




Brief Report

Development of Genomic Resources in Mexican *Bursera* (Section: *Bullockia*: Burseraceae): Genome Assembly, Annotation, and Marker Discovery for Three Copal Species

Yessica Rico ^{1,*} , Gustavo P. Lorenzana ² , Carlos A. Benítez-Pineda ¹ and Bode A. Olukolu ³ ¹ Instituto de Ecología, A.C., Red de Diversidad Biológica del Occidente Mexicano, Pátzcuaro 61600, Mexico² Independent Consultant, Pátzcuaro 61608, Mexico³ Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN 37996, USA

* Correspondence: yessica.rico@inecol.mx

Abstract: *Bursera* comprises ~100 tropical shrub and tree species, with the center of the species diversification in Mexico. The genomic resources developed for the genus are scarce, and this has limited the study of the gene flow, local adaptation, and hybridization dynamics. In this study, based on ~155 million Illumina paired-end reads per species, we performed a de novo genome assembly and annotation of three *Bursera* species of the *Bullockia* section: *Bursera bipinnata*, *Bursera cuneata*, and *Bursera palmeri*. The total lengths of the genome assemblies were 253, 237, and 229 Mb for *B. cuneata*, *B. palmeri*, and *B. bipinnata*, respectively. The assembly of *B. palmeri* retrieved the most complete and single-copy BUSCOs (87.3%) relative to *B. cuneata* (86.5%) and *B. bipinnata* (76.6%). The ab initio gene prediction recognized between 21,000 and 32,000 protein-coding genes. Other genomic features, such as simple sequence repeats (SSRs), were also detected. Using the de novo genome assemblies as a reference, we identified single-nucleotide polymorphisms (SNPs) for a set of 43 *Bursera* individuals. Moreover, we mapped the filtered reads of each *Bursera* species against the chloroplast genomes of five Burseraceae species, obtaining consensus sequences ranging from 156 to 160 kb in length. Our work contributes to the generation of genomic resources for an important but understudied genus of tropical-dry-forest species.

Keywords: whole-genome sequencing; genomic resources; OmeSeq-qRRS (quantitative reduced-representation sequencing); tropical dry forest; Mexico



Citation: Rico, Y.; Lorenzana, G.P.; Benítez-Pineda, C.A.; Olukolu, B.A. Development of Genomic Resources in Mexican *Bursera* (Section: *Bullockia*: Burseraceae): Genome Assembly, Annotation, and Marker Discovery for Three Copal Species. *Genes* **2022**, *13*, 1741. <https://doi.org/10.3390/genes13101741>

Received: 31 August 2022

Accepted: 23 September 2022

Published: 27 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The development of genomic resources, such as the assembly of reference genomes and their annotations, is still rare for nonmodel organisms of tropical forest ecosystems, which are hotspots of global biodiversity. The availability of species genomes facilitates a wide range of ecological and evolutionary studies, including insights into candidate genes under selection and local adaptation, assessments of the genome-wide diversity and divergence, the characterization of the gene composition and function, marker discovery, among others [1–3]. Reference genomes provide a basis for subsequent population genomic studies, which can inform the conservation status and trends of vulnerable species [4,5].

Bursera comprises a taxonomic genus of approximately 100 deciduous and resinous tree and shrub species, with a distribution range that spans from the southwestern United States to Peru, the Bahamas, the Galapagos, and the Greater Antilles. The genus has its center of species diversity and endemism in the Balsas Basin of western Mexico, and it is a distinctive element of the tropical dry forests (TDFs) [6]. *Bursera* species are good candidates for the restoration of disturbed TDF sites because of their capacity for asexual propagation through rooted cuttings [7]. Many *Bursera* species have also had economic and cultural importance since pre-Columbian times. Its aromatic resins are used for religious and medicinal purposes, while the wood is utilized for the elaboration of handcrafts [8], and

thus they are subject to selective logging in some rural regions. Due to their overexploitation and the increasing loss of the TDF by anthropogenic land-use changes [9,10], 67% of *Bursera* species are under some threat category within the IUCN Red List [11].

Genetic studies on *Bursera* are scarce, and most of them are focused on its phylogenetic relationships [12–15], with only some examples on the phylogeography [16,17], population genetics [18], interspecific hybridization [19,20], and marker discovery for taxonomic applications [21]. The genome assembly and annotation in Mexican *Bursera* would facilitate an understanding of the patterns of the local adaptation, genome-wide diversity and differentiation, and hybridization dynamics. We report the first de novo genome assemblies for three copal species of the *Bullockia* section, which are widespread in Mexico: *B. bipinnata*, *B. cuneata*, and *B. palmeri* [22]. The three species are diploids [23]. *Bursera bipinnata* has the widest distribution across the TDF along the Pacific coast of Mexico, with an altitudinal interval from 1650 to 2200 masl. *Bursera palmeri* is distributed at the southwest of the Mexican Plateau, with an altitudinal interval from 1600 to 2300 masl, while *B. cuneata* occurs mostly in the Bajío region, with an altitudinal interval from 1850 to 2500 masl. The three species occur in sympatry in the remnant TDFs of Michoacán and Guanajuato, where they exchange gene flows and produce fertile hybrids [20]. *Bursera bipinnata* is the most common species that is commercially exploited for the extraction of aromatic resins. *Bursera palmeri* is used locally as a fuel source, while the wood of *B. cuneata* is utilized for the elaboration of religious figures and traditional masks in Michoacán.

