



Article The Complete Plastid Genome Sequences of the Belian (Eusideroxylon zwageri): Comparative Analysis and Phylogenetic Relationships with Other Magnoliids

Wen Zhu¹, Yunhong Tan^{2,3}, Xinxin Zhou⁴, Yu Song^{5,6,*} and Peiyao Xin^{1,*}

- Southwest Research Center for Landscape Architecture Engineering, National Forestry and Grassland Administration, Southwest Forestry University, Kunming 650224, China
- ² Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666303, China
- ³ Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Yezin, Nay Pyi Taw 05282, Myanmar
- ⁴ South China Botanical Garden Herbarium, Chinese Academy of Sciences, Guangzhou 510650, China
- ⁵ Key Laboratory of Ecology of Rare and Endangered Species and Environmental Protection (Ministry of Education), Guangxi Normal University, Guilin 541004, China
- ⁶ Guangxi Key Laboratory of Landscape Resources Conservation and Sustainable Utilization in Lijiang River Basin, Guangxi Normal University, Guilin 541004, China
- * Correspondence: songyu@gxnu.edu.cn (Y.S.); xpytgyx@163.com (P.X.)

Abstract: The Belian (*Eusideroxylon zwageri* Teijsm. & Binn.) is a commercially important timber species in Southeast Asia that was listed on the IUCN Red List of threatened species in 1998. Six years ago, we published an article in Genome Biology Ecology entitled "Evolutionary Comparisons of the Chloroplast Genome in Lauraceae and Insights into Loss Events in the Magnoliids" in which one complete plastid genome of Belian was assembled for comparative analyses of the plastomes in Lauraceae. However, a recent study concluded that our sequenced Belian individual can be located in the clade of Myristicaceae instead of that of Lauraceae. Here, we performed reanalyses of an additional two Belian plastomes, along with 42 plastomes from plants spanning 10 families of the Magnoliids. The three Belian plastomes are 39% CG and vary in length from 157,535 to 157,577 bp. A total of 37 tRNA genes, 8 rRNA genes, and 85 protein-coding genes were among the 130 annotated genes. There were 95–101 repeat sequences and 56–61 simple repeat sequences (SSRs). Comparative genomic analysis revealed 170 mutation sites in their plastomes, which include 111 substitutions, 53 indels, and 6 microinversions. Phylogeny was reconstructed using maximum-likelihood and Bayesian approaches for 44 magnoliids species, indicating that the 3 Belian individuals were nested among the species in the Lauraceae family rather than Myristicaceae.

Keywords: chloroplast genome; comparative genomics; *Eusideroxylon*; Lauraceae; phylogenetic relationships

1. Introduction

Lauraceae, belonging to the Magnoliids Laurales, with about 50 genera and 2500–3000 species worldwide, are mainly distributed in the tropical and subtropical regions of the world [1,2]. Lauraceae plants are widely used in many fields and have great economic value. For example, *Camphora officinarum* and *C. parthenoxylon* are important raw materials for light industry and medicine; *Phoebe* is a valuable wood; *Cinnamomum cassia* and *Lindera aggregata* are famous medicinal plants; avocados (*Persea americana*) are a very nutritious fruit; and so on (http://www.iplant.cn/info/Lauraceae?t=foc, accessed on 23 November 2023). Song et al. (2019) successfully constructed the system structure of six tribes and nine branches of Lauraceae, and *Eusideroxylon* Teijsm. & Binn. belongs to Cryptocaryeae [3]. As one of the monotypic genera in the family Lauraceae, *Eusideroxylon* includes the only



Citation: Zhu, W.; Tan, Y.; Zhou, X.; Song, Y.; Xin, P. The Complete Plastid Genome Sequences of the Belian (*Eusideroxylon zwageri*): Comparative Analysis and Phylogenetic Relationships with Other Magnoliids. *Forests* **2023**, *14*, 2443. https:// doi.org/10.3390/f14122443

Academic Editor: Claudia Mattioni

Received: 8 November 2023 Revised: 30 November 2023 Accepted: 11 December 2023 Published: 14 December 2023



Copyright: © 2023 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/). tree species *Eusideroxylon zwageri* Teijsm. & Binn. (https://www.worldfloraonline.org/, accessed on 10 October 2023), which is mainly distributed in the tropical rainforests of Brunei, Indonesia, Malaysia, and the Philippines. *E. zwageri* is a slow-growing (0.5 m per year) tall evergreen tree that grows up to 50 m in height (https://en.wikipedia.org/wiki/ Eusideroxylon, accessed on 23 November 2023). The flowers are pale yellow to yellow in color and are hermaphrodite; the fruits are drupes; the leaves are dark green and leathery; and the young leaves are reddish brown to yellowish red (https://en.wikipedia.org/wiki/ Eusideroxylon, accessed on 23 November 2023). *E. zwageri*, known as "Belian" in Malaysia, "Ulin" in Indonesia, and "Tambulan" in the Philippines by local people, is recognized for its globally durable wood, which is used for heavy construction and building of traditional houses [4]. Due to its density, *E. zwageri* is one of the very few wood species that do not float in water. In 1998, Belian (*Eusideroxylon zwageri*) was included as vulnerable in the IUCN Red List [5]. Currently, Belian, under national protection in Indonesia, exhibits decreasing populations due to past over-exploitation [6]. Therefore, acquiring the genetic resources and population data of the tropical tree Belian is urgent.

Chloroplasts are vital organelles in the cells of green plants and certain algae, where photosynthesis occurs [7]. The chloroplast genome of higher plants is usually quadripartite structure, consisting of one large single-copy region (LSC), one small single-copy region (SSC), and two reverse repeat regions (IRs) [8]. Terrestrial plant chloroplast genomes are typically 120 kb to 160 kb in size and comprise 110–130 genes [9–11]. Chloroplast genome size changes are frequently related to reverse repeat region expansion and contraction, gene deletion, intron deletion, gene spacer region differences, and short-segment repeat sequences [12,13]. Compared with the nuclear genome and the mitochondrial genome, the chloroplast genome has moderate evolutionary speed, small molecular weight, relatively conservative structure, and primarily single-parent inheritance, rarely recombination. The success rate of obtaining complete genome by modern sequencing technology is high, and it is an ideal material for studying plant genetics and evolution [14,15].

