, GC contents, gene order, and codon usage were comparable to other cp genomes in angiosperm. Approximately, 198 and 175 perfect simple sequence repeats were detected in *C. gileadensis* and *C. foliacea* genomes, respectively. Similarly, 30 and 25 palindromic, 15 and 25 forward, and 20 and 25 tandem repeats were determined in both the cp genomes, respectively. Comparison of these complete cp genomes with *C. wightii* and *B. sacra* revealed significant sequence resemblance and comparatively highest deviation in intergenic spacers. The phylo-genomic comparison showed that *C. gileadensis* and *C. foliacea* form a single clade with previously reported *C. wightii* and *B. sacra* from family *Burseraceae*. Current study reports for the first time the cp genomics of species from *Commiphora*, which could be helpful in understanding genetic diversity and phylogeny of this myrrh producing species.

Introduction

The family *Burseraceae* comprises 18 genera and about 700 species [1]. The family has pan-tropical nature and is known for its fragrant resin, such as myrrh and frankincense. The family comprises of timber trees, small trees and shrubs [2,3]. The genus *Commiphora* comprises 190 plant species and distributed in southern Arabia (Yemen, Oman), northeastern Africa (Somalia, Ethiopia, Sudan) and subcontinent (India, Pakistan) [4–6]. The resin obtained from the tree by tapping is widely used in perfume, fragrance and medicinal products [3]. In indigenous medicine, resin based recipes are used for gastrointestinal, arthritis, wounding, obesity, pain and parasitic infections [7]. In the Sultanate of Oman, several *Commiphora* species are reported such as *C. gileadensis*, *C. foliacea*, and *C. habessinica*, [3].

C. gileadensis is widely known in the Mediterranean basin, especially on border of Oman, Saudi Arabia, Yemen and Somalia [1]. It is also known as balsam and commonly used for production of expensive perfumes [8,9]. Its sap, wood bark and seeds are used for medicinal purposes. Similarly, *C. gileadensis* yields in the production of very fragrant gum type resin, when the bark of the tree is damaged [10]. *C. gileadensis* was recognized in ancient times as a perfume and incense plant [11]. *C. gileadensis* also possess antibacterial properties and the people use it for treatment of infections [12]. *Commiphora* is used for the treatment of an opportunistic fungal infection in many countries of Africa [12]. *C. foliacea* was initially considered as endemic to Oman [1], but this specie was also reported in southern coast line of Yemen and Somalia [13,14].

Studying the genomics of ecologically and medicinally important wild trees can help in understanding the tree life, evolution, taxonomy and genetic diversity. In this regard, chloroplast (an important player of photosynthesis) genomics have been widely used in phylogenetic studies due to its maternal inheritance and recombination free nature [15]. The high conserved structure of chloroplast facilitates; primer designing, sequencing and used as a barcode for the identification of plants [16,17]. It contains its own independent genome, which encodes for specific proteins [18]. The genome is circular in structure that varies from 120 kb to 170kb and quadripartite configurations [19]. The chloroplast genome is composed of small single copy (SSC) region and large single copy (LSC) regions, separated by two copies of inverted repeats (IRa and IRb) [19]. They also provide important information in taxonomic and phylogenetic context on basis of differences in the sequences among plant species [20,21]. Chloroplast is haploid, maternally inherited and possess high conservation in gene content, which make it a good choice for studying evolutionary relationship in plants at any taxonomic levels [20]. The first complete chloroplast genome of the angiosperms were reported in tobacco [22]. Advances and rapid evolution in NGS (next-generation sequencing) technologies have made it possible the rapid sequencing of complete chloroplast genome sequences at much cheaper price. Up till now over 2700 cp genome sequences are submitted to National Center for Biotechnology Information (NCBI) including all of major groups of the plant kingdoms. However, still there are numerous economically and medicinally important plants species, which needs to be explored and understood in term of their chloroplast genome structure, organization and genetic evolution. Current study is our first effort to understand the two unexplored species *C. gileadensis* and *C. foliacea*. We sequenced the cp genomes and performed a detailed comparison with *C. wightii* and *B. sacra* to understand the genome structure, variation and phylogenetic placements.

Material and method

Ethic statement

The leaf samples were collected with care and trees were treated ethically. During sample collection, the local environment was not harmed. Permission was granted by Ministry of

Environment, Muscat, Sultanate of Oman to collect leaf samples for research purpose. The current study did not involve endangered or protected species.

Sample collection

Leaf samples were collected from Wadi Darbaat, Dhofar-Oman (17 31.237°N 55' 12.923°E). The samples include fresh and young photosynthetic leaves of *C. gileadensis* and *C. foliacea*. The collected samples were kept immediately in liquid nitrogen and then stored at -80°C until chloroplast DNA extraction.

Chloroplast DNA extraction and sequencing

Leaf samples of *C. gileadensis* and *C. foliacea* were cleaned and washed with sterilized water, air dried and kept in dark for 48 hrs in order to reduce the starch content in leaf tissues. Chloroplast DNA was extracted by the protocol of Shi et al, [23] with modifications to remove the traces of resinous content from tissues. The workflow of Ion Torrent S5 Sequencer (Life Technologies, USA) was used for extracted cp DNA sequencing. Chloroplast DNA were enzymatically sheared for 400 bp using the Ion Shear Plus Reagents and library were prepared following the protocol of Ion S5 with Ion Xpress Plus DNA Fragment Library kit. Prepared libraries were checked on Qubit fluorimeter and bioanalyzer (Agilent 2100, CA, USA) for quality check and standardization. Ion One Touch 2 instrument was used for template amplification, post template amplification, whereas the enrichment process was carried out with Ion One Touch ES enrichment system. The sample was loaded onto the Ion S5 Chip and sequencing were performed according to the protocol of Ion Torrent S5.

Genome assembly

The quality of raw reads were evaluated by using the FastQC [24]. Adapters were removed from both end of the contigs and Platanus_trim (v.1.0.7) [25] with phred score >30 was used to trim high quality reads. The chloroplast genomes of both *Commiphora* species were first *de novo* assembled. In order to get contamination free read of chloroplast genome from mitochondrial and nuclear genomes, the *Commiphora* species genomes paired end reads were obtained by mapping the high quality reads to a selected reference genome of *C. wightii* (NC036978) with Bowtie2 (v.2.2.3) [26]. The selected resultant reads were assembled using Spades (v.3.7.1) software [27] and the parameters were set to default. The regions which was uncertain in these genomes such as IR junctions region were picked out from the already published genome of *C. wightii* and *B. sacra* (NC036978 and NC029420, respectively), to adjust the sequence length, iteration method was used with software MITObim (v.1.8) [28]. The complete genome sequences were deposited in Gene Bank of NCBI, where *C. gileadensis* and *C. foliacea* were given MH042752 and MH041484 accession numbers, respectively.

Genome annotation

Chloroplast genomes were annotated by using Dual Organellar Genome Annotator (DOGMA) [29] and BLASTX and BLASTN were used to identify the positions of ribosomal RNAs, transfer RNAs and coding genes, tRNAscan-SE77 software was used to annotate tRNA genes. Furthermore, for manual adjustment, Geneious Pro (v.10.2.3) [30] and tRNAscan-SE [31] were used to compare it with previously reported *C. wightii* genome. Similarly, the start and stop codon and intron boundaries were also manually adjusted compared with pre sequenced *C. wightii* and *B. sacra*. Furthermore, the structural features of both *Commiphora* species cp genome were illustrated using OGDRAW [32]. Similarly, MEGA6 software [33] was used to determine the relative

synonymous codon usage and divergence in usage of identical codons. The divergence of these two *Commiphora* species cp genome with other related species were determined by using mVISTA [34] in Shuffle—LAGAN mode and using *C. wightii* as a reference genome.

Repeat identification

REPuter software [35] was used for the identification of palindromic, tandem and forward repeats present in genome. The criterion was minimum >15 base pairs with sequence identity of 90%. SSRs dataset was determined through PHOBOS ver3.3.12 [36] inclusive of attributed sets with (i) mononucleotide repeats ≥ 10 repeat units (ii) dinucleotide repeats ≥ 8 repeat units (iii) tri nucleotide and tetra nucleotide repeats ≥ 4 repeat units, and (iv) penta nucleotide and hexa nucleotide repeats ≥ 3 repeat units. Tandem Repeats Finder version 4.07 b [37] with default settings was used to determined tandem repeats.

Sequence-divergence and Phylo-genomic analysis

In this analysis, average-pairwise sequence divergence of complete plastomes and shared genes of *Commiphora* species with related species were determined. Missing and ambiguous gene annotations were confirmed by comparative sequence analysis after a multiple sequence alignment and gene order comparisons using Geneious Pro (v.10.2.3) [30] as reported previously [38,39]. These regions were aligned using MAFFT version 7.222 [40] with default parameters. Pairwise sequence divergence was calculated by selected Kimura's two-parameter (K2P) model [41]. Similarly, a custom Python script (<https://www.biostars.org/p/119214/>) and DnaSP 5.10.01 [42], were employed to determine single-nucleotide polymorphisms and *Indel* polymorphisms among the complete genomes respectively. To infer the phylogenetic position of both *C. gileadensis* and *C. foliacea* within the order Sapindales, 24 cp genomes were downloaded from the NCBI database for analysis. Multiple alignments were performed using complete cp genomes based on conserved structures and gene order [41] and 4 different methods were used to make the trees: Bayesian-inference (MrBayes v3.1.2 [43]), maximum parsimony (PAUP-4.0[44]), maximum-likelihood and neighbour joining (MEGA7.01[33]) according to the methods of Asaf et al [39,45]. For Bayesian posterior probabilities (PP) in the BI analyses, the best substitution model GTR + G model was tested according to the Akaike information criterion (AIC) by jModelTest verion 2102. The Markov Chain Monto Carlo (MCMC) was run for 1,000,000 generations with 4 incrementally heated chains, starting from random trees and sampling 1 out of every 100 generations. The first 30% of trees were discarded as burn-in to estimate the value of posterior probabilities. Furthermore, parameters for the ML analysis were optimized with a BIONJ tree as the starting tree with 1000 bootstrap replicates using the Kimura 2-parameter model with gamma-distributed rate heterogeneity and invariant sites. MP was run using a heuristic search with 1000 random addition sequence replicates with the tree-bisection-reconnection (TBR) branch-swapping tree search criterion. In the second phylogenetic analysis, 72 shared genes from the cp genomes of the twenty-six members of order Sapindales, were aligned using ClustalX with default settings, followed by manual adjustment to preserve reading frames. Similarly, the above4 mentioned phylogenetic inference models were utilized to build trees using 72 concatenated genes, using the same setting as described above and suggested by Asaf et al [45].