The main aim of this work was to generate draft reference genomes through de novo genome assembly and the annotation of *B. bipinnata*, *B. cuneata*, and *B. palmeri*. We also mapped the obtained filtered reads against the chloroplast genomes of five Burseraceae species: *Boswellia sacra* [24], *Canarium album* [25], *Commiphora wightii* [26], *C. gileadensis*, and *C. foliacea* [24]. Additionally, for highlighting the utility of these genomes for subsequent population genetic and genomic studies, we identified nuclear genetic markers, such as simple sequence repeats (SSRs), in each of the assemblies, and single-nucleotide polymorphisms (SNPs) in a group of 43 *Bursera* individuals of the three species, using the assemblies as a reference. Our work thus contributes to the generation of genomic resources for an important genus of tropical-dry-forest trees.

2. Materials and Methods

2.1. Sampling and Sequencing

To accomplish the genome sequencing, we collected leaf tissue from single individuals in the localities where each of the species occurs in allopatry (*B. bipinnata*: Jalisco; *B. palmeri*: Querétaro; *B. cuneata*: Ciudad de México) to avoid sampling potentially introgressed individuals resulting from interspecific hybridization [20]. Additionally, we collected leaf-tissue samples from 43 *Bursera* trees (*B. cuneata*: $n = 23$; *B. palmeri*: $n = 4$; *B. bipinnata*: $n = 16$), which were used to identify SNPs using the OmeSeq-qRRS (quantitative reduced-representation sequencing) method (see below). Most samples were from the state of Michoacán (Supplemental Table S1). Leaves were preserved in sealable plastic bags containing silica gel until DNA extractions were performed.

Genomic DNA from 20 mg of dried tissue were extracted using the CTAB extraction protocol with prewash steps to eliminate the excess polyphenols [27]. DNA integrity was evaluated by observing a unique DNA band through 1% agarose gel (i.e., not multiple bands). Yield was estimated using a fluorimeter method according to the manufacturer's instructions (Qubit; Thermo Fisher Scientific, Waltham, MA, USA). For the whole-genome sequencing, 100 ng of high-molecular-weight DNA was used for the Illumina Truseq Nano DNA Library Prep Kit (Illumina, San Diego, CA, USA), according to the company's guidelines. The resulting libraries were subsequently sequenced in two lanes of the Illumina NovaSeq S4 flow-cell system to obtain ~140–170 million reads (150 bp paired-end reads) per each species.

We used 25 ng of high-molecular-weight DNA, which was extracted from the 43 *Bursera* individuals. DNA was sequentially double digested with the restriction enzymes

NsiI-HF and NlaIII. Barcoded adapters were incorporated into genomic fragments following the OmeSeq-qRRS method [28]. Resulting libraries were then diluted to 10 nmol/L for sequencing on a single lane of the NovaSeq S4 flow-cell system (150 bp paired-end reads). The whole-genome and OmeSeq library preparations and sequencing were performed by the NC State Genomic Sciences Laboratory (Raleigh, NC, USA).

2.2. Draft Genome Assembly and Annotation

Raw reads from whole-genome sequencing were preprocessed to eliminate adapters and low-quality bases from paired-end short reads using FastQC [29] and Fastp [29,30]. We performed preliminary runs with three de novo genome-assembly methods using their respective default parameters, with k -mer sizes ranging from 75 to 90: ABySS v.2.3.4 [31], SGA [32], and Platanus v.1.2.4 [33]. After the preliminary runs, we selected ABySS as the best method based on the computer run time, contig size, and BUSCO completeness. Assembly optimization using ABySS was then performed by modifying the k -mer size (k) and k -mer minimum coverage multiplicity cutoff (kc) to obtain the largest N50 and BUSCO (Benchmarking Universal Single-Copy Orthologs) completeness metrics, evaluated against the Eudicotyledoneae-lineage database (eudicots_odb10), downloaded from BUSCO v.3.0.2 [34]. We then proceeded to close the gaps between the contigs for each of the selected assemblies (Fasta format) using the ABySS-sealer module and BESST [35]. Genome features, such as genome size and single-copy regions, were estimated using the k -mer method [36,37], which relies on counting the total nonerroneous k -mers in the filtered raw reads. For this end, we used Jellyfish v.2.3.0 [38], and we averaged the results for the odd-number words ranging from 17 to 31. Consequently, we performed the contiguity test in QUASt v.5.0.2 [39] and BUSCO completeness analyses to obtain the definitive metrics for each assembly.

