



Article

# Comparative Chloroplast Genomics, Phylogenomics, and Divergence Times of *Sassafras* (Lauraceae)

Zhiyuan Li <sup>1,†</sup> , Yunyan Zhang <sup>1,†</sup> , David Y. P. Tng <sup>2</sup> , Qixun Chen <sup>1</sup> , Yahong Wang <sup>1</sup> , Yongjing Tian <sup>1</sup>,  
Jingbo Zhou <sup>1</sup> and Zhongsheng Wang <sup>1,\*</sup>

<sup>1</sup> College of Life Sciences, Nanjing University, Nanjing 210023, China; 502023300067@smail.nju.edu.cn (Z.L.); zyyynju@nju.edu.cn (Y.Z.); 502024300060@smail.nju.edu.cn (Q.C.); 502024300072@smail.nju.edu.cn (Y.W.); mg1830075@smail.nju.edu.cn (Y.T.); mg1830084@smail.nju.edu.cn (J.Z.)

<sup>2</sup> Centre for Rainforest Studies, The School for Field Studies, Yungaburra, QLD 4884, Australia; dtng@fieldstudies.org

\* Correspondence: wangzs@nju.edu.cn

† These authors contributed equally to this work.

## Abstract

In the traditional classification system of the Lauraceae family based on morphology and anatomy, the phylogenetic position of the genus *Sassafras* has long been controversial. Chloroplast (cp) evolution of *Sassafras* has not yet been illuminated. In this study, we first sequenced and assembled the complete cp genomes of *Sassafras*, and conducted the comparative cp genomics, phylogenomics, and divergence time estimation of this ecological and economic important genus. The whole length of cp genomes of the 10 *Sassafras* ranged from 151,970 bp to 154,011 bp with typical quadripartite structure, conserved gene arrangements and contents. Variations in length of cp were observed in the inverted repeat regions (IRs) and a relatively high usage frequency of codons ending with T/A was detected. Four hypervariable intergenic regions (*ccsA-ndhD*, *trnH-psbA*, *rps15-ycf1*, and *petA-psbJ*) and 672 cp microsatellites were identified for *Sassafras*. Phylogenetic analysis based on 106 cp genomes from 30 genera within the Lauraceae family demonstrated that *Sassafras* constituted a monophyletic clade and grouped a sister branch with the *Cinnamomum* sect. *Camphora* within the tribe Cinnamomeae. Divergence time between *S. albidum* and its East Asian siblings was estimated at the Middle Miocene (16.98 Mya), *S. tzumu* diverged from *S. randaiense* at the Pleistocene epoch (3.63 Mya). Combined with fossil evidence, our results further revealed the crucial role of the Bering Land Bridge and glacial refugia in the speciation and differentiation of *Sassafras*. Overall, our study clarified the evolution pattern of *Sassafras* cp genomes and elucidated the phylogenetic position and divergence time framework of *Sassafras*.

**Keywords:** *Sassafras*; Lauraceae; comparative chloroplast genomics; divergence times; phylogeny of *Sassafras*



Academic Editor: Sergei A. Subbotin

Received: 9 July 2025

Revised: 25 July 2025

Accepted: 28 July 2025

Published: 30 July 2025

**Citation:** Li, Z.; Zhang, Y.; Tng, D.Y.P.; Chen, Q.; Wang, Y.; Tian, Y.; Zhou, J.; Wang, Z. Comparative Chloroplast Genomics, Phylogenomics, and Divergence Times of *Sassafras* (Lauraceae). *Int. J. Mol. Sci.* **2025**, *26*, 7357. <https://doi.org/10.3390/ijms26157357>

**Copyright:** © 2025 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 genus *Sassafras* J. Presl belongs to the Lauraceae family and includes three extant deciduous trees; *Sassafras albidum* (Laurales: Lauraceae) (Nutt.) Nees., *Sassafras tzumu* (Laurales: Lauraceae) (Hemsl.) Hemsl., *Sassafras randaiense* (Laurales: Lauraceae) (Hayata.) Rehder. This genus was firstly established by the Czech botanist Jan Svatopluk Presl [1] in 1825 who described *S. albidum*, a species endemic to eastern North America. For over eight decades, *S. albidum* was considered monotypic. This view changed in 1907, when

British botanist William Hemsley [2] named *S. tzumu* from mainland China. Later, in 1920, the American botanist Rehder [3] named *S. randaiense* in Taiwan, China. The three *Sassafras* species currently present typical East Asia–North America disjunct distribution with ecological and economic importance [4,5]. Regarding their fossil records, Berry [6] first discovered the extinct species *S. hesperia*, which was excavated from Late Miocene deposits in eastern Washington and northwestern Idaho. Poole et al. [7] found another fossil species, *S. oxylon gottwaldii*, with a potential affinity to *Sassafras* from the wood fossils in the Late Cretaceous sediments in the northern region of the Antarctic Peninsula, indicating that *Sassafras* may have Gondwanan origins.