Aiso-Sanada et al. (2020) studied the basic characteristics of Belian wood in plantations [16]. Previously, Chanderbali et al. (2001) compared the morphological characteristics of Eusideroxylon, Potoxylon, and other Cryptocarya genera. Eusideroxylon and Potoxylon, with semi-inferior ovaries, lie sister to genera with superior ovaries [2]. Kimoto et al. (2006) studied the embryology of Eusideroxylon and found that Eusideroxylon is consistent with Aspidostemon, the core Cryptocaryeae, Caryodaphnopsis, and Cassytha in having a glandular anther tapetum [17]. Embryologically, *Eusideroxylon* appears to have an intermediate state between *Hypodaphnis* and the core Cryptocaryeae [17]. The first molecular approach of Chanderbali et al. (2001) reported the chloroplast sequences including trnL-trnF, psbA*trnH*, *trnT*–*trnL*, and *rpl16* and nuclear ribosomal sequences 26S and ITS for Belian, which grouped with the Aspidostemon, Beilschmiedia, Cryptocarya, Endiandra, Eusideroxylon, Hypodaphnis, Potameia, and Potoxylon species [2]. The second phylogenetic research, which used the chloroplast marker *trnK* intron to build a Bayesian tree of 49 species, revealed a well-supported Cryptocaryeae group that included Aspidostemon, Beilschmiedia, Cryptocarya, Endiandra, Eusideroxylon, and Potameia species [18]. Ten years later, Hiroyuki et al. (2015) identified 16 chloroplast polymorphic marker sequences, totaling 10,618 bp in length [19]. Then, Nurtjahjaningsih et al. (2017) developed 16 polymorphic markers using 6 chloroplast DNA regions (*atpE*, *ccsA*–*ndhD*, *matK*, *trnL*–*trnF*, *rpl2*, and *ycf3*) to investigate the genetic structure from 72 Belian trees from 9 populations [4]. Finally, Md-Isa et al. (2021) analyzed Belian genetic variation using 4 microsatellite markers in 52 samples from three populations: Nirwana Rehabilitation Forest (NRF) and Tatau, Sarawak [20]. These genome fragments were significant for the genetic conservation and population management of Belian, yet they are not a substitute for the complete chloroplast genome.

Six years ago, we reported a complete chloroplast genome sequence of Belian (Gen-Bank accession No. MF939351) and a monophyletic phylogenetic group of Belian and 45 other Lauraceae species [21]. Since then, the phylogenetic relationships of Lauraceae, or magnoliids, have been studied using this genome. Liu et al. (2021) constructed the largest plastid genome dataset of Lauraceae, combining Belian and 190 plastome genomes from 131 species of 25 genera, and generated a phylogenetic tree using ML and NJ methods [22]. Similarly, Yang et al. (2023) constructed two phylogenetic trees using the complete chloroplast sequence of Belian, along with other Lauraceae species by ML and BI methods [23]. In both studies, Belian was consistently placed within the Cryptocaryeae tribe, which comprises species from the genera *Beilschmiedia*, *Cryptocarya*, *Endiandra*, *Eusideroxylon*, *Potameia*, *Sinopora*, and *Syndiclis*. However, a recent study by Ariati et al. (2023) concluded that our sequenced Belian individual might be classified within the Myristicaceae clade instead of Lauraceae [24]. We were particularly surprised by the result.

Here, we assembled two new complete chloroplast genomes of Belian and reanalyzed them alongside previously published one, including genome characterization, codon usage, repeat sequences, IR boundaries, mutational events, and nucleotide polymorphism (Pi) analysis. Finally, 44 plastomes from 10 magnoliids families were used to reconstruct the phylogenetic trees. This study was conducted with the following specific objectives: (a) to investigate the general features of the Belian chloroplast genome; (b) to explore the genome differences among the three Belian individuals through comparative genomic analysis; and (c) two newly assembled Belian chloroplast genomes were combined with previously published Belian genomes to construct phylogenetic trees and identify phylogenetic locations. In previous studies, the structure of the Belian chloroplast genome was not analyzed. Our results will deepen our understanding of the Belian chloroplast genome and provide a good basis for the conservation and utilization of genetic resources. Through comparative genomic analysis, the differences between Belian individuals were explored, which provided conditions for further study of Belian genetic variation and ecological adaptation. By reconstructing the phylogenetic relationship, the Belian phylogenetic location was revealed and the phylogenetic relationship of Ariati et al. (2023) was verified [24].

2. Materials and Methods

2.1. Plant Materials, Extraction, and Sequencing of DNA

Two accessions of fresh Belian leaves were collected in Java, Indonesia and Kalimantan, Malaysia (Figures 1 and S1). Belian II (SY6477) was collected in Sulawesi, Indonesia and published in NCBI in 2017, and Ariati et al. (2023) used this sequence [24]. The Xishuangbanna Tropical Botanical Garden's Herbarium, part of the Chinese Academy of Sciences, is home to the gathered plant specimens. Using a modified version CTAB approach, 2 g of leaves was used to extract total DNA [25]. Before sequencing, 0.5 μ g of pure DNA was fragmented to create short-insert (500 bp) libraries according to the manufacturer's specifications (Illumina, San Diego, CA, USA). DNA samples were tagged, pooled together, and sequenced in one lane of a genome analyzer (Illumina HiSeq 2000) at BGI-Shenzhen.

2.2. Annotation and Assembly of the Chloroplast Genome

Illumina paired-end sequencing generated 2.93 Gb and 3.17 Gb raw reads of 150 bp in length for Belian I and Belian III, respectively. The total number of reads in Belian I was 40,221,884; that in Belian III was 47,514,230. We applied stringent sequence filtering with an NGS QC Tool Kit version 2.3.3 [26] to select clean reads. A total of 32,840,077 clean reads were produced in Belian I and 39,620,935 clean reads in Belian III. The kmer-coverage of Belian I was 119.7, and the kmer-coverage of Belian III was 121.7. Following the filtering of the sequencing data, the chloroplast genomes were automatically assembled using GetOrganelle version 1.7.5 [27]. Bandage version 0.9.0 [28] was used to identify the circular maps to assess the assembly quality. The whole chloroplast genome sequence of *Cryptocarya chinensis* (GenBank accession No. LC212965) was used as a reference, automatically annotated using CPGAVAS2 (https://www.herbalgenomics.org/cpgavas2, accessed on 29 August 2023) [29]. Start and stop codons, as well as intron/exon borders of protein-coding genes, were manually verified. OGDRAW (https://chlorobox.mpimp-



golm.mpg.de/OGDraw.html, accessed on 29 August 2023) [30] was used to create a circular chloroplast genome map.

Figure 1. Map of distribution and sample sites of Belian. The orange circle represents Belian I, located in Java, Indonesia; the green circle represents Belian II in Kalimantan, Malaysia; the purple circle represents Belian III located in Sulawesi, Indonesia; the blue circles represent the distribution of Belian. The dash line allows us to observe the distance between Belian I and Belian II and the distance between Belian I and Belian III.

2.3. Chloroplast Genome Analysis

All protein-coding sequences from the three Belian chloroplast genomes were extracted using PhyloSuite version 1.2.3 [31]. Relative synonymous codon usage (RSCU) of three chloroplast genomes was calculated using CodonW version 1.4.2 (https://codonw. sourceforge.net, accessed on 30 August 2023). The repeat sequences were predicted using the online program REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer, accessed on 2 September 2023) [32]. The parameters were set as Maximum computed repeats = 5000, Minimal repeatsize = 25, and Hamming distance = 3. Simple repeat sequences (SSRs) were predicted using the online software MISA (https://webblast.ipk-gatersleben.de/misa, accessed on 2 September 2023) [33], which included mono-nucleotide, di-nucleotide, trinucleotide, tetra-nucleotide, penta-nucleotide, and hexa-nucleotide, with repeat thresholds set at 10, 5, 4, 3, 3, and 3, respectively.