For the annotation process, we ran the pipeline in the GenSAS (Genome Sequence Annotation Server) online platform [40]. Due to constraints imposed for uploading data to the server, we performed the annotation on subsets of the assemblies, which encompassed those contigs larger than 2.5 kbps. The ab initio gene prediction was automated through the *Structural* module in GenSAS, using AUGUSTUS v.3.4.0 [41], with the parameter settings for: *Arabidopsis thaliana* species annotation, to report genes on both strands, and to predict complete genes, using the GeneMark-ES v.4.48 self-training algorithm for novel gene identification [42], and specifying a minimum contig length of 50 kbp, and a maximum gap of 5 kbp.

Using BWA-MEM [43], we mapped the preprocessed filtered reads against the genome assembly of *Bowsellia sacra* (NCBI: GCA_013180625.1), which is an African tree of the Burseraceae family, and which is the closest species to *Bursera* with an available whole-genome sequence [44]. Additionally, we aligned the preprocessed filtered reads to the chloroplasts of five species belonging to Burseraceae: *B. sacra* (NCBI: NC_029420.1); *C. album* (NCBI: NC_048982.1); *C. foliacea* (NCBI: NC_041103.1); *C. gileadensis* (NCBI: NC_041104.1); *C. wightii* (NCBI: NC_036978.1).

2.3. Marker Discovery

The *SSR Finder* module in GenSAS [41] was employed to identify dinucleotide to hexanucleotide SSRs for each species in their respective assemblies. We specified three repetitions for di, tri, and tetranucleotides: four in pentanucleotides motifs, and five in hexanucleotide motifs.

To identify the SNPs, the OmeSeq raw reads for the 43 *Bursera* samples were demultiplexed and quality filtered for removing erroneous base calls and contaminating adapter sequences using the automated pipeline ngsComposer [45]: <https://github.com/bodeolukolu/ngsComposer>; accessed on 16 May 2022). Subsequently, the assembled genomes of the three *Bursera* species were used as a reference for the variant calling and filtering with the automated pipeline GBSapp (<https://github.com/bodeolukolu/GBSapp>;

accessed on 16 May 2022). SNPs with minor allele frequencies (MAF) lower than 0.02 were removed, and the SNPs had to be called in at least 80% of the individuals.

To observe the genetic relationships between the three species, we performed a discriminant analysis of principal components (DAPC) in the R package *Adegenet* [46]. The DAPC is a multivariate analysis that is free of HW and LD assumptions, which maximizes the genetic variation among groups, which, in our case, are the species. The optimal number of PCs to retain was optimized with the *xvalDapc* function [46].

3. Results

3.1. Genome Assembly and Annotation

A total of 34.5 Gb of raw 151 bp Illumina reads were sequenced for the tree copal species. The resulting draft genome of *B. cuneata* was 253 Mb in size, with a GC content of 34.97%, and an N50 length of 0.014 Mb. The assembled genome size of *B. bipinnata* was 229 Mb, with a GC content of 34.14%, and an N50 length of 0.006 Mb. The assembled genome size of *B. palmeri* was 237 Mb, with a GC content of 34.11%, and an N50 length of 0.011 Mb. The assembly statistics are shown in Table 1. The genome assembly of *B. palmeri* recovered the most complete and single-copy BUSCOs (87.3%) relative to *B. cuneata* (86.5%) and *B. bipinnata* (76.6%). The completeness assessments of the three assembled genomes are shown in Figure 1.

Table 1. Assembly metrics for three *Bursera* de novo nuclear genomes.

Species	Genome Assembly Size (Mb)	Contigs (n)	N50 (bp)	L50 (n)	Largest Contig (bp)	GC Content (%)
<i>B. cuneata</i>	252.9	70,005	13,532	4324	203,078	34.97
<i>B. bipinnata</i>	229.2	93,723	5699	9652	104,624	34.14
<i>B. palmeri</i>	237.4	67,240	11,112	4918	230,450	34.11

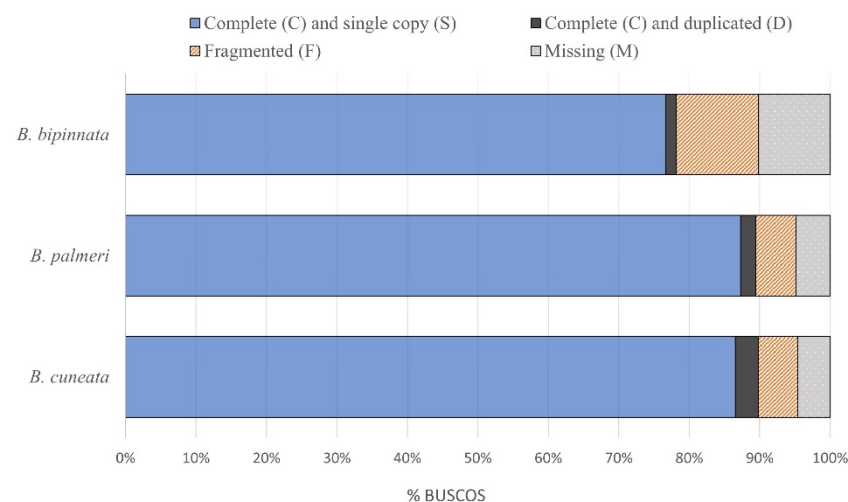


Figure 1. Results from three *Bursera* genome contiguity assessments using BUSCO.