Result and discussion

Genome features, content and organization

The chloroplast genomes of *C. gileadensis* (MH042752) and *C. foliacea* (MH041484) were identical to typical angiosperms genomes of 160,268 bp and 160,249 bp, respectively (Fig 1).

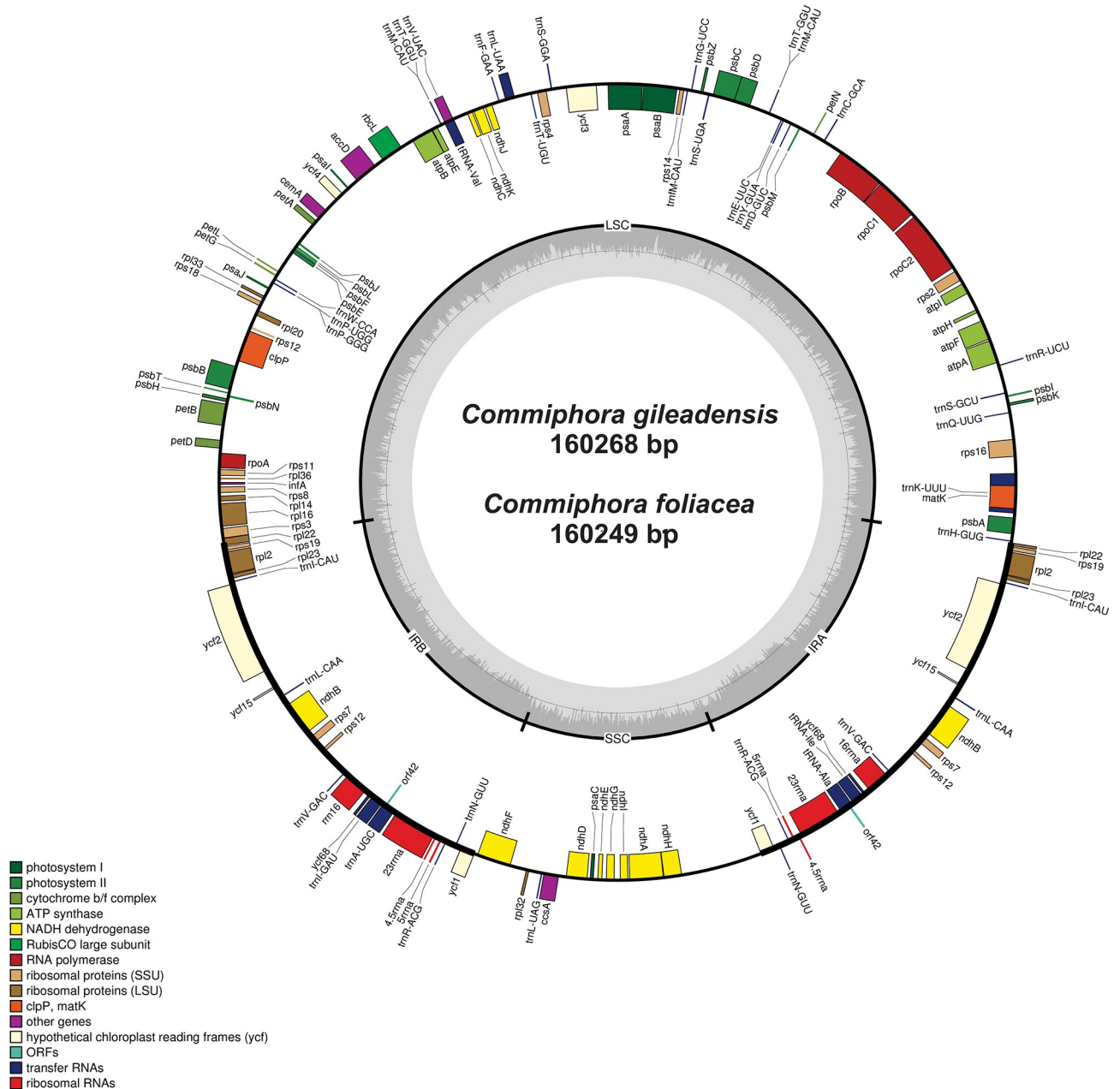


Fig 1. Genome map of the *C. gileadensis* and *C. foliacea* cp genomes. Thick lines indicate the extent of the inverted repeat regions (IRA and IRb), which separate the genome into small (SSC) and large (LSC) single copy regions. Genes drawn inside the circle are transcribed clockwise, and those outside are transcribed counter clockwise. Genes belonging to different functional groups are color-coded. The dark grey in the inner circle corresponds to the GC content and the light grey corresponds to the AT content.

<https://doi.org/10.1371/journal.pone.0208511.g001>

The size of these cp genomes were almost similar with previously reported chloroplast genome of *B. sacra* (160,543 bp) [46], *Azadirachta indica* (160,737 bp) [47], *Citrus sinensis* (160,129 bp) [48] and *Ailanthus altissima* [49], which belong to order Sapindales. Both of these genomes possess the quadripartite structures comprises a pair of inverted repeats (IRA and IRb) separated by small single copy region (SSC) and large single copy region (LSC). The LSC regions in these genomes varies from 87,885 bp to 88,054 bp, SSC varies from 18,746bp to 18,962bp, and

Table 1. Summary of complete *Commiphora gileadensis* and *C. foliacea* chloroplast genomes.

Parameters	<i>C. gileadensis</i>	<i>C. foliacea</i>	<i>C. wightii</i>	<i>B. sacra</i>
Size (bp)	160268	160249	156064	160543
Overall GC contents (%)	37.9	37.8	38	37.6
LSC size in bp	87885	87923	-	88054
SSC size in bp	18769	18746	-	18962
IR size in bp	26807	26790	-	26763
Protein coding regions size in bp	78238	78119	60006	80361
tRNA size in bp	2934	2934	2919	2195
rRNA size in bp	9050	9050	9394	9050
Number of genes	142	140	123	142
Number of protein coding genes	95	93	86	93
Number of rRNA	8	8	8	8
Number of tRNA	39	39	29	39

<https://doi.org/10.1371/journal.pone.0208511.t001>

the inverted repeat region varies from 26,763bp to 26,807bp (Fig 1). Similarly, the length of LSC, SSC and IR regions was also similar with previously reported genomes for order Sapindales [48,49].

Furthermore, the average GC content of *C. gileadensis* and *C. foliacea* genomes were found 37.8% which is almost similar to *B. sacra* (37.8%) and *C. wightii* (38%). The GC content of these cp genomes were also found similar with previously reported *Sesamum indicum* L. which is approximately 38% [50]. The AT content of both the cp genomes were 62.2%. This is in correlation to the other species from order Sapindales, for example *A. miaotaiense* (62.12%) [51], *A. davidii* (62.10%) [52], *C. sinensis* (61.52%) [48] and *P. amurense* (61.60%) [53]. Overall, the A+T content of 62.14% in both the cp genomes are closely related to order Sapindales (Table 1).

The GC content was unevenly present in the *C. gileadensis* and *C. foliacea* cp genomes where it was low (32.3 and 32.4%, respectively) in the SSC regions, high (42.9%) in IR regions and moderate (35.8%) in the LSC regions. In synergy to the previously published reports on cp genomes, the presence of ribosomal RNA (rRNA) sequences enhance the GC contents in the IR regions [54–56]. In addition, about 43.72% of *C. gileadensis* and 46.91% of *C. foliacea* cp genomes were found noncoding. In case of coding regions, the protein coding genes were 48.81 and 45.62%, tRNA genes were 1.83 and 1.83%, and rRNA genes were 5.64 and 5.64% found in the *C. gileadensis* and *C. foliacea* cp genomes, respectively.

The total coding DNA sequences (CDSs) of *C. gileadensis* and *C. foliacea* were 78,238 bp and 73,119bp in size which encodes 94 and 93 genes respectively (S1 Table). This also includes 26,078 and 24,273bp codons respectively (S2 Table). Similarly, the codon-usage frequency of the both *C. gileadensis* and *C. foliacea* cp genomes were determined on the basis of protein-coding and tRNA-related gene sequence (S3 Table, S4 Table). Like previously reported cp genomes, the cysteine (1.2%) and leucine (10.3%) were the least and most commonly encoded amino acids [39,54]. Furthermore, The AT contents of both *C. gileadensis* and *C. foliacea* cp genomes at the 1st, 2nd, and 3rd codon position of CDS were 54.6 and 55.1%, 61.4 and 58.4%, and 65.99 and 67.3%, respectively (S2 Table <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0182281> - pone-0182281-t003). This is in correlation with previous reports showing that the terrestrial plant's cp genome with highest AT-content at the 3rd codon-position [54,57]

The total number of genes in the *C. gileadensis* and *C. foliacea* were 140 and 141 respectively, in which 94 and 93 genes were protein coding genes, while 39 were tRNAs and 8 were

rRNAs genes. Similar results were reported in previous reported cp genomes of *B. sacra* has 142 genes [46], *A. miaotaiense* has 137 [51], *A. wangii* has 135 [58], *A. buergerianum* has 134 [59], and in *Meliaceae species* has 112 genes [60], which is from the same order Sapindales [51]. *Camellia species* contains 146 genes [60]. The protein-coding genes present in *C. gileadensis* and *C. foliacea* cp genomes include twelve genes-encoding small-ribosomal proteins (*rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps16, rps18, rps19*), 9 genes-encoding large ribosomal proteins (*rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36*), 10 genes of photo-system-II, five genes-encoding photosystem-I components, and 6 genes (*atpA, atpB, atpE, atpF, atpH, atpI*) ATP-synthase and electron-transport chain components (S1 Table). Similarly, the chloroplast genomes of *C. gileadensis* and *C. foliacea* contains introns containing genes. There were 11 genes containing intron inclusive of nine which have single-intron and 3 (*clpP, ycf3* and *rps12*) which have two introns (Table 2). These results are similar with previously reported cp genome of angiosperms. The smallest intron in both *C. gileadensis* and *C. foliacea* cp genomes were 518bp and 526 bp respectively, whereas the longest intron was determined in *trnK-UUU* (2507 bp) in both cp genomes that included the entire *matK* gene. Introns can be a useful tool for successful transformational effectiveness and play a vital role in the regulation of gene expression [61]. Like other angiosperms cp genomes, *rps12* gene was unequally distributed, with single copy of its 3' exon/intron, located at the IR regions and 5'-exon, located in the LSC region. A similar correlation in the results were observed in previously reported cp genomes of *C. platymamma* [62], *C. aurantiifolia* [63] and *Dipteronia species* [62]. Moreover, there are 4 ribosomal RNA genes and 30 transfer RNA genes. The *infA gene*, which code for transcription factor of initiation was present in both *Commiphora species*, while it is absent in *Citrus sinensis* (L.) cp genome [64].