2.4. Comparative Sequence Analysis

CPJSdraw version 1.158 [34] was used to analyze the boundary differences of IR regions of the chloroplast genome and draw a comparison map. The microstructural mutations of the sequences were detected in Genenious Prime version 2023.2.1 [35] and manually examined in BioEdit version 7.0.9 [36], especially for the inversion sites. We performed a sliding window analysis using DnaSP version 6.12 [37] to evaluate the variability (Pi) across the plastomes. The window length was set to 600 bp, and the step size was set as 200 bp.

2.5. Phylogenetic Analysis

A total of 48 chloroplast genomes from 10 families of the Magnoliids and 5 species of Chloranthaceae were aligned using MAFFT version 7 (https://mafft.cbrc.jp/alignment/ server/, accessed on 26 September 2023) [38], of which 39 sequences were downloaded from the NCBI and 9 were from LCGDB (Supplementary Table S1). The complete chloroplast genome matrix was obtained and manually adjusted by BioEdit version 7.0.9 [36]. Phylogenetic relationships were reconstructed based on the maximum-likelihood (ML) method in IQ tree version 2.2.2 [39] and the Bayesian inference (BI) method in MrBayes version 3.2.7 [40]. The best model for the ML analysis was calculated using IQ Tree version 2.2.2 to obtain the best model under the AIC as GTR + F + R8, and 1000 bootstrap replicates were run to provide confidence support. Using jModelTest version 2.1.10 [41] software, the whole chloroplast genome dataset was examined for BI analysis, and the best TVM + I + G model was chosen. The combined data were run for 1 million generations, sampling every 1000 generations. The first 25% of the tree was discarded as burn-in, and the remaining trees were used to generate a majority-rule consensus tree. In all phylogenetic analyses, Chloranthus erectus (GenBank accession No. MH394412), Chloranthus spicatus (GenBank accession No. EF380352), Chloranthus henryi (GenBank accession No. MK922064), and Chloranthus japonicus (GenBank accession No. KP256024) were used as the outgroup. All phylogenetic trees were visualized or further edited in Fig Tree version 1.4.4 (http://tree.bio.ed.ac.uk/publications/, accessed on 27 September 2023).

3. Results

3.1. General Features of the Belian Chloroplast Genomes

The newly assembled two plastomes had quadripartite structures forming circular molecules, and the size of both genomes was 157,535 bp of Belian I and 157,536 bp of Belian III (Figure 2), respectively. They were 41 bp and 42 bp more minor than that of Belian II. Both genomes included a pair of inverted repeats (IRs) of 24,717 bp in Belian I and 24,706 bp in Belian III, separated by a small single copy (SSC) region of 18,912 bp in Belian I and 18,916 bp in Belian III, and a large single copy (LSC) region of 89,189 bp in Belian II and 89,208 bp in Belian III (Table 1). Their GC content, similar to that of Belian II, was 39%. As predicted, we annotated 130 genes each in 2 plastids, including 113 unique genes and 17 duplicates in the IR regions (Supplementary Table S2). Among these genes, there were 85 PCGs, 37 tRNA genes, and 8 rRNA genes (Figure 2). Introns were also found in 12 protein-coding genes and 8 tRNA genes (Supplementary Table S2).

Table 1. Summary of 3 Complete Plastomes of Belian.

	Belian I	Belian II	Belian III
Sequencing no.	SY5992	SY6477	SY6583
Coordination and a second second	Java,	Sulawesi,	Kalimantan,
Geographic origin	Indonesia	Indonesia	Malaysia
Collector	TAN Yunhong	Michael Padmanaba	ZHOU Xinxin
Voucher	Y1019	SY34806	SY37637
Accession no.	LAU00162	MF939381	LAU00163
Herbarium	HITBC-BRG	HITBC-BRG	HITBC-BRG
Total cpDNA size (bp)	157,535	157,577	157,536
Length of LSC region (bp)	89,189	89,231	89,208
Length of IR region (bp)	24,717	24,717	24,706
Length of SSC region (bp)	18,912	18,912	18,916
Total GC content	39.06%	39.05%	39.05%
LSC	37.75%	37.74%	37.74%
IR	43.24%	43.23%	43.24%
SSC	34.34%	34.32%	34.34%
Total number of genes (unique)	130 (113)	130 (113)	130 (113)
Protein encoding	85	85	85
tRNA	37	37	37
rRNA	8	8	8



other genes
 hypothetical chloroplast reading frames (ycf)

Figure 2. Gene map of the Belian plastomes. Genes drawn inside the circle are transcribed clockwise, whereas genes drawn outside the circle are transcribed anticlockwise. Different colored bars represent genes with different functions. The plastome's GC content is indicated by the dashed dark gray region in the inner circle, while the AT content is displayed by the light gray area.

3.2. Codon Usage

The relative synonymous codon usage (RSCU) of the three Belian chloroplast genomes was calculated using CodonW 1.4.2, and a total of 78,289 codons were detected, averaging 26,096 codons per individual. All the codons are divided into 64 types, coding for 20 amino acids. There were 31 high-frequency codons with RSCU > 1, comprising 13 codons ending in A and 16 ending in U, accounting for 93.55% (Figure 3). The codons AGU and UGG had an RSCU value of 1, indicating an unbiased usage. RSCU > 1.6 was found in codons GCU, CCA, UUA, and AGA, suggesting that they appeared more frequently and were overused. Leucine, arginine, and serine were encoded by six codons, while methionine and tryptophan were encoded by only one codon each. The rest of the amino acids were encoded by two to four codons.



Figure 3. Codon usage patterns of Belian plastomes. The *y*-axis represents the relative synonymous codon usage, whereas the *x*-axis represents the codons. The symbol "*" is the stop codon.

3.3. Repeat Sequences and Simple Sequence Repeats Analysis

We identified 95 to 101 long repeat sequences (>25 bp) in three Belian chloroplast genomes using REPuter, including 43–45 palindromic repeats, 37–41 forward repeats, 10–12 reverse repeats, and 4–5 complement repeats (Figure 4A). Palindromic repeats were the most common in the genomes of the three individuals, followed by forward repeats and complement repeats. Most of the repeats ranged from 25 to 29 bp in length and with nine repeats being 40 to 49 bp in length. In addition, the reverse repeats and complement repeats were only detected in the range of 25–29 bp, but not in the range of 30–39 bp and 40–49 bp (Figure 4B).



Figure 4. The number of long repeats in three Belian plastomes. (**A**) The number of different repeat types. (**B**) The number of different length ranges. The letters "C" stand for complementary repeat type, "R" for reverse repeat type, "P" for palindromic repeat type, and "F" for forward repeat type.