The estimated genome sizes based on the *k-mer* method applied to the filtered reads were 565 ± 189 , 591 ± 172 , and 840 ± 80 Mb for *B. cuneata*, *B. palmeri*, and *B. bipinnata*, respectively. Similarly, the estimated single-copy regions were 301 ± 117 , 290 ± 94 , and 351 ± 71 Mb, respectively, implying that the final assemblies recovered between 65% and 84% of the nonrepetitive regions along the genomes. This allowed for the identification of 31,357 protein-coding genes in *B. cuneata*, 29,558 in *B. palmeri*, and 21,043 in *B. bipinnata* through the ab initio prediction approach.

The alignment and mapping of the filtered reads against the nuclear genome of *B. sacra* resulted in approximately 53–63 million mapped reads, which represented ~40% of the

total filtered reads in each piece of the *Bursera* sequencing data (Table 2). Similarly, the identification of the chloroplast genomes resulted in between approximately 3.9 and 7.5 million mapped reads against the five reference genomes, which represents from ~2.8 (*B. bipinnata*) to 4.4% (*B. cuneata*) of the total filtered reads. *C. wightii* was the reference plastid genome with the most matching reads against the three *Bursera* species (Table 3). The chloroplast genomes were 156–160 kb in length, with a GC content of ~27%.

Table 2. Alignment and mapping of the preprocessed filtered reads of the three *Bursera* species against the nuclear genome of *Boswellia sacra*.

Species	Total Primary Reads (Million)	Mapped Reads (Million)	Properly Paired Reads (Million)
<i>B. cuneata</i>	169.8	62.9 (37.0%)	43.5 (25.6%)
<i>B. bipinnata</i>	141.9	53.6 (37.8%)	35.6 (25.1%)
<i>B. palmeri</i>	145.2	62.9 (43.4%)	48.6 (33.5%)

Table 3. Alignment and mapping of the preprocessed filtered reads of the three *Bursera* species against the chloroplast genomes of five Burseraceae species. The percentages of the total mapped filtered reads are in parentheses.

Species	Reads (M)	<i>Boswellia sacra</i>	<i>Canarium album</i>	<i>Commiphora foliacea</i>	<i>Commiphora gileadensis</i>	<i>Commiphora wightii</i>
<i>B. cuneata</i>	169.8	7,481,910 (4.40%)	7,253,921 (4.27%)	7,304,511 (4.30%)	7,364,709 (4.34%)	7,514,759 (4.42%)
<i>B. bipinnata</i>	141.9	4,080,000 (2.87%)	3,908,183 (2.75%)	3,948,175 (2.78%)	3,970,894 (2.80%)	4,122,751 (2.90%)
<i>B. palmeri</i>	145.2	5,187,716 (3.57%)	4,953,793 (3.41%)	5,043,110 (3.47%)	5,042,119 (3.47%)	5,238,798 (3.61%)

3.2. SSR and SNP Discovery

A total of 107,270, 100,614, and 76,766 dinucleotide to hexanucleotide SSRs were identified for *B. cuneata*, *B. palmeri*, and *B. bipinnata*, respectively (Supplementary Table S2). For the three species, dinucleotides were the most common motifs, accounting for ~36% of all the identified SSRs, followed by the tetranucleotide (~29%) and trinucleotide motifs (~19%). Smaller numbers of pentanucleotide (~9%) and hexanucleotide (~4%) motifs were also observed among the three species (Table 4). Within the dinucleotide motifs, the AT/TA were the most frequent, accounting for ~65% of the motifs identified in this category, while, for the tetranucleotides, the AAAT/TTTA motifs were the most common (~20%); these trends were shared among the three species.

Table 4. Frequencies of dinucleotide to hexanucleotide simple sequence repeats identified in three *Bursera* species.

Species	Total	Di	Tri	Tetra	Penta	Hexa
<i>B. cuneata</i>	107,270	39,769 (37.1%)	20,606 (19.2%)	31,871 (29.7%)	10,011 (9.3%)	5013 (4.7%)
<i>B. bipinnata</i>	76,766	27,915 (36.4%)	15,049 (19.6%)	23,067 (30%)	6988 (9.1%)	3747 (4.9%)
<i>B. palmeri</i>	100,614	37,480 (37.3%)	19,431 (19.3%)	29,818 (29.6%)	9212 (9.2%)	4673 (4.6%)

The alignment and mapping of the OmeSeq filtered reads against the three draft genomes in a set of 43 *Bursera* individuals resulted in a high proportion of reads (i.e., from 95 to 99%) that mapped to the reference genomes. The variant calling and filtering resulted in 5543 biallelic SNPs based on a minor-allele-frequency (MAF) threshold of 0.02, and no more than 20% missing SNPs across individuals. The results from the DAPC showed the distinction among the three species (Figure 2), although there was a large overlap between

some individuals of the three species: *B. cuneata* and *B. bipinnata*, being the two species with greater genetic similarity.