Expansion and contraction of IRs

Expansion and contraction of the IR (a&b) repeats were compared among different species belonging to order Sapindales. The chloroplast genomes of angiosperm are highly conserved,

Table 2. The genes with introns in the *C. gileadensis* (*C. g*) and *C. foliacea* (*C. f*) chloroplast genome and the length of exons and introns.

Gene	Location	Exon I (bp)		Intron I (bp)		Exon II (bp)		Intron II (bp)		Exon III (bp)	
		<i>C. g</i>	<i>C. f</i>	<i>C. g</i>	<i>C. f</i>	<i>C. g</i>	<i>C. f</i>	<i>C. g</i>	<i>C. f</i>	<i>C. g</i>	<i>C. f</i>
<i>atpF</i>	LSC	159	159	719	718	438	438				
<i>petB</i>	LSC	6	6	753	749	648	645				
<i>rpl2*</i>		393	393	662	662	432	435				
<i>rpl16</i>	LSC	9	9	1088	1086	411	411				
<i>rps16</i>	LSC	48	48	871	864	225	225				
<i>rpoC1</i>	LSC	432	435	791	791	1620	1620				
<i>rps12*</i>		114	114	-	-	243	243				
<i>clpP</i>	LSC	69	69	873	862	267	267	663	656	218	228
<i>ndhA</i>		553	553	1109	1104	539	534				
<i>ndhB*</i>		777	777	681	681	753	756				
<i>ycf3</i>	LSC	129	129	719	716	228	228	768	768	150	150
<i>trnA-UGC*</i>	IR	38	38	805	803	35	35				
<i>trnI-GAU*</i>	IR	42	42	956	957	35	35				
<i>trnL-UAA</i>	LSC	37	37	518	526	50	50				
<i>trnK-UUU</i>	LSC	37	37	2507	2507	35	35				

C. g = *C. gileadensis*, *C. f* = *C. foliacea*

<https://doi.org/10.1371/journal.pone.0208511.t002>

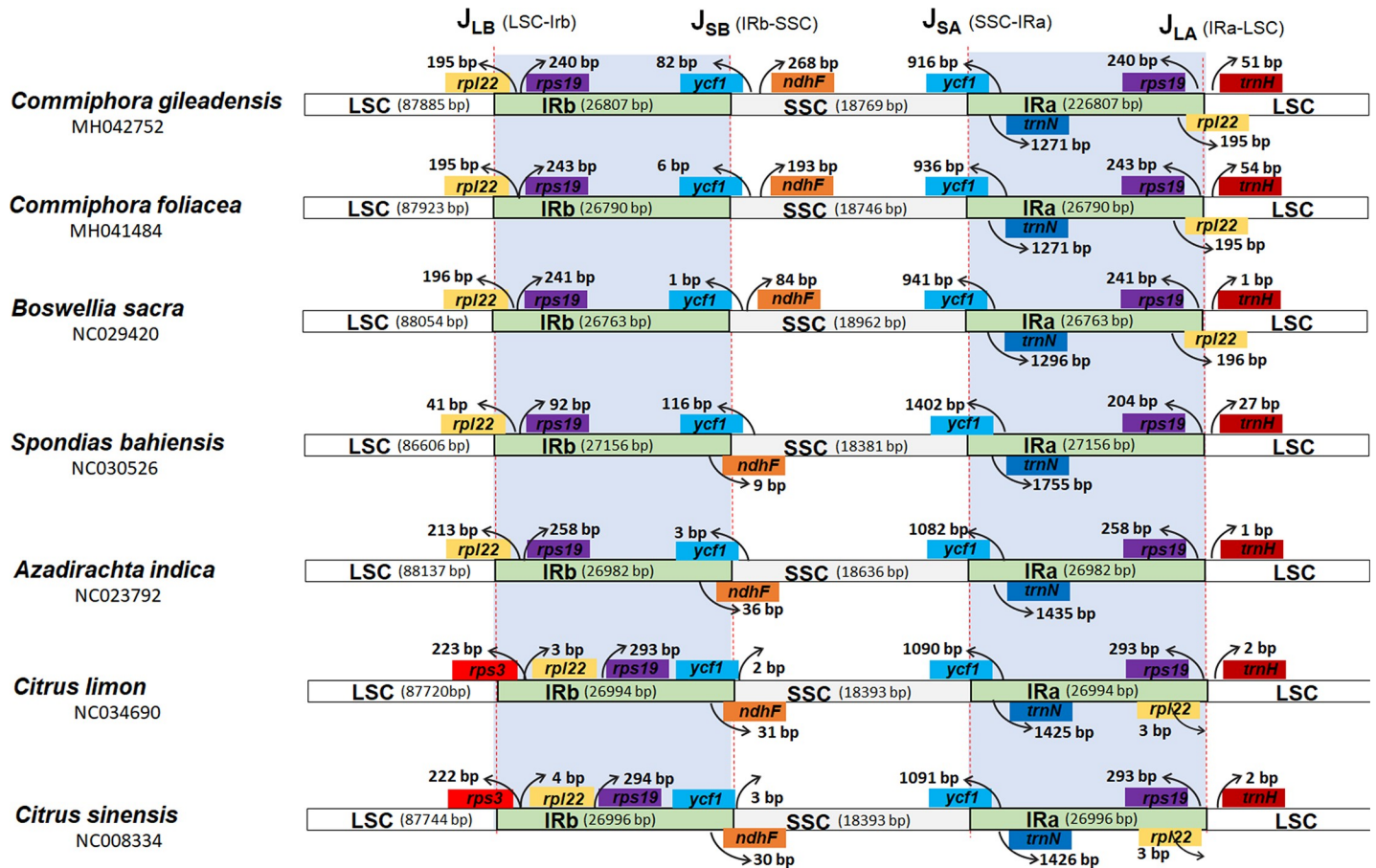


Fig 2. Distance between adjacent genes and junctions of the small single-copy (SSC), large single-copy (LSC), and two inverted repeat (IR) regions among seven plastid genomes within order Sapindales. Boxes above and below the main line indicate the adjacent border genes. The figure is not to scale regarding sequence length, and only shows relative changes at or near the IR/SC borders.

<https://doi.org/10.1371/journal.pone.0208511.g002>

but there is still some variation due to contraction or expansion of SSC and IR boundary region [64]. Due to these contraction and expansion, the size variation and rearrangement occurs in the LSC/SSC/IRA/IRB [64]. In this study we carried out a detail comparison of 4-junctions (JLA, JLB, JSA, and JSB) between LSC and SSC regions and both the IRa and IRb regions of the *C. gileadensis* and *C. foliacea* species and five other species from order Sapindales were performed (Fig 2). Despite the similar IR regions lengths of *C. gileadensis* and *C. foliacea* with other related species, some contraction and expansion were determined with the IR regions ranging from 26,763 bp in *B. sacra* to 27,156 bp in *Spondias bahiensis*. The genes present at starts and end of IR-regions were partly repeated, including 195 bp of *rpl22* in both *C. gileadensis* and *C. foliacea*, 196 bp in *B. sacra*, 4bp and 213 bp in *S. bahiensis* and *A. indica* respectively. However, in *Citrus lemon* and *Citrus sinensis* the duplicated gene was *rps3* which is located 223 and 222 bp in inverted repeat region from JLB (Fig 2). Correspondingly, the *ycf1* gene which is considered as a hypothetical is duplicated partially, 916 bp and 936 bp in *C. gileadensis* and *C. foliacea*, 941bp, 1402 bp, 1082bp, 1090 bp and 1091 bp in *B. sacra*, *S. bahiensis*, *A. indica*, *C. lemon* and *C. sinensis* respectively. J_{LA} is positioned between *trnH* and *rps19*, whereas the deviation in gaps between J_{LA} and *rps19* range from 240 to 293 bp throughout compared species. Similarly, the detachment in *C. gileadensis* and *C. foliacea* was 240 bp and 243 bp correspondingly. The distance between *trnH* and J_{LA} was 51 bp and 54 bp in *C. gileadensis* and *C.*

foliacea, which is 1 bp in *B. sacra* and *A. indica*. Furthermore, variation was observed in the location of *ndhF* genes which is present at 268 bp, 193 bp and 84 bp away from J_{SB} in SSC regions in *C. gileadensis*, *C. foliacea* and *B. sacra* cp genomes. However, in other four species cp genomes *ndhF* was located at the junction of IRb-SSC. Furthermore, there is 76 bp variation was observed in location of *ycf1* gene at J_{SB} border in both *C. gileadensis* and *C. foliacea*. However, in *B. sacra* cp genome this distance was calculated 1 bp away from J_{SB} border [46]. Similar to previously reported cp genome from Sapindales these cp genomes having well-maintained genomic structure in term of cp genome length, IR regions, gene order and gene numbers [49]. However, some of the deviation in sequence might be due to the result of boundary contraction and expansion between the boundaries of IR and single copy regions among different plant species as reported by Wang et al. [64].

Structural variation in genomic regions

In order to determine the sequence divergence among the four chloroplast genomes viz. *C. gileadensis*, *C. foliacea*, *C. wightii* and *B. sacra*, the annotation of *C. gileadensis* cp genome was used as a reference for determination of the sequence similarity in the cp genomes of the three species through mVISTA program (Fig 3). The results showed that high degree of synteny and comparatively lower sequence similarity were noted among these cp genome of these four species especially in *rpoC2*, *rpoB*, *petB*, *psaB*, *ndhB*, *ndhF*, *ccsA*, *ycf1*, *ycf2*, *rpl22* and *atpF* genes (Fig 3). Furthermore, like previous reported genomes the LSC and SSC regions were more divergent as compared to IR regions in the compared species and less similarity in the coding region were observed. Similarly, various deviating regions included *matK*, *ycf3-psaA*, *clpP*, *accD*, *atpF*, *rpoC1*, *petA-psbI*, *ycf1-rps15*, *rps19* and *ndhF* were reported previously in various cp genomes [54,56]. Differences in the coding regions were similar in this study to the previously analyzed cp genome by Kumar et al. [33]. Similarly, for the shared genes the average pairwise sequence differentiation was calculated among these four species (Fig 3 and S9 Table). The results revealed that the 13 most divergent genes among these genomes were *infA*, *rps8*, *rpl32*, *rpl22*, *rpl16*, *psaI*, *ndhH*, *ndhG*, *matK*, *ccsA*, *atpH*, *accD* and *psbN*. The *rpl22* gene showed the greatest average sequence divergence (0.029), after that *rps3* (0.028), *ndhH* (0.027), and *ccsA* (0.020), majority of these were located in the LSC region. Similar results were observed in previously reported angiosperm cp genomes [56,65]. Furthermore, comparison of the cp genome of *C. gileadensis* with *C. foliacea*, *C. wightii* and *B. sacra* revealed 3,032, 8,787 and 5,120 SNPs as well as 3,580, 10,460 and 17,122 *Indels* respectively (Fig 4). Similarly, the *C. foliacea* cp genome also showed 8,194 and 5,182 SNPs while 7,632 and 17,970 *Indel* with *C. wightii* and *B. sacra* respectively. These Results shows that even the most conserved genome possesses some interspecific mutations which provides an important information in analyzing the phylogenetic and genetic diversity among the species [56].