Simple repeat sequences (SSRs) are widely distributed in the chloroplast genome. In all, 176 SSRs were found in the three Belian chloroplast genomes, with an average of 59 per individual. Mono-nucleotides were the most numerous, accounting for 73.1% to 74.58% of the total, followed by five di-nucleotides, two tri-nucleotides, eight tetra-nucleotides, and one penta-nucleotide; no hexa-nucleotides were detected (Figure 5). More than 90% of mono-nucleotides belong to A or T base repeats. SSRs are not equally distributed across the genome. There were 48/50/45, 7, and 4 SSR repeats distributed in LSC, SSC, and IRs of the three Belian genomes, respectively. The variability in SSR numbers among the three individuals was primarily observed in the LSC region, while the counts in the SSC and IR regions remained consistent.



Figure 5. Number and distribution of SSR motifs in the three Belian genomes. The bar chart shows different SSR motif types, and the pie chart shows the distribution of SSR in LSC, SSC, and IR regions.

3.4. Inverted Repeat Contraction and Expansion

Differences in IR boundaries of chloroplast genomes of *Beilschmiedia turbinate* (LAU00039), *Cryptocarya densiflora* (LAU00050), *Endiandra dolichocarpa* (LAU00053), *Syndiclis chinensis* (LAU00155), and three Belian individuals were compared using CPJSdraw version 1.158. The genes that link the IR boundaries are similar in them. LSC/IRb(JLB), IRb/SSC(JSB), SSC/IRa(JSA), and IRa/LSC(JLA) boundary junctions are mainly associated with six genes, *rsp19, rpl23, ycf1, ndhF, trnN*, and *rpl2*. The locations of IR boundary genes in the three Belian individuals were basically the same, but the linking genes at the IR boundary between the other four species were different (Figure 6). The *ycf1* gene fragment at LSC/IRb(JLB) showed great changes, ranging in length from 1011 to 1137 bp, and also had great changes in location.



Figure 6. Comparison of the boundary patterns of the large single-copy (LSC) region, small single-copy (SSC) region, and inverted repeat (IR) regions among *Beilschmiedia turbinate* (LAU00039), *Crypto-carya densiflora* (LAU00050), *Endiandra dolichocarpa* (LAU00053), *Syndiclis chinensis* (LAU00155), and three Belian chloroplast genomes.

3.5. Mutations in the Belian Chloroplast Genomes

Between the Belian I and Belian III plastomes, we found 143 mutation events, including 5 microinversions, 40 InDels, and 98 substitutions. Compared with Belian II, we accurately located 6 microinversions, 53 InDels, and 111 substitutions in these Belian plastomes. Among the SNPs, 46 were located in the gene-coding regions, which included 21 transitions (Ts) and 25 transversions (Tv), while 65 were found in non-coding regions, comprising 27 Ts and 38 Tv (Supplementary Table S3). The transition/transversion ratio (Ts/Tv) was calculated to be 0.76. Of the 53 InDels, 40 were in the LSC region, 13 in the SSC region, and 4 in the IR regions. The sizes of all InDels varied from 1 to 15 bp. There were 50 in intergenic regions, 3 in the gene-coding regions, and none in introns. The greatest InDel, 15 bp in size, was found in the *atpF-atpH* intergenic region. There were also four microinversions in the *psbC-trnS*, *petA-psbJ*, *rrn5S-trnR*, and *ccsA-ndhD* intergenic areas, as well as two microinversions in the *rpl16* and *ycf1* coding regions.

SNP and InDel sites in Belian II and Belian III were detected using Belian I as a reference, and their densities in each region were calculated (Table 2). Belian II possesses 31 SNPs (5 Ts and 26 Tv) and 20 InDels, while Belian III holds 98 SNPs (44 Ts and 54 Tv) and 40 InDels. The number of SNPs in Belian III was 3.16 times that of Belian II, and the SNP count was double that of Belian II. Compared with the other two regions, the SSC region of Belian II had the highest density of SNP and InDel. The highest SNP density was observed in the LSC regions of Belian III, whereas the highest InDel density was identified in the SSC region. Across all samples, the density of SNP and InDel in the IR regions is low.

	Туре	Total	LSC Density/kb	SSC Density/kb	IR Density/kb	Total Density/kb
Belian II	SNP	31 (Ts5, Tv26)	0.213	0.529	0.04	0.197
Belian III	SNP	98 (Ts44, Tv54)	0.874	0.74	0.121	0.622
Belian II	InDel	20	0.108	0.159	0	0.127
Belian III	InDel	40	0.325	0.37	0.081	0.254

Table 2. Density statistics of SNPs/InDels region.

3.6. Nucleotide Diversity (Pi) Analysis

The nucleotide diversity (Pi) values of three Belian plastomes were calculated using DNAsp version 6.12 software. In the three Belian plastomes, Pi values ranged from 0 to 0.0044, with a mean of 0.0005. A total of 35 variable loci (Pi > 0.002) were identified (Figure 7), of which 10 were hypervariable regions (Pi > 0.003), namely *trnK*, *trnK*–*rps16*, *psbM*–*trnD*, *trnT*–*psbD*, *psaA*, *trnF*, *trnM*–*atpE*, *atpE*, *ndhF*, and *ndhF*–*rpl32*. Among these variable loci, eight were located in the LSC, two in the SSC, and none in the IRs. The value of *psbM*–*trnD* was the highest (Pi > 0.004). These 10 variable loci were all found in the LSC and SSC regions and were ideal candidates for phylogenetic study.



Figure 7. Nucleotide variability (Pi) values of Belian (Lauraceae). Line segments parallel to the *x*-axis include the LSC, SSC, and IR regions; InDels are shown by black circles, while SNPs are indicated by red circles.

3.7. Phylogenetic Analysis

To determine the phylogenetic location of Belian in Lauraceae, the phylogenetic relationships among 2 Belian plastomes and 42 plastomes from 10 families of the Magnoliids were reconstructed based on the chloroplast genomes, with *Chloranthus erectus*, *C. spicatus*, *C. henryi*, and *C. japonicus* as the outgroup. The topologies, using the Bayesian inference (BI) and maximum-likelihood (ML) methods, were nearly identical and both well supported (Figure 8). Overall, the 44 samples were divided into 10 branches corresponding to Lauraceae, Hernandiaceae, Calycanthaceae, Myristicaceae, Annonaceae, Magnoliaceae, Saururaceae, Piperaceae, Aristolochiaceae, and Winteraceae (Figure 8). The three plastomes of Belian were clustered into one branch in the Lauraceae clade with high support (ML-BS = 100%, BI-PP = 1.0).





100/1

0.03

100/1.0

100/1.0

100/1.0

100/1.0

100/1.0

Figure 8. Phylogenetic tree of 44 species of magnoliids based on complete plastome sequences by maximum likelihood (ML) and Bayesian inference (BI) with *Chloranthus erectus, C. spicatus, C. henryi,* and *C. japonicus* as the outgroup. Numbers at each node are the maximum-likelihood bootstrap support/Bayesian posterior probabilities. Ten families of species are represented in five colors. The red font represents the three Belians, and the blue font is the species published by Ariati et al. (2023) [24].