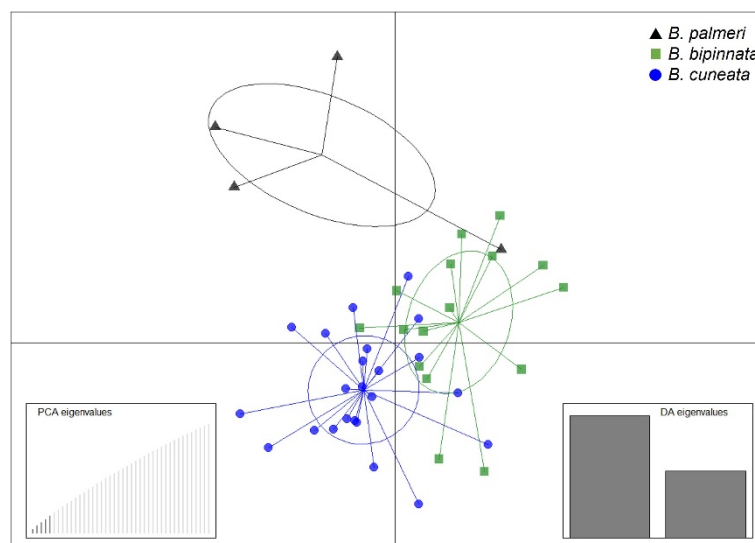


Figure 2. DAPC plot based on 5543 biallelic SNPs for 43 *Bursera* individuals. The left inset shows the number of retained PCs, and the right inset denotes the first two discriminant functions used.

4. Discussion

The availability of sequenced genomes is lacking for the genus *Bursera*, despite its ecological and cultural importance in neotropical-dry-forest ecosystems. So far, the only available genome within the family Burseraceae is for *Boswellia sacra*, which occurs in Asia and Africa. Here, we generated the first de novo genome assemblies for three *Bursera* species of the *Bullockia* section: *B. bipinnata*, *B. cuneata*, and *B. palmeri*, which are native to Mexico.

Even though the draft genomes were assembled by relying on short paired-end Illumina reads, the quality obtained was acceptable (BUSCOs ranging from 76 to 87%), and we could recover between 21,043 and 31,357 genes based on the ab initio gene prediction. The lower contiguity and BUSCO scores obtained for *B. bipinnata* relative to the other two *Bursera* species were likely associated with the quality of the DNA and its larger estimated genome size (it is noteworthy that *Bursera* species are diploids, and they have small genomes compared with other forest tree species [23]). The estimated genome sizes for the three *Bursera* species based on the *k*-mer count ranged from 565 to 840 Mb, which are larger than the genome assembly size reported for *B. sacra* (432 Mb). The read mapping of the three *Bursera* species against the nuclear genome of *B. sacra* was relatively low (<50%), likely reflecting their large phylogenetic distance within the Burseraceae family [13]. This highlights the importance of developing genus- and species-specific genomic resources for *Bursera*, as the available genomic data in other genera within the Burseraceae family are not representative of the genome-wide variation in *Bursera*. It is especially important to develop draft genomes for species within each of the two *Bursera* sections, as they have distinct evolutionary histories [47].

Moreover, we identified thousands of nuclear SSR markers in each species, the most common being the dinucleotide repeats. Specifically, the AT/TA motifs were by far the most common, which agreed with previous studies on SSR discovery in plant species [48,49]. SSRs are cost-effective and highly variable codominant markers that are very popular in conservation genetic studies [50]. The lack of available polymorphic markers, such as SSRs, in species of *Bursera* has precluded the implementation of intraspecific genetic studies for assessments of the genetic diversity, gene flow, and structure in vulnerable species. A subsequent step of the SSRs identified here would be their PCR validation and assessments of the levels of polymorphisms. Such polymorphic markers could then be potentially cross

amplified in closely related *Bursera* species, thus contributing to increase our knowledge on the population genetic diversity and structure in several species.

Additionally, to exemplify the utility of our nuclear genome assemblies as reference genomes, we performed the identification of biallelic SNPs in a set of 43 *Bursera* individuals, independently sequenced through the OmeSeq-qRRS approach. The genome-wide markers that were identified constitute a valuable tool to evaluate the neutral and adaptive genetic variation for intra- and interspecific-level comparisons in genomic studies. In our case, the DAPC not only showed the species clustering, but also the overlapping among the individuals of the three species. These three *Bursera* species are known to hybridize in co-occurring locations [20], which may explain the observed intermixing resulting from the interspecific gene flow. The development of genomic resources for three Mexican *Bursera* species is a first step toward increasing the scientific investigations on population genomics and comparative phylogenetic studies.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13101741/s1>, Table S1. Locality information of 43 sampled *Bursera* individuals used for SNP calling. Table S2. List of nuclear SSRs identified for each *Bursera* species.