SSR Polymorphism in the cp Plastomes

Diversity exist in the copies of SSRs present in the chloroplast genome and these SSRs are vital molecular markers in the plant evolutionary, population genetics and studying the ecology of the plants [66]. In the present study, we detected complete SSRs in *C. gileadensis*, *C. foliacea* cp genomes together with *C. wightii* and *B. sacra* (Fig 5) and detail SSR analysis of *C. gileadensis*, *C. foliacea*, *C. wightii* and *B. sacra* were also performed (S5 Table, S6 Table, S7 Table, S8 Table). Specific parameters were set for the SSRs present in genome because SSR of more than 10bp are liable to slip strand mispairing, which is considered to be the basic reason for SSR polymorphism. [67–69]. The results reveled a total of 196, 175, 153 and 191 SSRs in the *C. gileadensis*, *C. foliacea*, *C. wightii* and *B. sacra* cp genomes, respectively. The majority of SSRs

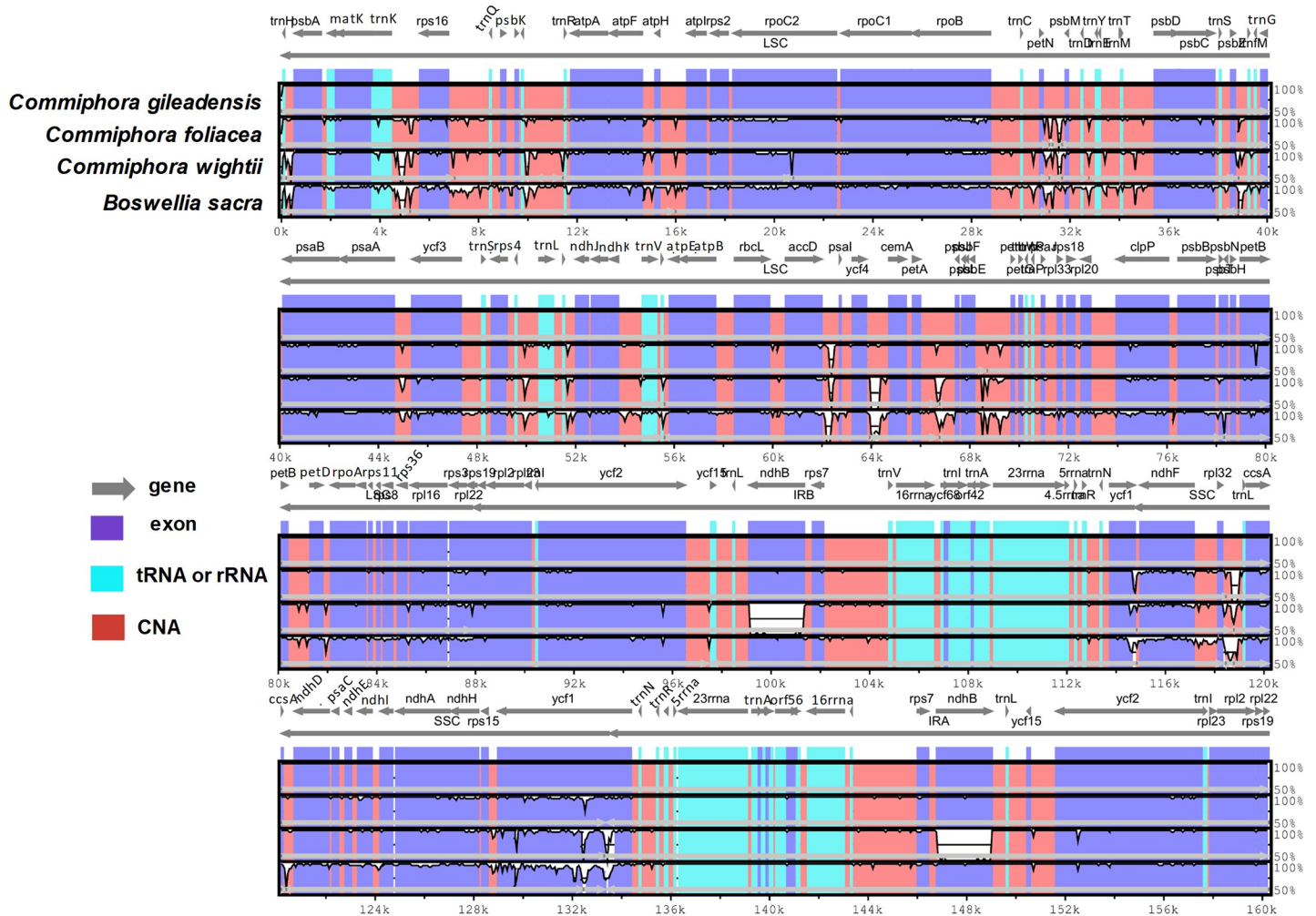


Fig 3. Visual alignment of plastid genomes from *C. gileadensis* and *C. foliacea* with previously reported *C. wightii* and *B. sacra*. VISTA-based identity plot showing sequence identity among seven species, using *C. gileadensis* as a reference.

<https://doi.org/10.1371/journal.pone.0208511.g003>

75 (38.2%) in *C. gileadensis* cp genome was mono-nucleotide repeat motifs. However, in other three cp genome the majority of SSRs were tri nucleotides motif, varying from quantity from 71 (40.57%) in *C. foliacea* to 75 (39.26%) in *B. sacra*. Tri-nucleotide repeat motif was found the second most common 69 (35.2%) in *C. gileadensis*. Using our search criterion, 3, 2 and 2 penta nucleotide were detected in *C. gileadensis*, *C. foliacea* and *C. wightii* cp genome respectively. However, in hexa nucleotide was only detected in *B. sacra* cp genome. Furthermore, in *C. gileadensis* and *C. foliacea*, most common mononucleotide SSRs are A (93.33% and 94.1%) motif, respectively. Approximately, 52% and 67.3% of SSRs are sited in non-coding regions, 2.04% and 5.71% are located in rRNA sequences in both *C. gileadensis* and *C. foliacea* respectively. These results suggest that SSRs are irregularly disseminated in the chloroplast genome and provides valuable information to select the effective molecular markers for spotting inter and intra specific polymorphisms [70–72]. The abundance of ‘A’ and ‘T’ nucleotide in the cp genomes as compared to ‘G’ and ‘C’ is due to the fact that mono and dinucleotide is only consist of ‘A’ and ‘T’ nucleotide which contributes to the bias in the cp genome base composition [68]. The finding from these *Commiphora* genomes reveals that SSRs in the cp genomes are

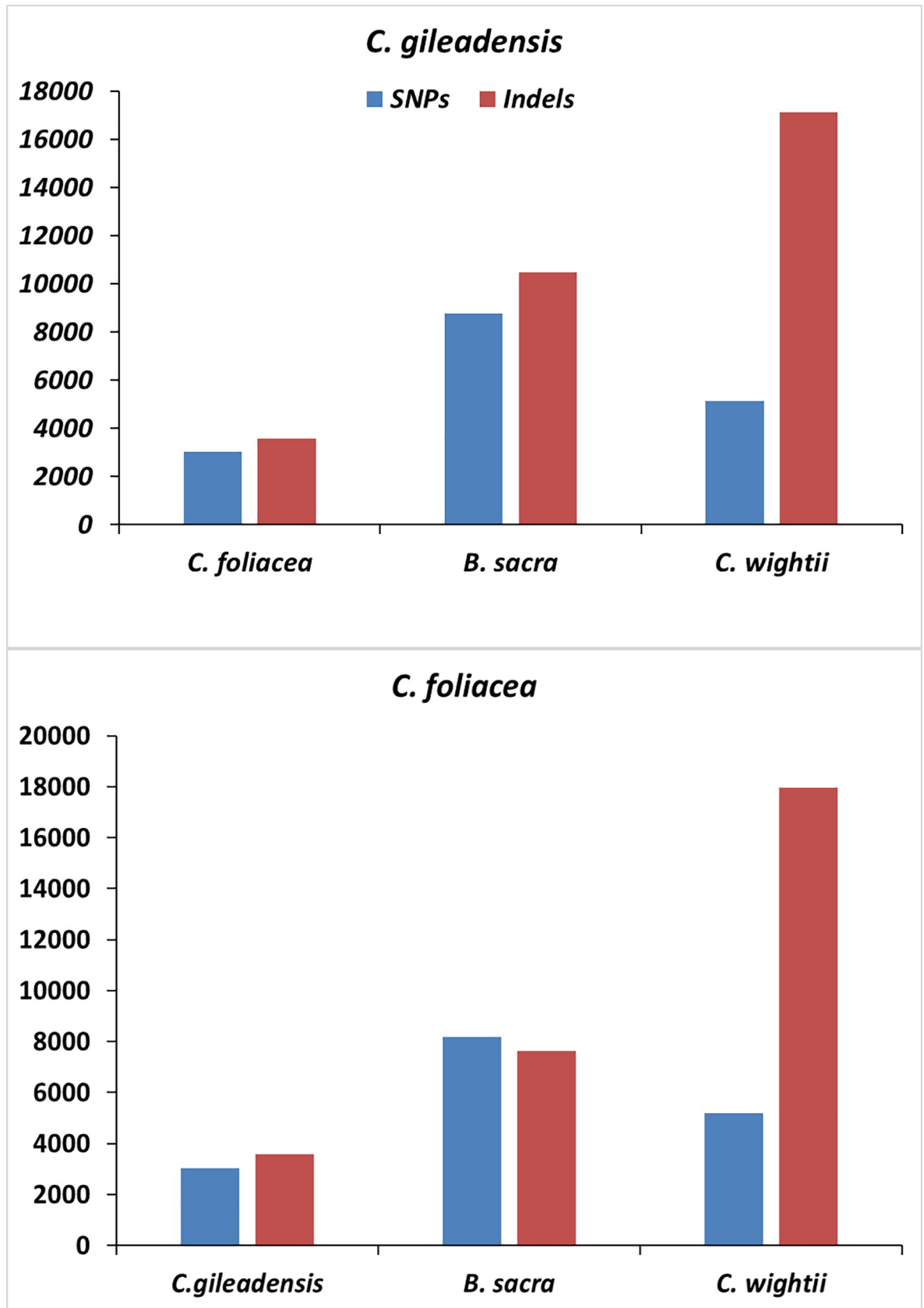


Fig 4. Single nucleotide polymorphism (SNP) and insertion and deletion in *Commiphora* species with related species *Commiphora wightii* and *Boswellia Sacra*.