In addition, we compared the *trnL–trnF* (GenBank accession No. AF268718), *psbA–trnH* (GenBank accession No. AF268820), and *rpl16* (GenBank accession No. AF268252) sequences used by Chanderbali et al. (2001) and the *trnK* (GenBank accession No. AJ627926) sequence used by Rohwer and Rudolph (2005) with the Belian sequence (GenBank accession

No. MF939351) [2,18]. The *trnL–trnF* sequence was identical, and the *psbA–trnH*, *rpl16*, and *trnK* sequences contained 1 to 3 bp mutations. When we compared the genomes of the three Belian, we found similar mutations in these three regions. Furthermore, we constructed the phylogenetic tree using the same methods and sequences as in the Ariati et al. (2023) study and added two additional Belian sequences [24]. We obtained a different topological structure than they did. The three Belians clustered with *Cinnamonum camphora*, *litsea coreana*, and *litsea auriculata*, all located in the Lauraceae family, not Myristiaceae (Supplementary Figure S2). We also constructed a phylogenetic tree based on the complete genome using ML and BI methods to obtain the same result.

4. Discussion

In this study, the chloroplast genomes of two Belians from Java and Kalimantan were newly assembled and compared with another previously published one. The chloroplast genome of Belian had a typical quadripartite structure with a size of 157,535–157,577 bp, containing a total of 130 genes, including 85 CDS genes, 37 tRNA genes, and 8 rRNA genes, which was similar to the chloroplast genomes of other Lauraceae species [42,43]. The total length of the chloroplast genomes of the three Belians varied, and there were differences in the length of the LSC, SSC, and IR regions. InDel mutations for these size differences were identified through comparative genomics analysis. For example, in the intergenic region *atpB–rbcL*, Belian II inserted a GATGTAC repeat fragment; in *trnH–psbA*, Belian II inserted an AAATG fragment; in *atpF–atpH*, Belian II inserted TTAATATTAATTTCC; and in *trnE–trnT*, Belian II inserted CTATG and T. These InDel mutations affect not only the length of the genome but also the GC content of the genome together with SNP mutations. Nurtjahjaningsih et al. (2017) found that Belian presents high genetic diversity at the chloroplast DNA level in Indonesia and that geographical distance does not always correlate with Belian's phylogenetic distance [4]. They believe that human activity has affected the genetic structure of Belian populations in Indonesia [4]. Our results support the view that Belian I and Belian II are closely related, but they are spatially distant (Figures 1 and 8), both of which are distributed in Indonesia. Therefore, we speculate that there may be two reasons for the size differences in the three Belian chloroplast genomes. The first was human activity, which accelerated the spread of Belian seeds and facilitated the exchange of genes due to their large and heavy size and slow spread. The second may be related to the history of Belian dispersal. We look forward to collecting more samples and studying the genetic structure of the Belian population. In addition, Song et al. (2017) found that the chloroplasts of the core Lauraceae were 150,749 bp to 152,739 bp in length, with trnI–CAU, rpl23, rpl2, and ycf2 fragments and their intergenic regions lost in the IRb region, and that the chloroplasts of the basal Lauraceae were 157,577 bp to 158,530 bp in length, with *rpl2* lost in the IRa region [21], which is supported by the results of the present study.

Codon preference significantly varies among different species and different genes in the same species. The utilization of biased codons is a significant indication in studying species evolution [44]. Relative synonymous codon usage (RSCU) is an important index to study codon usage preference. A total of 78,289 codons were detected in the proteincoding genes of the three Belian chloroplast genomes, with 26,095–26,099 codons detected in each genome. There are 31 codons with high relative synonymous usage with RSCU > 1, 29 of which end in A/U, indicating that the chloroplast genome of Belian is more inclined to employ codons ending in A or U, which is consistent with the results of previous studies [45].

Although the chloroplast plastids of most land plants are generally similar in size, content, and structure, certain plant species have developed significantly rearranged chloroplast plastids. Inversions, internal inversions, and IR boundary shifts are the principal causes of such structural alterations, with repeat sequences playing a key role in the evolution and rearrangement of chloroplast genomes [46,47]. In the chloroplast genome of Belian, the number of repeats with a length of 25–29 bp was the largest, followed by 30–39 bp. The

majority of the repeats were palindromic repeats and forward repeats, with reverse repeats and complement repeats primarily discovered in the range of 25 bp to 29 bp and in a small number, which is similar to the results of avocados (*Persea americana*) [48]. Furthermore, the repeat distribution in Belian chloroplast genomes were unbalanced, with the most repeats in the LSC region and fewer repeats in the SSC and IR regions. This unbalanced distribution may be related to the distribution of chloroplast genes, where genes related to photosynthesis were predominantly located in the LSC region, while rRNAs were all located in the IR region.

Simple sequence repeats in chloroplast genomes are highly variable among different species within the same genus, play an important role in the identification of plant genetic relationships and taxonomic status, and are considered to be one of the main sources of molecular markers [49]. A total of 176 SSRs were detected in the three Belian chloroplast genomes, with 56 to 61 SSRs detected in each Belian individual. The number of mono-nucleotide repeats was the largest, predominantly consisting of A/T repeats, while the di-nucleotide repeats were TA, AT, AG, and CT. In the genomes of the three Belians, the number of tetra-nucleotide repeats was higher than the tri-nucleotide repeats, and the penta- and hexa-nucleotides were very less. Similar results were seen in *Pinus* genomes [50]. The preference of SSRs to be rich in A/T bases may be due to the fact that there are only two hydrogen bonds between A/T, whereas there are three hydrogen bonds between G/C, so it is more difficult to break the G/C bond to produce mutations [51].

The IR region is the most conserved region in the chloroplast genomes; however, the expansion and contraction of the IR region boundary is a common evolutionary phenomenon, which is the primary mechanism leading to the change in the chloroplast genome size [52,53]. By comparing the differences in the chloroplast genome IR boundaries of *Beilschmiedia turbinate, Cryptocarya densiflora, Endiandra dolichocarpa, Syndiclis chinensis,* and the three Belian individuals, it was found that the boundaries of the three Belian individuals, it reaffirms the highly conserved nature of the IR region. In addition, it has been demonstrated that the *ycf1* and *ycf2* genes are located at the junction of the IR regions with the LSC and SSC regions, and these two genes have partial duplication [42,54]. In this study, only the *ycf1* gene occurs at the junction of the SSC and IRa, and the *ycf1* gene fragment occurs at the junction of the LSC and IRb (Figure 6). This result was also seen in the genus *Atractylodes* [55].

SNP and InDel are essential sources of genetic variation, leading to differences in gene structure, which reflect the adaptability of individuals to environmental changes [56]. In this study, we examined mutations in two other Belian chloroplast genomes using Belian I as a reference sequence, and 170 mutation events were identified, namely 6 microinversions, 53 InDels, and 111 substitutions. The distribution of mutation events in Belian is similar to that in other angiosperms; most of them are located in the non-coding region; the number of mutation events is the highest in the LSC region, and the number of mutation events is the least or even none in the IR regions. However, the density of SNP and InDel in each region of Belian II and Belian III showed that the maximum SNP and InDel density in the SSC region of Belian II was 5.56/kb, and the maximum density of the LSC region of Belian III was 8.43/kb. The number of SNP and InDel in Belian is consistent with the level of the same species. Muraguri et al. (2020) identified a total of 162 SNPs and 92 InDels in the chloroplast genomes of 12 individuals from *Ricinus communis* [57]; Zhang et al. (2020) detected a total of 77 SNPs and 255 InDels in 3 individuals of *Quercus acutissima* [58]. We note that there were 53 mutation events in Belian II and 143 mutation events in Belian III, 3.33 times as many as in Belian II. The mutations of Belian II and Belian III in CDS are mainly in the form of SNP, and Belian III contains most of the mutations in Belian II. In several CDS, Belian III has more mutation sites, such as *accD* and *ycf*2. These may be some evidence of Belian III adaptation to the environment.