Author Contributions: Y.R. conceived and designed the study, conducted the field sampling and laboratory procedures, provided the reagents and materials, and wrote the manuscript; G.P.L. analyzed the genomic data; C.A.B.-P. analyzed the SNP data; B.A.O. generated the SNP data. All the authors carefully edited and approved the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by Consejo Nacional de Ciencia y Tecnología (CONACYT) (Ciencia Básica CB2016-283237 to Y.R.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The nuclear genomes of three *Bursera* species have been deposited in the NCBI under the project number PRJNA828589, available at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA828589> (accessed on 22 September 2022).

Acknowledgments: We are thankful for the assistance in the field of Benjamín Castillo Ponce, Bruno A. Gutiérrez Becerril, Stephanie Aguilera López, Tania Andrade Ortiz, and Victor Reyes Pino, and to several field assistants from the local communities. Moreover, we also want to thank Mayra Castro Morales for performing the DNA extractions; Marisol Zurita Solis, Ingrid Lara, and Antonio González Rodríguez for the DNA quantification using Qubit; Carlos Cultid-Medina for providing the computer equipment to perform the genome assemblies and annotation.

Conflicts of Interest: The authors declare no conflict of interest.

Sampling Permit: SEMARNAT SGPA/DGGFS/712/1062/18.

References

1. Ekblom, R.; Wolf, J.B.W. A Field Guide to Whole-Genome Sequencing, Assembly and Annotation. *Evol. Appl.* **2014**, *7*, 1026–1042. [[CrossRef](#)] [[PubMed](#)]
2. Bellin, D.; Ferrarini, A.; Chimento, A.; Kaiser, O.; Levenkova, N.; Bouffard, P.; Delledonne, M. Combining Next-Generation Pyrosequencing with Microarray for Large Scale Expression Analysis in Non-Model Species. *BMC Genom.* **2009**, *10*, 1–9. [[CrossRef](#)] [[PubMed](#)]
3. Unamba, C.I.N.; Nag, A.; Sharma, R.K. Next Generation Sequencing Technologies: The Doorway to the Unexplored Genomics of Non-Model Plants. *Front. Plant Sci.* **2015**, *6*, 1074. [[CrossRef](#)] [[PubMed](#)]
4. Hoban, S.; Campbell, C.D.; da Silva, J.M.; Ekblom, R.; Funk, W.C.; Garner, B.A.; Godoy, J.A.; Kershaw, F.; MacDonald, A.J.; Mergeay, J.; et al. Genetic Diversity Is Considered Important but Interpreted Narrowly in Country Reports to the Convention on Biological Diversity: Current Actions and Indicators Are Insufficient. *Biol. Conserv.* **2021**, *261*, 109233. [[CrossRef](#)]
5. Formenti, G.; Theissinger, K.; Fernandes, C.; Bista, I.; Bombarely, A.; Bleidorn, C.; Ciofi, C.; Crottini, A.; Godoy, J.A.; Höglund, J.; et al. The Era of Reference Genomes in Conservation Genomics. *Trends Ecol. Evol.* **2022**, *37*, 197–202. [[CrossRef](#)]