<https://doi.org/10.1371/journal.pone.0208511.g004>

normally composed of polyadenine (polyA) or polythymine (polyT) repeats and irregularly contains the tandems guanine (G) or cytosine (C) repeats [73], which is similar to the previous results thus a possible reason for AT richness [46,55,56]. The presence of SSRs in cp genomes will give useful information for primer designing used for phylogeography and population structure at specie level or SSRs can also be used for obtaining useful and important information used for phylogenetic relationship and population genetics [74]. Previously reported *D. viscoa* contains 249 SSRs, having the mononucleotide SSRs in highest number followed by tri nucleotide repeats [74,75]. The cp genome of *globe Artichoke* contains 127 repeats is lesser than our findings [76].

Repeats analysis of *Commiphora* plastomes

Repetitive sequences in the plastomes plays role in the rearrangement of genomes which provide an important information about phylogenetic studies [50,77] From the previously analyzed cp genomes it is evident that for the induction of indels and substitutions these repeat sequence is essential. Additionally, analysis of different cp-genomes exposed that repeat sequence is important to produce indels/substitutions [78]. Similarly, in our study repeat analysis of the *C. gileadensis* and *C. foliacea* identified 30 and 25 palindromic repeat, 15 and 25 forward, 20 and 25 tandem repeat respectively. Similarly, 21 and 20 palindromic repeats, 27 and 20 tandem repeats were spotted in *C. wightii* and *B. sacra* respectively. However, in *C. wightii* only 6 forward repeats were detected while in *B. sacra* it was 29 in number. Overall 65 and 75 repeats of different length were found in both *C. gileadensis* and *C. foliacea*, respectively. In *C. gileadensis* four palindromic repeats were 75-89bp and 21 repeats were > 90 length. However, in *C. foliacea* the number of >90 repeats were less and only 2 palindromic repeats were found. On the other hand, among the forward repeats 10 repeats of >90 bp were detected in both *C. gileadensis* and *C. foliacea* cp genome (Fig 6). Earlier reports recommend that deviation in sequences and genome arrangement occur due to the slipped-strand mispairing and inappropriate recombination of repetitive sequences [77,79]. Moreover, the occurrence of the repeats shows that this locus is a key hot-spot for re-configuration of the genome [50,80]. Also, the Information from these repeats are a source of valuable information for constructing genetic markers for population studies and phylogenetic analysis [50].

Phylogenomic analysis

Several aspects of *Commiphora* natural history have impeded efforts to resolve its species-level taxonomy and investigate its systematic biology [81]. Previously, the two species have examined species-level phylogenetic relationships in *Commiphora* and tested the monophyly of some of these infrageneric taxonomic groups [5,82]. Gostel et al. [83] reconstruct phylogenetic relationship in *Commiphora* species using genes from nuclear as well as from chloroplast genome. However, hypothesis regarding higher level relationship among *Commiphora* specie are similarly unresolved [83]. To resolve the phylogenetic relationship among different species, the complete chloroplast genome sequencing provides more detailed information about the phylogenetics [84,85]. Therefore, in this study the phylogenetic position of both *C. gileadensis* and *C. foliacea* within order Sapindales was established by analyzing the complete cp genomes (Fig 7 and S1) and 72 shared genes (form all twenty-six species). Phylogenetic analysis using MP, BI, NJ and ML methods were performed. The results revealed that both complete cp

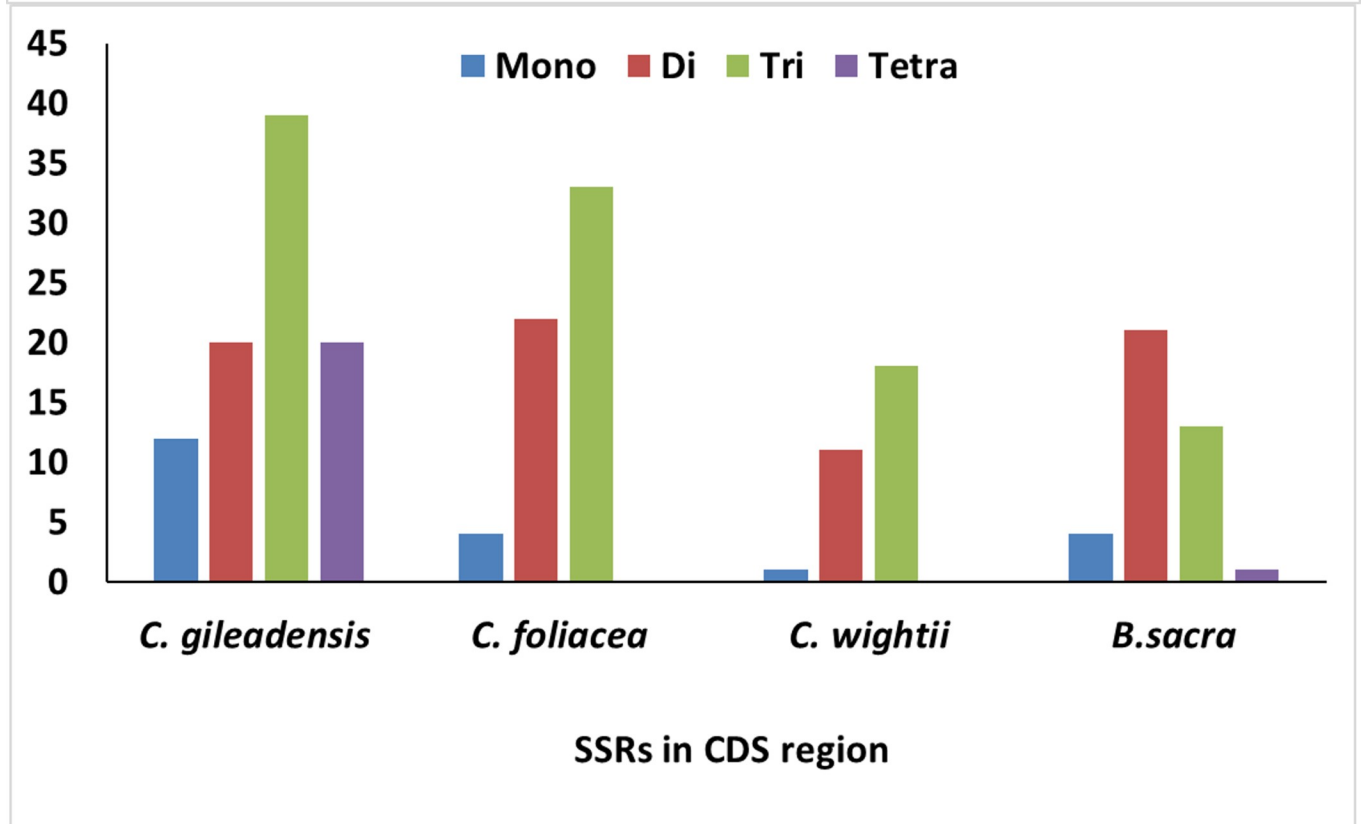
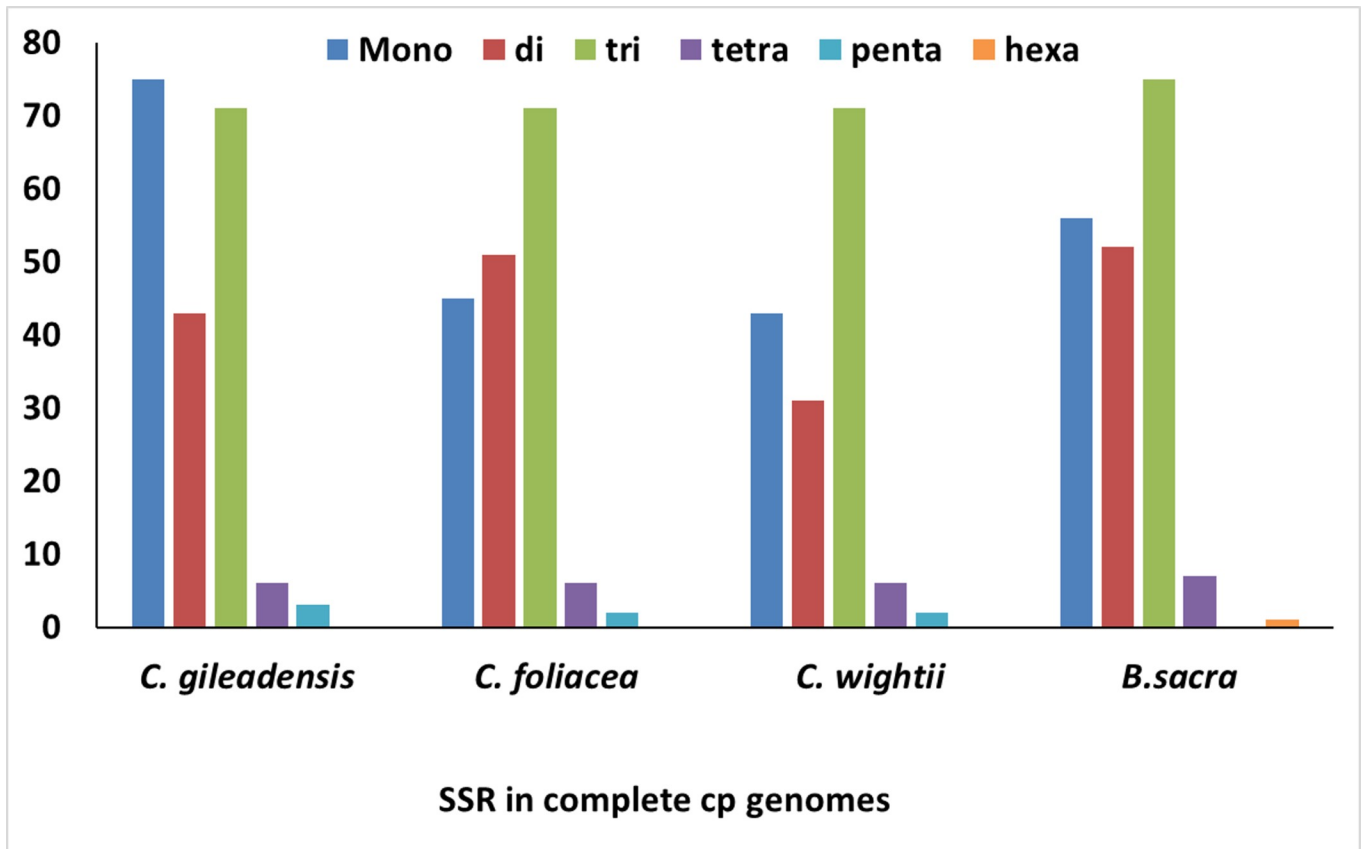


Fig 5. Analysis of simple sequence repeat (SSR) in the *C. gileadensis* and *C. foliacea* plastid genome. A, Number of SSR types in complete genome, coding, and non-coding regions; B, Frequency of identified SSR motifs in different repeat class types.