Chloroplast genomes contain highly variable regions that help to distinguish closely related species or genera and are considered potential molecular marker material for phylogenetic analyses [59]. In the three Belian genomes, 35 variable loci (Pi > 0.002) were identified by DnaSP version 6.12, of which 10 were hypervariable regions (Pi > 0.003), namely *trnK*, *trnK–rps16*, *psbM–trnD*, *trnT–psbD*, *psaA*, *trnF*, *trnM–atpE*, *atpE*, *ndhF*, and *ndhF–rpl32*. The highly variable region *petA–psbJ* is considered a hotspot of variation in *Neocinnamomum* [60] and *Litsea glutinosa* [61], but not in Belian. The IR region has lower genetic polymorphism than the LSC and SSC regions, and the coding region is more conserved than the non-coding region in Belian chloroplast genomes. Song et al. (2015) compared *Machilus yunnanensis* and *Machilus balansae* and identified 297 mutation sites, with Pi values ranging from 0 to 0.0133 and an average of 0.00154 [42]. Song et al. (2017) compared the chloroplast genomes of *Phoebe sheareri* and *Phoebe omeiensis* and identified 222 mutation sites with Pi values ranging from 0 to 0.0083 with an average of 0.0010 [62]. However, Pi values in the three Belian plastomes ranged from 0 to 0.0044, with an average value of 0.0005, which was very low.

The chloroplast genome is essential for the study of phylogenetic relationships and species identification of angiosperms and for determining their taxonomic status [63,64]. Ariati et al. (2023) constructed phylogenetic trees using the chloroplast CDS of Belian II (GenBank accession No. MF939351) and 20 species of Myristicaceae using ML and BI methods, with Uvaria littoralis as the outgroup [24]. The results showed that Belian II belonged to Endocomia macrocoma in Myristicaceae. Both Lauraceae and Myristiaceae belong to Magnoliids. In Myristicaceae, all genera are dioecious, except Endocomia and some Iryanthera. We need to focus on the morphological difference between Belian and Endocomia macrocoma. The seeds of Belian are the largest dicotyledonous species ever recorded, measuring about 14 cm long (Supplementary Figure S1), weighing about 230 g, and shaped like a rugby ball, while the seeds of Endocomia macrocoma have a red aril and a fruit measuring about 3.6 cm long [65]. It also has a stem with watery, reddish sap. In this study, the phylogenetic relationships among 2 Belian plastomes and 42 plastomes from 10 families of Magnoliids were reconstructed based on the chloroplast genomes. The topology of Lauraceae clades is similar to that of previous studies, especially the location of Belian, which is located at the base of Lauraceae [2,18,21,23]. Therefore, there is morphological and molecular evidence that the sequence (GenBank accession No. MF939351) we published was not misidentified.

5. Conclusions

The present study sequenced, assembled, and annotated two Belian chloroplast genomes from Java and Kalimantan and analyzed them with another previously published Belian genome from Sulawesi. Comparative analysis revealed that the Belian genomes from three different locations were conserved in terms of gene content, gene sequence, and GC content. The rapidly evolving differentiation regions, repeats, and mutation sites identified in this study may serve as potential molecular markers for phylogenetic studies. The location of Belian in Lauraceae was determined based on the whole chloroplast genome sequence, which further confirmed the placement of our previously published Belian sequence within Lauraceae, not Myristicaceae. In summary, our study has deepened our understanding of the Belian chloroplast genome and provided a foundation for taxonomic identification, phylogenetic studies, and conservation of genetic resources.

Supplementary Materials: The following supporting information can be downloaded at: https: //www.mdpi.com/article/10.3390/f14122443/s1, Figure S1: Sample images of Belian. (A) Belian seeds and young leaves. (B) Mature leaves of Belian. (A) and (B) were collected from Sulawesi. (C) Collection site marking. (D) Belian III leaves. (C) and (D) were collected from Kalimantan.; Figure S2: The phylogenetic tree based on complete chloroplast sequences and 75 protein-coding genes with ML and BI methods, including 24 species in the Magnoliids; Table S1: Plastome sequences obtained from NCBI and LCGDB for this study.; Table S2: Genes present in the Belian chloroplast genome; Table S3: Mutation present in the Belian chloroplast genome. **Author Contributions:** Y.S. and P.X. conceived and designed the experiments. Y.S. and P.X. revised the manuscript. W.Z. performed the data analysis and drafted the manuscript. Y.T. and X.Z. contributed materials. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Nos. 32260060 and 32060710), Key Technologies Research for the Germplasm of Important Woody Flowers in Yunnan Province (No. 202302AE090018), and Yunnan Science and Technology Talents and Platform Program (No. 202205AF150022).

Data Availability Statement: The complete plastid genome sequence data of the 2 Belian individuals will be submitted to LCGDB (https://lcgdb.wordpress.com, accessed on 23 November 2023) through the revision process.