6. McVaugh, R.; Rzedowski, J. Synopsis of the *Genus bursera* L. in Western Mexico, with Notes on the Material of *Bursera* Collected by Sesse & Mocino. *Kew Bull.* **1965**, *18*, 317. [[CrossRef](#)]
7. Bonfil-Sanders, C.; Cajero-Lázaro, I.; Evans, R.Y. Germinación de Semillas de Seis Especies de *Bursera* Del Centro de México. *Agrociencia* **2008**, *42*, 827–834.
8. Aurora Montufar, L. Identidad y Simbolismo Del Copal Prehispánico y Reciente. *Arqueología* **2004**, *33*, 60–71.
9. Trejo, I.; Dirzo, R. Deforestation of Seasonally Dry Tropical Forest: A National and Local Analysis in Mexico. *Biol. Conserv.* **2000**, *94*, 133–142. [[CrossRef](#)]
10. Hernández-Oria, J.G. Desaparición Del Bosque Seco En El Bajío Mexicano: Implicaciones Del Ensamblaje de Especies y Grupos Funcionales En La Dinámica de Una Vegetación Amenazada. *Zonas Áridas* **2007**, *11*, ág. 13–31. [[CrossRef](#)]
11. Fuentes, A.C.D.; Samain, M.S.; Martínez-Salas, E. The IUCN Red List of Threatened Species 2019: E.T137371772A137376559. 2019. Available online: <https://www.iucnredlist.org/es/species/137371772/137376559> (accessed on 2 May 2022).
12. Becerra, J.X.; Venable, D.L. Nuclear Ribosomal DNA Phylogeny and Its Implications for Evolutionary Trends in Mexican *Bursera* (Burseraceae). *Am. J. Bot.* **1999**, *86*, 1047–1057. [[CrossRef](#)] [[PubMed](#)]
13. Becerra, J.X. Evolution of Mexican *Bursera* (Burseraceae) Inferred from ITS, ETS, and 5S Nuclear Ribosomal DNA Sequences. *Mol. Phylogenet. Evol.* **2003**, *26*, 300–309. [[CrossRef](#)]
14. Weeks, A.; Simpson, B.B. Molecular Phylogenetic Analysis of *Commiphora* (Burseraceae) Yields Insight on the Evolution and Historical Biogeography of an “Impossible” Genus. *Mol. Phylogenet. Evol.* **2007**, *42*, 62–79. [[CrossRef](#)]
15. Rosell, J.A.; Olson, M.E.; Weeks, A.; De-Nova, J.A.; Lemos, R.M.; Camacho, J.P.; Fera, T.P.; Gómez-Bermejo, R.; Montero, J.C.; Eguarte, L.E. Diversification in Species Complexes: Tests of Species Origin and Delimitation in the *Bursera* Simaruba Clade of Tropical Trees (Burseraceae). *Mol. Phylogenet. Evol.* **2010**, *57*, 798–811. [[CrossRef](#)] [[PubMed](#)]
16. Weeks, A.; Tye, A. Phylogeography of Palo Santo Trees (*Bursera Graveolens* and *Bursera Malacophylla*; Burseraceae) in the Galápagos Archipelago. *Bot. J. Linn. Soc.* **2009**, *161*, 396–410. [[CrossRef](#)]
17. Poelchau, M.F.; Hamrick, J.L. Comparative Phylogeography of Three Common Neotropical Tree Species. *J. Biogeogr.* **2013**, *40*, 618–631. [[CrossRef](#)]
18. Dunphy, B.K.; Hamrick, J.L. Estimation of Gene Flow into Fragmented Populations of *Bursera Simaruba* (Burseraceae) in the Dry-Forest Life Zone of Puerto Rico. *Am. J. Bot.* **2007**, *94*, 1786–1794. [[CrossRef](#)]
19. Weeks, A.; Simpson, B.B. Molecular Genetic Evidence for Interspecific Hybridization among Endemic Hispaniolan *Bursera* (Burseraceae). *Am. J. Bot.* **2017**, *91*, 976–984. [[CrossRef](#)]
20. Melecio, E.Q.; Rico, Y.; Noriega, A.L.; Rodríguez, A.G. Molecular Evidence and Ecological Niche Modeling Reveal an Extensive Hybrid Zone among Three *Bursera* Species (Section *Bullockia*). *PLoS ONE* **2021**, *16*, e0260382. [[CrossRef](#)]
21. Collins, E.S.; Gostel, M.R.; Weeks, A. An Expanded Nuclear Phylogenomic PCR Toolkit for Sapindales. *Appl. Plant Sci.* **2016**, *4*, 1600078. [[CrossRef](#)]
22. Rzedowski, J.; Guevara-Féfer, F. Burseraceae. *Flora Del Bajío Y De Reg. Adyac.* **1992**, *3*, 46.
23. Guimarães, R.; Forni-Martins, E.R. Chromosome Numbers and Their Evolutionary Meaning in the Sapindales Order: An Overview. *Braz. J. Bot.* **2021**, *45*, 77–91. [[CrossRef](#)]
24. Khan, A.; Asaf, S.; Khan, A.L.; Al-Harrasi, A.; Al-Sudairy, O.; AbdulKareem, N.M.; Khan, A.; Shehzad, T.; Alsaady, N.; Al-Lawati, A.; et al. First Complete Chloroplast Genomics and Comparative Phylogenetic Analysis of *Commiphora Gileadensis* and *C. Foliacea*: Myrrh Producing Trees. *PLoS ONE* **2019**, *14*, e0208511. [[CrossRef](#)] [[PubMed](#)]
25. Lai, R.-L.; Feng, X.; Chen, J.; Chen, Y.-T.; Wu, R.-J. The Complete Chloroplast Genome Characterization and Phylogenetic Analysis of *Canarium Album*. *Mitochondrial. DNA B Resour.* **2019**, *4*, 2948–2949. [[CrossRef](#)] [[PubMed](#)]
26. Monpara, J.; Thaker, V. Phylogenetic Position and Marker Studies Using CpDNA of *C. Wightii*: A Critically Endangered and Medicinally Important Plant in India. *Vegetos* **2021**, *34*, 300–308. [[CrossRef](#)]
27. Healey, A.; Furtado, A.; Cooper, T.; Henry, R.J. Protocol: A Simple Method for Extracting next-Generation Sequencing Quality Genomic DNA from Recalcitrant Plant Species. *Plant Methods* **2014**, *10*, 21. [[CrossRef](#)] [[PubMed](#)]
28. Yencho, G.C.; Olukolu, B.A. Compositions and Methods Related to Quantitative Reduced Representation Sequencing. 2022. Available online: <https://patents.google.com/patent/US20220243267A1/> (accessed on 6 June 2022).
29. Andrews, S. FASTQC. 2010. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 9 March 2022).
30. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884. [[CrossRef](#)] [[PubMed](#)]
31. Simpson, J.T.; Wong, K.; Jackman, S.D.; Schein, J.E.; Jones, S.J.M.; Birol, I. ABySS: A Parallel Assembler for Short Read Sequence Data. *Genome Res.* **2009**, *19*, 1117. [[CrossRef](#)] [[PubMed](#)]
32. Simpson, J.T.; Durbin, R. Efficient de Novo Assembly of Large Genomes Using Compressed Data Structures. *Genome Res.* **2012**, *22*, 549–556. [[CrossRef](#)] [[PubMed](#)]
33. Kajitani, R.; Toshimoto, K.; Noguchi, H.; Toyoda, A.; Ogura, Y.; Okuno, M.; Yabana, M.; Harada, M.; Nagayasu, E.; Maruyama, H.; et al. Efficient de Novo Assembly of Highly Heterozygous Genomes from Whole-Genome Shotgun Short Reads. *Genome Res.* **2014**, *24*, 1384–1395. [[CrossRef](#)]
34. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]