<https://doi.org/10.1371/journal.pone.0208511.g005>

genomes and 72 shared genes of *C. gileadensis* and *C. foliacea* contain the same phylogenetic signals and generated phylogenetic trees with identical topologies (Fig 6, S1 Fig). The results show that both *C. gileadensis* and *C. foliacea* form a single clade with previously reported *C. wightii* and *B. sacra* from family *Burseraceae* with high BI and bootstrap support values (Fig 7, S1 Fig). The tree topology showed that these four species from family *Burseraceae* are more closely related to *Spondias* species from Family *Anacardiaceae* and *Azadirachta indica* from *Meliaceae* (Fig 7, S1 Fig). Furthermore, the phylogenetic analysis validated the relationship inferred from the phylogenetic work reported by Saina et al. [86] that the families *Burseraceae* and *Anacardiaceae* formed a sister group/clade, which further branched forming sister clade with *Meliaceae*, *Rutaceae*, *Simaroubaceae* and *Sapindaceae* families. Therefore, for future phylogenetic studies must incorporate additional species for better understanding of *Commiphora* species evolution and phylogeny. This study offers a basis for future phylogenetic of family *Burseraceae*.

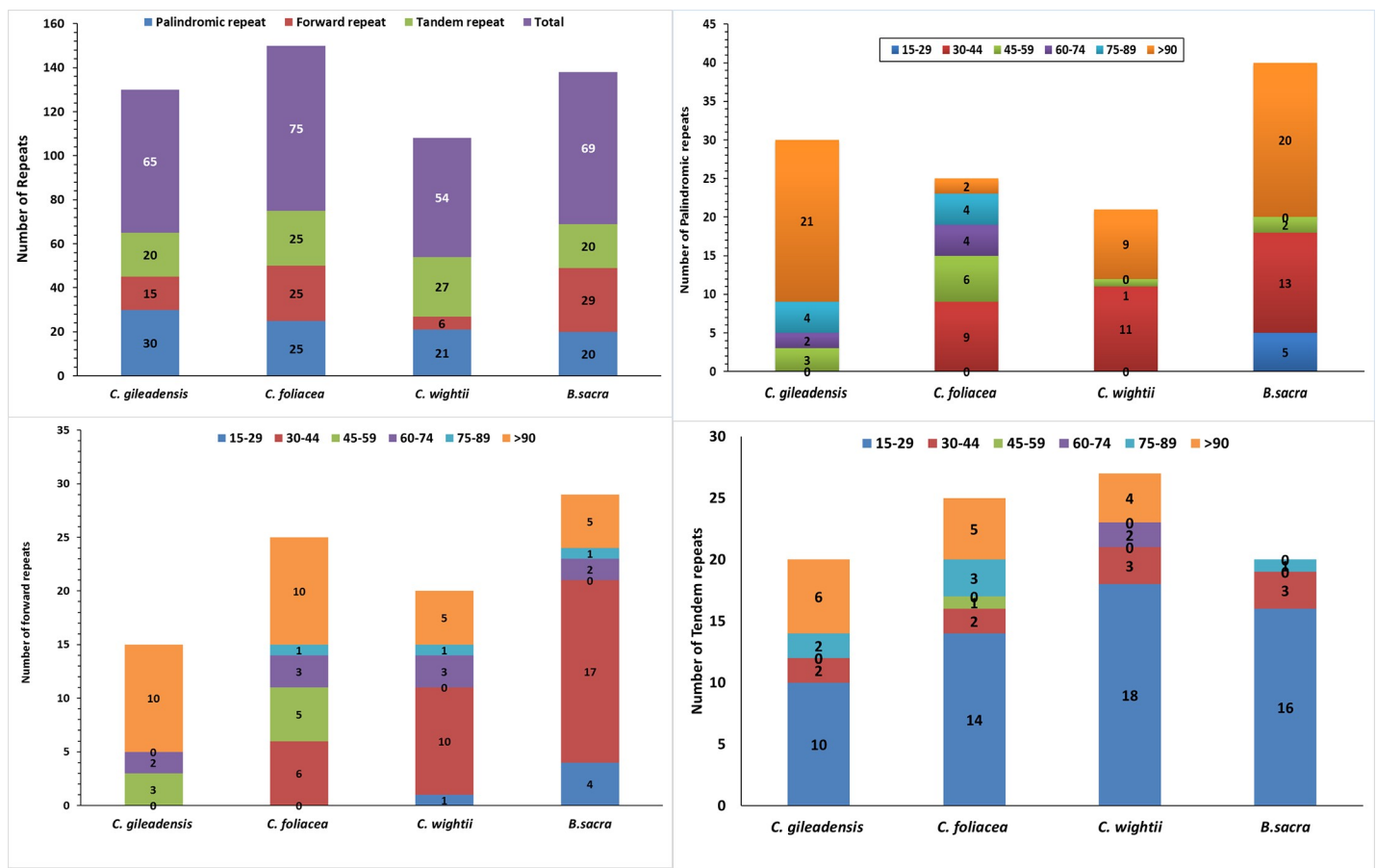


Fig 6. Analysis of repeated sequences in *C. gileadensis* and *C. foliacea* plastid genome. Totals of three repeat types, Frequency of palindromic repeats by length, Frequency of forward repeats by length and Frequency of tandem repeats by length.

<https://doi.org/10.1371/journal.pone.0208511.g006>

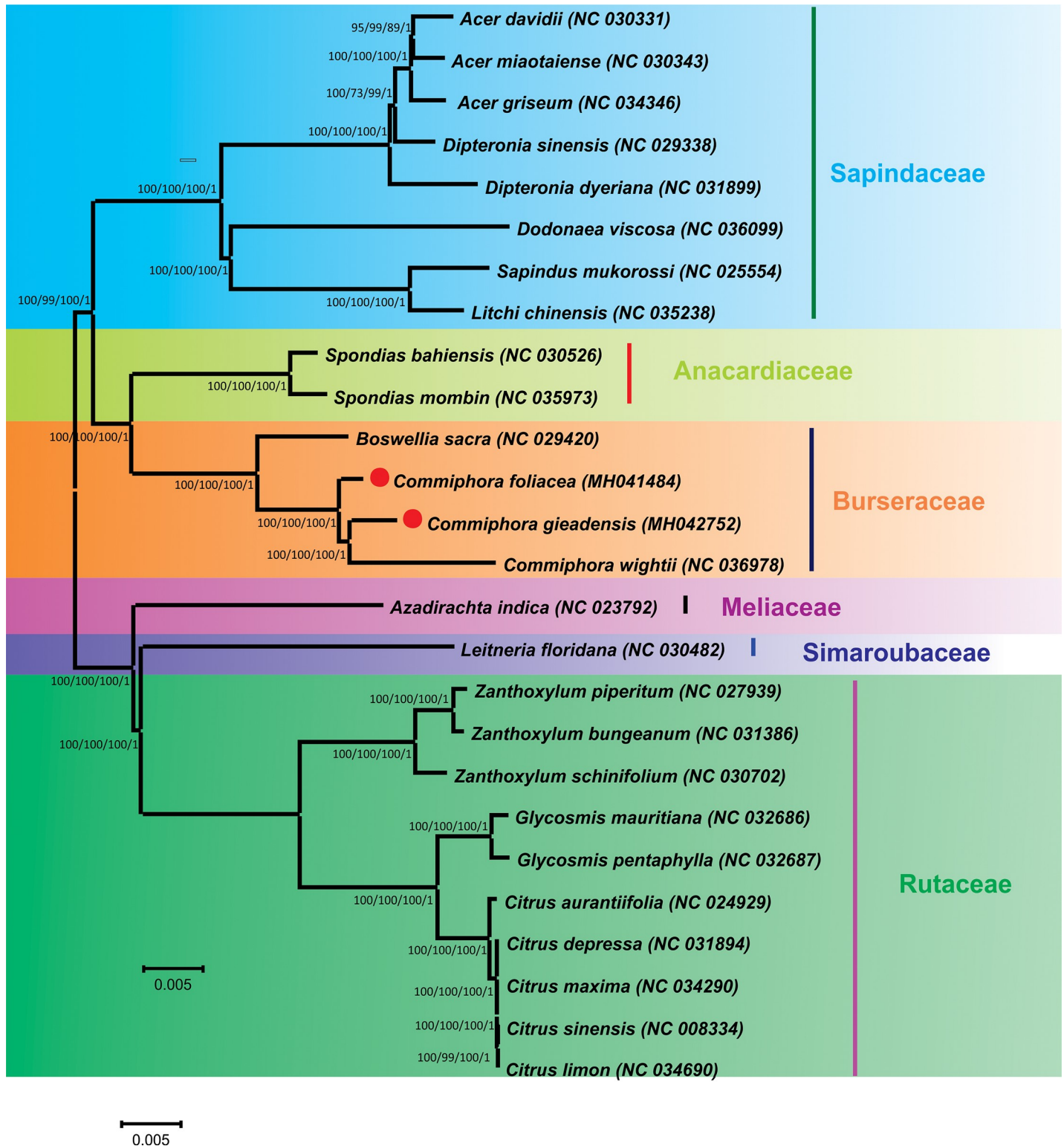


Fig 7. Phylogenetic trees of *C. gileadensis* and *C. foliacea* within order Sapindales. The entire genome dataset was analyzed using four different methods: Bayesian inference (BI), maximum parsimony (MP), maximum likelihood (ML), and neighbor-joining (NJ). Numbers above the branches represent bootstrap values in the MP, ML, and NJ trees and posterior probabilities in the BI trees, whereas the number below the branches represents branch length. The red dot represents the position of *C. gileadensis* and *C. foliacea*.