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

References

- The Angiosperm Phylogeny Group; Chase, M.W.; Christenhusz, M.J.M.; Fay, M.F.; Byng, J.W.; Judd, W.S.; Soltis, D.E.; Mabberley, D.J.; Sennikov, A.N.; Soltis, P.S.; et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. 2016, 181, 1–20. [CrossRef]
- 2. Chanderbali, A.S.; van der Werff, H.; Renner, S.S. Phylogeny and Historical Biogeography of Lauraceae: Evidence from the Chloroplast and Nuclear Genomes. *Ann. Mo. Bot. Gard.* **2001**, *88*, 104–134. [CrossRef]
- 3. Song, Y.; Yu, W.B.; Tan, Y.H.; Jin, J.J.; Wang, B.; Yang, J.B.; Liu, B.; Corlett, R.T. Plastid phylogenomics improve phylogenetic resolution in the Lauraceae. *J. Syst. Evol.* **2019**, *58*, 423–439. [CrossRef]
- 4. Nurtjahjaningsih, I.L.G.; Sukartiningsih; Kurokochi, H.; Saito, Y.; Ide, Y. Genetic Structure of the Tropical Tree *Eusideroxylon zwageri* in Indonesia Revealed by Chloroplast DNA Phylogeography. *Forests* **2017**, *8*, 229. [CrossRef]
- Asian Regional Workshop. Conservation & Sustainable Management of Trees, Viet Nam, August 1996. Eusideroxylon zwageri. The IUCN Red List of Threatened Species. 1998. Available online: https://www.iucnredlist.org/species/31316/9624725 (accessed on 29 September 2023).
- 6. Irawan, B.; Gruber, F.; Finkeldey, R.; Gailing, O. Linking indigenous knowledge, plant morphology, and molecular differentiation: The case of ironwood (*Eusideroxylon zwageri* Teijsm. et Binn.). *Genet. Resour. Crop Evol.* **2015**, *63*, 1297–1306. [CrossRef]
- Daniell, H.; Lin, C.S.; Yu, M.; Chang, W.J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* 2016, 17, 134. [CrossRef] [PubMed]
- Jansen, R.K.; Raubeson, L.A.; Boore, J.L.; dePamphilis, C.W.; Chumley, T.W.; Haberle, R.C.; Wyman, S.K.; Alverson, A.J.; Peery, R.; Herman, S.J.; et al. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods Enzym.* 2005, 395, 348–384. [CrossRef]
- 9. Palmer, J.D. Comparative organization of chloroplast genomes. Annu. Rev. Genet. 1985, 19, 325–354. [CrossRef]
- Daniell, H.; Lee, S.B.; Grevich, J.; Saski, C.; Quesada-Vargas, T.; Guda, C.; Tomkins, J.; Jansen, R.K. Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theor. Appl. Genet.* 2006, 112, 1503–1518. [CrossRef]
- 11. Walker, T. Plant Diversity and Evolution. Genotypic and Phenotypic Variation in Higher Plants. *Exp. Agric.* **2006**, *42*, 121–126. [CrossRef]
- 12. Palmer, J.D.; Nugent, J.M.; Herbon, L.A. Unusual structure of geranium chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 769–773. [CrossRef] [PubMed]
- Dyall, S.D.; Brown, M.T.; Johnson, P.J. Ancient invasions: From endosymbionts to organelles. *Science* 2004, 304, 253–257. [CrossRef]
 [PubMed]
- Clegg, M.T.; Gaut, B.S.; Learn, G.H., Jr.; Morton, B.R. Rates and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci. USA* 1994, 91, 6795–6801. [CrossRef] [PubMed]
- Raubeson, L.A.; Peery, R.; Chumley, T.W.; Dziubek, C.; Fourcade, H.M.; Boore, J.L.; Jansen, R.K. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms Nuphar advena and Ranunculus macranthus. *BMC Genom.* 2007, *8*, 174. [CrossRef] [PubMed]
- Aiso-Sanada, H.; Nezu, I.; Ishiguri, F.; Jaffar, A.N.N.B.M.; Ambun, D.B.A.; Perumal, M.; Wasli, M.E.; Ohkubo, T.; Abe, H. Basic wood properties of Borneo ironwood (*Eusideroxylon zwageri*) planted in Sarawak, Malaysia. *Tropics* 2020, 28, 99–103. [CrossRef]
- 17. Kimoto, Y.; Utami, N.; Tobe, H. Embryology of *Eusideroxylon* (Cryptocaryeae, Lauraceae) and character evolution in the family. *Bot. J. Linn. Soc.* **2006**, *150*, 187–201. [CrossRef]
- 18. Rohwer, J.G.; Rudolph, B. Jumping Genera: The Phylogenetic Positions of *Cassytha*, *Hypodaphnis*, and *Neocinnamomum* (Lauraceae) Based on Different Analyses of *trnK* Intron Sequences. *Ann. Mo. Bot. Gard.* **2005**, *92*, 153–178.
- 19. Kurokochi, H.; Nurtjahjaningsih, I.L.G.; Sukartiningsih; Tan, E.; Asakawa, S.; Saito, Y.; Ide, Y. Development of polymorphic chloroplast DNA markers for the endangered tree *Eusideroxylon zwageri* through chloroplast isolation and next-generation sequencing. *Conserv. Genet. Resour.* **2015**, *7*, 845–850. [CrossRef]

- 20. Md-Isa, S.F.; Yien Yong, C.S.; Saleh, M.N.; Go, R. An assessment of genetic variation in vulnerable Borneo Ironwood (*Eusideroxylon zwageri* Teijsm. & Binn.) in Sarawak using SSR markers. *J. Threat. Taxa* **2021**, *13*, 18588–18597. [CrossRef]
- Song, Y.; Yu, W.B.; Tan, Y.; Liu, B.; Yao, X.; Jin, J.; Padmanaba, M.; Yang, J.B.; Corlett, R.T. Evolutionary Comparisons of the Chloroplast Genome in Lauraceae and Insights into Loss Events in the Magnoliids. *Genome Biol. Evol.* 2017, *9*, 2354–2364. [CrossRef]
- 22. Liu, Z.-F.; Ma, H.; Ci, X.-Q.; Li, L.; Song, Y.; Liu, B.; Li, H.-W.; Wang, S.-L.; Qu, X.-J.; Hu, J.-L.; et al. Can plastid genome sequencing be used for species identification in Lauraceae? *Bot. J. Linn. Soc.* **2021**, *197*, 1–14. [CrossRef]
- Yang, Z.; Ferguson, D.K.; Yang, Y. New insights into the plastome evolution of Lauraceae using herbariomics. *BMC Plant Biol.* 2023, 23, 387. [CrossRef] [PubMed]
- 24. Ariati, S.R.; Priyadi, A.; Hariri, M.R.; Risna, R.A. The complete chloroplast genome of *Myristica teysmannii* (Myristicaceae), an endemic and endangered species from Indonesia. *J. Asia-Pac. Biodivers.* **2023**, *16*, 734–738. [CrossRef]
- 25. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 1987, 19, 11–15.
- Patel, R.K.; Jain, M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE* 2012, 7, e30619. [CrossRef] [PubMed]
- Jin, J.J.; Yu, W.B.; Yang, J.B.; Song, Y.; dePamphilis, C.W.; Yi, T.S.; Li, D.Z. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 2020, 21, 241. [CrossRef]
- Wick, R.R.; Schultz, M.B.; Zobel, J.; Holt, K.E. Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics* 2015, *31*, 3350–3352. [CrossRef]
- Shi, L.; Chen, H.; Jiang, M.; Wang, L.; Wu, X.; Huang, L.; Liu, C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 2019, 47, W65–W73. [CrossRef]
- 30. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64. [CrossRef]
- Zhang, D.; Gao, F.; Jakovlic, I.; Zou, H.; Zhang, J.; Li, W.X.; Wang, G.T. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* 2020, 20, 348–355. [CrossRef]
- 32. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 2001, 29, 4633–4642. [CrossRef] [PubMed]
- Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 2003, 106, 411–422. [CrossRef] [PubMed]
- 34. Li, H.; Guo, Q.; Xu, L.; Gao, H.; Liu, L.; Zhou, X. CPJSdraw: Analysis and visualization of junction sites of chloroplast genomes. *PeerJ* 2023, 11, e15326. [CrossRef] [PubMed]
- Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012, 28, 1647–1649. [CrossRef]
- Hall, T.A. Bioedit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/ nt. Nucleic Acids Symp. Ser. 1999, 41, 95–98.
- Rozas, J.; Ferrer-Mata, A.; Sanchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sanchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* 2017, *34*, 3299–3302. [CrossRef] [PubMed]
- Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 2019, 20, 1160–1166. [CrossRef]
- Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 2015, *32*, 268–274. [CrossRef]
- Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Hohna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 2012, *61*, 539–542. [CrossRef]
- 41. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [CrossRef]
- Song, Y.; Dong, W.; Liu, B.; Xu, C.; Yao, X.; Gao, J.; Corlett, R.T. Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. *Front. Plant Sci.* 2015, 6, 662. [CrossRef]
- 43. Bai, X.; Peng, J.; Yang, Y.; Xiong, B. The Complete Chloroplast Genome Sequence of *Machilus chuanchienensis* (Lauraceae): Genome Structure and Phylogenetic Analysis. *Genes* **2022**, *13*, 2402. [CrossRef] [PubMed]
- 44. Shabalina, S.A.; Spiridonov, N.A.; Kashina, A. Sounds of silence: Synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Res.* 2013, *41*, 2073–2094. [CrossRef] [PubMed]
- Zhou, M.; Long, W.; Li, X. Patterns of synonymous codon usage bias in chloroplast genomes of seed plants. For. Stud. China 2008, 10, 235–242. [CrossRef]
- Xiao-Ming, Z.; Junrui, W.; Li, F.; Sha, L.; Hongbo, P.; Lan, Q.; Jing, L.; Yan, S.; Weihua, Q.; Lifang, Z.; et al. Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci. Rep.* 2017, *7*, 1555. [CrossRef] [PubMed]