35. Sahlin, K.; Vezzi, F.; Nystedt, B.; Lundeberg, J.; Arvestad, L. BESST—Efficient Scaffolding of Large Fragmented Assemblies. *BMC Bioinform.* **2014**, *15*, 1–11. [[CrossRef](#)] [[PubMed](#)]
36. Lu, M.; An, H.; Li, L. Genome Survey Sequencing for the Characterization of the Genetic Background of *Rosa Roxburghii* Tratt and Leaf Ascorbate Metabolism Genes. *PLoS ONE* **2016**, *11*, e0147530. [[CrossRef](#)] [[PubMed](#)]
37. Guo, X.M.; Wang, Z.F.; Zhang, Y.; Wang, R.J. Chromosomal-Level Assembly of the *Leptodermis Oblonga* (Rubiaceae) Genome and Its Phylogenetic Implications. *Genomics* **2021**, *113*, 3072–3082. [[CrossRef](#)] [[PubMed](#)]
38. Marçais, G.; Kingsford, C. A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of k-Mers. *Bioinformatics* **2011**, *27*, 764–770. [[CrossRef](#)]
39. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)] [[PubMed](#)]
40. Humann, J.L.; Lee, T.; Ficklin, S.; Main, D. Structural and Functional Annotation of Eukaryotic Genomes with GenSAS. *Methods Mol. Biol.* **2019**, *1962*, 29–51. [[CrossRef](#)] [[PubMed](#)]
41. Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: Ab Initio Prediction of Alternative Transcripts. *Nucleic Acids Res.* **2006**, *34*, W435. [[CrossRef](#)]
42. Lomsadze, A.; Ter-Hovhannisyanyan, V.; Chernoff, Y.O.; Borodovsky, M. Gene Identification in Novel Eukaryotic Genomes by Self-Training Algorithm. *Nucleic Acids Res.* **2005**, *33*, 6494. [[CrossRef](#)]
43. Li, H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
44. Khan, A.L.; Al-Harrasi, A.; Wang, J.P.; Asaf, S.; Riethoven, J.J.M.; Shehzad, T.; Liew, C.S.; Song, X.M.; Schachtman, D.P.; Liu, C.; et al. Genome Structure and Evolutionary History of Frankincense Producing *Boswellia Sacra*. *iScience* **2022**, *25*, 104574. [[CrossRef](#)] [[PubMed](#)]
45. Kuster, R.D.; Yench, G.C.; Olukolu, B.A. NgsComposer: An Automated Pipeline for Empirically Based NGS Data Quality Filtering. *Brief. Bioinform.* **2021**, *22*, 1–10. [[CrossRef](#)] [[PubMed](#)]
46. Jombart, T.; Devillard, S.; Dufour, A.-B.; Pontier, D. Revealing Cryptic Spatial Patterns in Genetic Variability by a New Multivariate Method. *Heredity* **2008**, *101*, 92–103. [[CrossRef](#)] [[PubMed](#)]
47. De-Nova, J.A.; Medina, R.; Montero, J.C.; Weeks, A.; Rosell, J.A.; Olson, M.E.; Eguiarte, L.E.; Magallón, S. Insights into the Historical Construction of Species-Rich Mesoamerican Seasonally Dry Tropical Forests: The Diversification of *Bursera* (Burseraceae, Sapindales). *New Phytol.* **2012**, *193*, 276–287. [[CrossRef](#)] [[PubMed](#)]
48. Zhang, Y.; He, J.; Zhao, P.X.; Bouton, J.H.; Monteros, M.J. Genome-Wide Identification of Microsatellites in White Clover (*Trifolium repens* L.) Using FIASCO and PhpSSRMiner. *Plant Methods* **2008**, *4*, 1–8. [[CrossRef](#)] [[PubMed](#)]
49. Cardle, L.; Ramsay, L.; Milbourne, D.; Macaulay, M.; Marshall, D.; Waugh, R. Computational and Experimental Characterization of Physically Clustered Simple Sequence Repeats in Plants. *Genetics* **2000**, *156*, 847. [[CrossRef](#)]
50. Rossetto, M.; Rymer, P.D. Applications of Molecular Markers in Plant Conservation. In *Molecular Markers in Plants*, 1st ed.; Henry, R.J., Ed.; John Wiley & Sons, Inc.: Chichester, UK, 2013; pp. 81–98. [[CrossRef](#)]