<https://doi.org/10.1371/journal.pone.0208511.g007>

Supporting information

S1 Fig. Phylogenetic trees of *C. gileadensis* and *C. foliacea* within order Sapindales. The 72 shared gene dataset was analyzed using four different methods: Bayesian inference (BI), maximum parsimony (MP), maximum likelihood (ML), and neighbor-joining (NJ). Numbers above the branches represent bootstrap values in the MP, ML, and NJ trees and posterior probabilities in the BI trees, whereas the number below the branches represents branch length. The red dot represents the position of *C. gileadensis* and *C. foliacea*.
(TIF)

S1 Table. Genes in the sequenced *C. gileadensis* and *C. foliacea* chloroplast genome.
(DOCX)

S2 Table. Base compositions in *C. gileadensis* (*C. g*), *C. foliacea* (*C. f*), *C. wightii* (*C. w*) and *B. sacra* (*B. s*) cp genomes.
(DOCX)

S3 Table. The codon–anticodon recognition pattern and codon usage for the *C. gileadensis* chloroplast genome.
(DOCX)

S4 Table. The codon–anticodon recognition pattern and codon usage for the *C. foliacea* chloroplast genome.
(DOCX)

S5 Table. Simple sequence repeats (SSRs) in the *C. wightii* chloroplast genome.
(DOCX)

S6 Table. Simple sequence repeats (SSRs) in *C. gileadensis* chloroplast genome.
(DOCX)

S7 Table. Simple sequence repeats (SSRs) in *C. foliacea* chloroplast genome.
(DOCX)

S8 Table. Simple sequence repeats (SSRs) in *Boswellia sacra* chloroplast genome.
(DOCX)

S9 Table. Pairwise distance of 77 shared genes in three *Commiphora* species (*Commiphora gileadensis*, *Commiphora foliacea*, *Commiphora wightii*) with *Boswellia sacra*.
(XLS)

Author Contributions

Data curation: Sajjad Asaf.

Formal analysis: Arif Khan, Sajjad Asaf.

Funding acquisition: Abdul Latif Khan, Nadiya Alsaady.

Methodology: Sajjad Asaf, Omar Al-Sudairy, Noor Mazin AbdulKareem, Adil Khan.

Project administration: Abdul Latif Khan, Ahmed Al-Harrasi.

Software: Sajjad Asaf.

Supervision: Abdul Latif Khan, Ahmed Al-Harrasi, Ahmed Al-Rawahi.

Visualization: Zabta Khan Shinwari.

Writing – original draft: Arif Khan.

Writing – review & editing: Abdul Latif Khan, Tariq Shehzad, Ali Al-Lawati, Ahmed Al-Rawahi.

References

1. Miller AG, Morris M (1988) Plants of Dhofar: the southern region of Oman, traditional, economic and medicinal uses. Oman: Office of the Adviser for Conservation of the Environment, Diwan of Royal Court Sultanate of Oman xxvii, 361p-col illus ISBN 715708082.
2. Thulin M, Beier BA, Razafimandimbison SG, Banks HI (2008) *Ambilobeia*, a new genus from Madagascar, the position of *Aucoumea*, and comments on the tribal classification of the frankincense and myrrh family (Burseraceae). *Nordic Journal of Botany* 26: 218–229.
3. Langenheim JH (2003) Plant resins: chemistry, evolution, ecology, and ethnobotany: Timber Press.
4. Shen T, Li G-H, Wang X-N, Lou H-X (2012) The genus *Commiphora*: a review of its traditional uses, phytochemistry and pharmacology. *Journal of ethnopharmacology* 142: 319–330. <https://doi.org/10.1016/j.jep.2012.05.025> PMID: 22626923
5. Weeks A, Simpson BB (2007) Molecular phylogenetic analysis of *Commiphora* (Burseraceae) yields insight on the evolution and historical biogeography of an “impossible” genus. *Molecular phylogenetics and evolution* 42: 62–79. <https://doi.org/10.1016/j.ympev.2006.06.015> PMID: 16904915
6. Mahr D (2012) *Commiphora*: An Introduction to the Genus: Part 1: Distribution, Taxonomy, and Biology. *Cactus and Succulent Journal* 84: 140–154.
7. Al-Harbi M, Qureshi S, Raza M, Ahmed M, Afzal M, et al. (1997) Gastric antiulcer and cytoprotective effect of *Commiphora molmol* in rats. *Journal of Ethnopharmacology* 55: 141–150. PMID: 9032627
8. Shen T, Li GH, Wang XN, Lou HX (2012) The genus *Commiphora*: a review of its traditional uses, phytochemistry and pharmacology. *J Ethnopharmacol* 142: 319–330. <https://doi.org/10.1016/j.jep.2012.05.025> PMID: 22626923
9. Mahr D (2012) *Commiphora*: An Introduction to the Genus. *Cactus and Succulent Journal* 84: 140–154.
10. Iluz D, Hoffman M, Gilboa-Garber N, Amar Z (2010) Medicinal properties of *Commiphora gileadensis*. *African Journal of Pharmacy and Pharmacology* 4: 516–520.
11. Groom N (1981) Frankincense and myrrh. A study of the Arabian incense trade. Longman: London & New York 285: 96–120.
12. Al-Sieni AI (2014) The antibacterial activity of traditionally used *Salvadora persica* L. (miswak) and *Commiphora gileadensis* (palsam) in Saudi Arabia. *African Journal of Traditional, Complementary and Alternative Medicines* 11: 23–27.
13. Thulin M (1999) Burseraceae, Flora of Somalia, Tiliaceae-Apiaceae (ed., by Thulin, M), 2: 183–228. Royal Botanic Gardens, Kew.
14. Eslamieh J (2011) *Commiphora gileadensis*. *Cactus and Succulent Journal* 83: 206–210.
15. Liu H-J, Ding C-H, He J, Cheng J, Pei LY, et al. (2018) Complete chloroplast genomes of *Archiclematis*, *Naravelia* and *Clematis* (Ranunculaceae), and their phylogenetic implications. *Phytotaxa* 343: 214–226.
16. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, et al. (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American journal of botany* 92: 142–166. <https://doi.org/10.3732/ajb.92.1.142> PMID: 21652394
17. Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, et al. (2014) Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *American Journal of Botany* 101: 1987–2004. <https://doi.org/10.3732/ajb.1400398> PMID: 25366863
18. Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. *American journal of botany*: 1205–1224.
19. Wicke S, Schneeweiss GM, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant molecular biology* 76: 273–297. <https://doi.org/10.1007/s11103-011-9762-4> PMID: 21424877
20. Jansen RK, Cai Z, Raubeson LA, Daniell H, Leebens-Mack J, et al. (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences* 104: 19369–19374.

21. Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences* 104: 19363–19368.
22. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, et al. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *The EMBO journal* 5: 2043–2049. PMID: [16453699](#)
23. Shi C, Hu N, Huang H, Gao J, Zhao Y-J, et al. (2012) An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. *Plos one* 7: e31468. <https://doi.org/10.1371/journal.pone.0031468> PMID: [22384027](#)
24. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.
25. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, et al. (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome research* 24: 1384–1395. <https://doi.org/10.1101/gr.170720.113> PMID: [24755901](#)
26. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* 9: 357. <https://doi.org/10.1038/nmeth.1923> PMID: [22388286](#)
27. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology* 19: 455–477. <https://doi.org/10.1089/cmb.2012.0021> PMID: [22506599](#)
28. Hahn C, Bachmann L, Chevreur B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic acids research* 41: e129–e129. <https://doi.org/10.1093/nar/gkt371> PMID: [23661685](#)
29. Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255. PMID: [15180927](#)
30. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199> PMID: [22543367](#)
31. Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic acids research* 33: W686–W689. <https://doi.org/10.1093/nar/gki366> PMID: [15980563](#)
32. Lohse M, Drechsel O, Bock R (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current genetics* 52: 267–274. <https://doi.org/10.1007/s00294-007-0161-y> PMID: [17957369](#)
33. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* 9: 299–306. <https://doi.org/10.1093/bib/bbn017> PMID: [18417537](#)
34. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic acids research* 32: W273–W279. <https://doi.org/10.1093/nar/gkh458> PMID: [15215394](#)
35. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, et al. (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic acids research* 29: 4633–4642. PMID: [11713313](#)
36. Kraemer L, Beszteri B, Gäbler-Schwarz S, Held C, Leese F, et al. (2009) S TAMP: Extensions to the S TADEN sequence analysis package for high throughput interactive microsatellite marker design. *BMC bioinformatics* 10: 41. <https://doi.org/10.1186/1471-2105-10-41> PMID: [19183437](#)
37. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* 27: 573. PMID: [9862982](#)
38. Liu H-Y, Yu Y, Deng Y-Q, Li J, Huang Z-X, et al. (2018) The Chloroplast Genome of *Lilium henrici*: Genome Structure and Comparative Analysis. *Molecules* 23: 1276.
39. Asaf S, Khan AL, Aaqil Khan M, Muhammad Imran Q, Kang S-M, et al. (2017) Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*) and nine other *Glycine* species. *PLOS ONE* 12: e0182281. <https://doi.org/10.1371/journal.pone.0182281> PMID: [28763486](#)
40. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30: 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: [23329690](#)
41. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution* 16: 111–120. PMID: [7463489](#)
42. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187> PMID: [19346325](#)

43. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574. PMID: [12912839](https://pubmed.ncbi.nlm.nih.gov/12912839/)
44. Swofford D (2002) PAUP*: phylogenetic analysis using parsimony (* and other methods). Sunderland, MA. Sinauer Associates.
45. Khan AL, Asaf S, Lee I-J, Al-Harrasi A, Al-Rawahi A (2018) First chloroplast genomics study of *Phoenix dactylifera* (var. *Naghal* and *Khanezi*): A comparative analysis. *PLOS ONE* 13: e0200104. <https://doi.org/10.1371/journal.pone.0200104> PMID: [30063732](https://pubmed.ncbi.nlm.nih.gov/30063732/)
46. Khan AL, Al-Harrasi A, Asaf S, Park CE, Park G-S, et al. (2017) The first chloroplast genome sequence of *Boswellia sacra*, a resin-producing plant in Oman. *PloS one* 12: e0169794. <https://doi.org/10.1371/journal.pone.0169794> PMID: [28085925](https://pubmed.ncbi.nlm.nih.gov/28085925/)
47. Krishnan NM, Jain P, Gupta S, Hariharan AK, Panda B (2016) An Improved Genome Assembly of *Azadirachta indica* A. Juss. *G3: Genes, Genomes, Genetics* 6: 1835–1840.
48. Bausher MG, Singh ND, Lee S-B, Jansen RK, Daniell H (2006) The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var. 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. *BMC Plant Biology* 6: 21. <https://doi.org/10.1186/1471-2229-6-21> PMID: [17010212](https://pubmed.ncbi.nlm.nih.gov/17010212/)
49. Saina JK, Li Z-Z, Gichira AW, Liao Y-Y (2018) The Complete Chloroplast Genome Sequence of Tree of Heaven (*Ailanthus altissima* (Mill.) Sapindales: Simaroubaceae), an Important Pantropical Tree. *International journal of molecular sciences* 19: 929.
50. Nie X, Lv S, Zhang Y, Du X, Wang L, et al. (2012) Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PloS one* 7: e36869. <https://doi.org/10.1371/journal.pone.0036869> PMID: [22606302](https://pubmed.ncbi.nlm.nih.gov/22606302/)
51. Zhang Y, Li B, Chen H, Wang Y (2016) Characterization of the complete chloroplast genome of *Acer miaotaiense* (Sapindales: Aceraceae), a rare and vulnerable tree species endemic to China. *Conservation Genetics Resources* 8: 383–385.
52. Jia Y, Yang J, He Y-L, He Y, Niu C, et al. (2016) Characterization of the whole chloroplast genome sequence of *Acer davidii* Franch (Aceraceae). *Conservation genetics resources* 8: 141–143.
53. Chen K-K (2018) Characterization of the complete chloroplast genome of the Tertiary relict tree *Phellodendron amurense* (Sapindales: Rutaceae) using Illumina sequencing technology. *Conservation Genetics Resources* 10: 43–46.
54. Qian J, Song J, Gao H, Zhu Y, Xu J, et al. (2013) The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PloS one* 8: e57607. <https://doi.org/10.1371/journal.pone.0057607> PMID: [23460883](https://pubmed.ncbi.nlm.nih.gov/23460883/)
55. Asaf S, Waqas M, Khan AL, Khan MA, Kang S-M, et al. (2017) The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species. *Frontiers in plant science* 8: 304. <https://doi.org/10.3389/fpls.2017.00304> PMID: [28326093](https://pubmed.ncbi.nlm.nih.gov/28326093/)
56. Asaf S, Khan AL, Khan MA, Imran QM, Kang S-M, et al. (2017) Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*) and nine other *Glycine* species. *PloS one* 12: e0182281. <https://doi.org/10.1371/journal.pone.0182281> PMID: [28763486](https://pubmed.ncbi.nlm.nih.gov/28763486/)
57. Morton BR (1998) Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. *Journal of molecular evolution* 46: 449–459. PMID: [9541540](https://pubmed.ncbi.nlm.nih.gov/9541540/)
58. Zheng W, Wang W, Harris A, Xu X (2017) The complete chloroplast genome of vulnerable *Aesculus wangii* (Sapindaceae), a narrowly endemic tree in Yunnan, China. *Conservation Genetics Resources*: 1–4.
59. Xu J-H, Wu H-B, Gao L-Z (2017) The complete chloroplast genome sequence of the threatened trident maple *Acer buergerianum* (Aceraceae). *Mitochondrial DNA Part B* 2: 273–274.
60. Yang J-B, Yang S-X, Li H-T, Yang J, Li D-Z (2013) Comparative chloroplast genomes of *Camellia* species. *PLoS One* 8: e73053. <https://doi.org/10.1371/journal.pone.0073053> PMID: [24009730](https://pubmed.ncbi.nlm.nih.gov/24009730/)
61. Xu J, Feng D, Song G, Wei X, Chen L, et al. (2003) The first intron of rice EPSP synthase enhances expression of foreign gene. *Science in China Series C: Life Sciences* 46: 561. <https://doi.org/10.1360/02yc0120> PMID: [18758713](https://pubmed.ncbi.nlm.nih.gov/18758713/)
62. Zhou T, Chen C, Wei Y, Chang Y, Bai G, et al. (2016) Comparative transcriptome and chloroplast genome analyses of two related Dipteronia Species. *Frontiers in plant science* 7: 1512. <https://doi.org/10.3389/fpls.2016.01512> PMID: [27790228](https://pubmed.ncbi.nlm.nih.gov/27790228/)
63. Su H-J, Hogenhout SA, Al-Sadi AM, Kuo C-H (2014) Complete chloroplast genome sequence of Omani lime (*Citrus aurantiifolia*) and comparative analysis within the rosids. *Plos one* 9: e113049. <https://doi.org/10.1371/journal.pone.0113049> PMID: [25398081](https://pubmed.ncbi.nlm.nih.gov/25398081/)

64. Wang R-J, Cheng C-L, Chang C-C, Wu C-L, Su T-M, et al. (2008) Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC evolutionary biology* 8: 36. <https://doi.org/10.1186/1471-2148-8-36> PMID: 18237435
65. Asaf S, Khan AL, Khan AR, Waqas M, Kang S-M, et al. (2016) Mitochondrial genome analysis of wild rice (*Oryza minuta*) and its comparison with other related species. *PLoS one* 11: e0152937. <https://doi.org/10.1371/journal.pone.0152937> PMID: 27045847
66. Huang H, Shi C, Liu Y, Mao S-Y, Gao L-Z (2014) Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evolutionary Biology* 14: 151. <https://doi.org/10.1186/1471-2148-14-151> PMID: 25001059
67. Rose O, Falush D (1998) A threshold size for microsatellite expansion. *Molecular biology and evolution* 15: 613–615. <https://doi.org/10.1093/oxfordjournals.molbev.a025964> PMID: 9580993
68. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, et al. (2007) Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8: 174. <https://doi.org/10.1186/1471-2164-8-174> PMID: 17573971
69. Huotari T, Korpelainen H (2012) Complete chloroplast genome sequence of *Elodea canadensis* and comparative analyses with other monocot plastid genomes. *Gene* 508: 96–105. <https://doi.org/10.1016/j.gene.2012.07.020> PMID: 22841789
70. Zhang Y, Iaffaldano BJ, Zhuang X, Cardina J, Cornish K (2017) Chloroplast genome resources and molecular markers differentiate rubber dandelion species from weedy relatives. *BMC plant biology* 17: 34–34. <https://doi.org/10.1186/s12870-016-0967-1> PMID: 28152978
71. Dong W, Liu J, Yu J, Wang L, Zhou S (2012) Highly Variable Chloroplast Markers for Evaluating Plant Phylogeny at Low Taxonomic Levels and for DNA Barcoding. *PLOS ONE* 7: e35071. <https://doi.org/10.1371/journal.pone.0035071> PMID: 22511980
72. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan A (2011) Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177: 309–334.
73. Kuang D-Y, Wu H, Wang Y-L, Gao L-M, Zhang S-Z, et al. (2011) Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome* 54: 663–673. <https://doi.org/10.1139/G11-026> PMID: 21793699
74. Saina JK, Gichira AW, Li Z-Z, Hu G-W, Wang Q-F, et al. (2018) The complete chloroplast genome sequence of *Dodonaea viscosa*: Comparative and phylogenetic analyses. *Genetica* 146: 101–113. <https://doi.org/10.1007/s10709-017-0003-x> PMID: 29170851
75. Provan J, Corbett G, Powell W, McNicol J (1997) Chloroplast DNA variability in wild and cultivated rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats. *Genome* 40: 104–110. PMID: 9061917
76. Curci PL, De Paola D, Danzi D, Vendramin GG, Sonnante G (2015) Complete chloroplast genome of the multifunctional crop globe artichoke and comparison with other Asteraceae. *PLoS One* 10: e0120589. <https://doi.org/10.1371/journal.pone.0120589> PMID: 25774672
77. Cavalier-Smith T (2002) Chloroplast evolution: secondary symbiogenesis and multiple losses. *Current Biology* 12: R62–R64. PMID: 11818081
78. Yi X, Gao L, Wang B, Su Y-J, Wang T (2013) The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. *Genome biology and evolution* 5: 688–698. <https://doi.org/10.1093/gbe/evt042> PMID: 23538991
79. Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki K-i (2004) Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. *DNA research* 11: 93–99. PMID: 15449542
80. Gao L, Yi X, Yang Y-X, Su Y-J, Wang T (2009) Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. *BMC evolutionary biology* 9: 130. <https://doi.org/10.1186/1471-2148-9-130> PMID: 19519899
81. Gillett JB (1973) *Commiphora* Jacq. (Burseraceae): Englerian Species Which "Disappear". *Kew Bulletin* 28: 25–28.
82. Becerra JX, Noge K, Olivier S, Venable DL (2012) The monophyly of *Bursera* and its impact for divergence times of Burseraceae. *Taxon* 61: 333–343.
83. Gostel MR, Phillipson PB, Weeks A (2016) Phylogenetic reconstruction of the myrrh genus, *Commiphora* (Burseraceae), reveals multiple radiations in Madagascar and clarifies infrageneric relationships. *Systematic Botany* 41: 67–81.
84. Wambugu PW, Brozynska M, Furtado A, Waters DL, Henry RJ (2015) Relationships of wild and domesticated rices (*Oryza* AA genome species) based upon whole chloroplast genome sequences. *Scientific Reports* 5: 13957. <https://doi.org/10.1038/srep13957> PMID: 26355750

85. Wu Z, Tembrock LR, Ge S (2015) Are Differences in Genomic Data Sets due to True Biological Variants or Errors in Genome Assembly: An Example from Two Chloroplast Genomes. PLOS ONE 10: e0118019. <https://doi.org/10.1371/journal.pone.0118019> PMID: 25658309
86. Saina JK, Li ZZ, Gichira AW, Liao YY (2018) The Complete Chloroplast Genome Sequence of Tree of Heaven (*Ailanthus altissima* (Mill.) (Sapindales: Simaroubaceae), an Important Pantropical Tree. Int J Mol Sci 19.