- 47. Zhu, A.; Guo, W.; Gupta, S.; Fan, W.; Mower, J.P. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. *New Phytol.* **2016**, 209, 1747–1756. [CrossRef] [PubMed]
- 48. Song, Y.; Yao, X.; Tan, Y.; Gan, Y.; Corlett, R.T. Complete chloroplast genome sequence of the avocado: Gene organization, comparative analysis, and phylogenetic relationships with other Lauraceae. *Can. J. For. Res.* **2016**, *46*, 1293–1301. [CrossRef]
- Ma, J.; Yang, B.; Zhu, W.; Sun, L.; Tian, J.; Wang, X. The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. *Gene* 2013, 528, 120–131. [CrossRef]
- Zeb, U.; Wang, X.; AzizUllah, A.; Fiaz, S.; Khan, H.; Ullah, S.; Ali, H.; Shahzad, K. Comparative genome sequence and phylogenetic analysis of chloroplast for evolutionary relationship among *Pinus* species. *Saudi J. Biol. Sci.* 2022, 29, 1618–1627. [CrossRef]
- 51. Dodsworth, S.; Chase, M.W.; Kelly, L.J.; Leitch, I.J.; Macas, J.; Novak, P.; Piednoel, M.; Weiss-Schneeweiss, H.; Leitch, A.R. Genomic repeat abundances contain phylogenetic signal. *Syst. Biol.* **2015**, *64*, 112–126. [CrossRef]
- 52. dePamphilis, C.W.; Palmer, J.D. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* **1990**, *348*, 337–339. [CrossRef] [PubMed]
- McCoy, S.R.; Kuehl, J.V.; Boore, J.L.; Raubeson, L.A. The complete plastid genome sequence of *Welwitschia mirabilis*: An unusually compact plastome with accelerated divergence rates. *BMC Evol. Biol.* 2008, *8*, 130. [CrossRef] [PubMed]
- 54. Li, R.; Ma, P.F.; Wen, J.; Yi, T.S. Complete sequencing of five araliaceae chloroplast genomes and the phylogenetic implications. *PLoS ONE* **2013**, *8*, e78568. [CrossRef] [PubMed]
- 55. Liu, J.; Shi, M.; Zhang, Z.; Xie, H.; Kong, W.; Wang, Q.; Zhao, X.; Zhao, C.; Lin, Y.; Zhang, X.; et al. Phylogenomic analyses based on the plastid genome and concatenated nrDNA sequence data reveal cytonuclear discordance in genus *Atractylodes* (Asteraceae: Carduoideae). *Front. Plant Sci.* 2022, *13*, 1045423. [CrossRef] [PubMed]
- 56. Han, B.; Xue, Y. Genome-wide intraspecific DNA-sequence variations in rice. Curr. Opin. Plant Biol. 2003, 6, 134–138. [CrossRef]
- 57. Muraguri, S.; Xu, W.; Chapman, M.; Muchugi, A.; Oluwaniyi, A.; Oyebanji, O.; Liu, A. Intraspecific variation within Castor bean (*Ricinus communis* L.) based on chloroplast genomes. *Ind. Crops Prod.* **2020**, *155*, 112779. [CrossRef]
- 58. Zhang, R.S.; Yang, J.; Hu, H.L.; Xia, R.X.; Li, Y.P.; Su, J.F.; Li, Q.; Liu, Y.Q.; Qin, L. A high level of chloroplast genome sequence variability in the Sawtooth Oak *Quercus acutissima*. *Int. J. Biol. Macromol.* **2020**, *152*, 340–348. [CrossRef]
- 59. Dong, W.; Liu, J.; Yu, J.; Wang, L.; Zhou, S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* **2012**, *7*, e35071. [CrossRef]
- Cao, Z.; Yang, L.; Xin, Y.; Xu, W.; Li, Q.; Zhang, H.; Tu, Y.; Song, Y.; Xin, P. Comparative and phylogenetic analysis of complete chloroplast genomes from seven *Neocinnamomum* taxa (Lauraceae). *Front. Plant Sci.* 2023, 14, 1205051. [CrossRef]
- 61. Hinsinger, D.D.; Strijk, J.S. Toward phylogenomics of Lauraceae: The complete chloroplast genome sequence of *Litsea glutinosa* (Lauraceae), an invasive tree species on Indian and Pacific Ocean islands. *Plant Gene* **2017**, *9*, 71–79. [CrossRef]
- 62. Song, Y.; Yao, X.; Tan, Y.; Gan, Y.; Yang, J.; Corlett, R.T. Comparative analysis of complete chloroplast genome sequences of two subtropical trees, *Phoebe sheareri* and *Phoebe omeiensis* (Lauraceae). *Tree Genet. Genomes* **2017**, *13*, 120. [CrossRef]
- Li, X.; Yang, Y.; Henry, R.J.; Rossetto, M.; Wang, Y.; Chen, S. Plant DNA barcoding: From gene to genome. *Biol. Rev. Camb. Philos.* Soc. 2015, 90, 157–166. [CrossRef] [PubMed]
- Zhang, Z.; Zhang, Y.; Song, M.; Guan, Y.; Ma, X. Corrigendum: Species Identification of Dracaena Using the Complete Chloroplast Genome as a Super-Barcode. *Front. Pharmacol.* 2020, 11, 51. [CrossRef] [PubMed]
- Göltenboth, F.; Timotius, K.; Milan, P.; Margraf, J. Ecology of Insular Southeast Asia; Elsevier: Amsterdam, The Netherlands, 2006; pp. 385–399.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.