Due to the disputable morphological characteristics of *Sassafras*, such as the flower sexuality, anther orientation (inward or outward), the inflorescence type, and the presence or absence of involucre, the phylogenetic position of *Sassafras* had been controversial in the traditional classification system of Lauraceae. Specifically, in the classification system of Kostermans [8], *Sassafras* is considered to belong to the raceme-false umbel group, which was closely related to the genera *Lindera* and *Litsea* which have involucres and false umbels. In contrast, Van der Werff and Richter [9] believed that *Sassafras* belonged to the raceme-thyrsoid cyme group, which is more closely related to genera such as *Cinnamomum* and *Ocotea* without involucres and with thyrsoid cymes.

With the wide application of molecular phylogenetics, phylogenetic trees constructed by Rohwer, Chanderbali et al., and Rohwer and Rudolph [10–12] using different gene fragments found the terminal branches including the *Persea* group and the Laureae–Cinnamomeae clade. Almost all genera with controversial systematic relationships in the Lauraceae family were clustered in this branch, including *Sassafras*. Therefore, it was named the core Lauraceae group. In further phylogenetic studies on the core Lauraceae group, the research results of Li et al. and Nie et al. [13,14] found stronger support for including *Sassafras* into the Cinnamomeae tribe, which is sister to the *Laurus* tribe. The Cinnamomeae tribe branch and the *Laurus* tribe branch formed a sister pair. In the phylogenetic tree constructed by Rohde et al. [15] based on nuclear ITS sequences and chloroplast intergenic regions *psbA-trnH* and *trnG-trnS* sequences, *Cinnamomum* was found to be paraphyletic group and was divided into three lineages. Based on ITS data, *Sassafras* appeared as the sister group of the *Cinnamomum* sect. *Cinnamomum* group, while in the *psbA-trnH* and *trnG-trnS* phylogenetic trees, it appeared as the sister group of the *Cinnamomum* sect. *Camphora* group. Since *Cinnamomum* is the core genus of the Cinnamomeae tribe, and *Sassafras* showed a sister relationship with different taxonomic groups of *Cinnamomum* based on different sequences, this provided molecular evidence for classifying *Sassafras* into the Cinnamomeae tribe.

Within the last decade, molecular phylogeneticists have started constructing phylogenetic trees using complete cp genome data, as this is more informative than single-gene or multi-gene fragment data. For instance, Song et al. [16] constructed a phylogenetic tree using the published complete chloroplast genome data of 34 species of Lauraceae, which demonstrated unequivocal support for the separation of the *Laurus* and the Cinnamomeae tribes, and also the classification of *Sassafras* into the Cinnamomeae tribe. Zhao et al. [17] used two *Endiandra* species as the outgroup and conducted a systematic analysis of the complete chloroplast genome data of 30 species in the *Persea*–*Laurus* tribe branch of the Lauraceae family, forming three major branches: the *Persea*–*Machilus* branch, the *Ocotea*–*Cinnamomum* branch, and the *Laurus* branch. *Sassafras* was located in the *Ocotea*–*Cinnamomum* branch, which is sister to the *Laurus* tribe branch. Jo et al. [18] conducted a phylogenetic analysis of 49 Lauraceae species using 77 protein-coding sequences and four rRNA gene sequences, and obtained six distinct branches: Cryptocaryeae, Neocinnamomeae, Caryodaphnopsidae, Perseeae, Cinnamomeae, *Laurus*. Among them, *Sassafras* was

classified into the Cinnamomeae tribe. Song et al. and Liu et al. [19,20] both divided the Lauraceae family into 9 branches (Hypodaphnideae, Cryptocaryeae, Caryodaphnopsidae, Neocinnamomeae, Cassytheae, Mezilaurus, Perseeae, Cinnamomeae, Laurus) based on plastid genomes, and *Sassafras* was classified into the Cinnamomeae tribe. Yang et al. [21] newly sequenced the plastid genomes of five species based on phytospecimenomics, and the results of phylogenetic analysis also supported the division of the Lauraceae family into nine branches, with *Sassafras* classified into the Cinnamomeae tribe. Song et al. [22] conducted a phylogenetic analysis of 91 species from 29 genera in the Lauraceae family based on mitochondrial genomes, and the results still supported that *Sassafras* and *Cinnamomum* as sister taxa.

Collectively, the currently constructed phylogenetic trees of the Lauraceae family basically support the establishment of the Cinnamomeae tribe, the classification of *Sassafras* into the Cinnamomeae tribe, and the sister-group relationship between the Cinnamomeae tribe and the Laurus tribe [23]. However, in most previous studies, there were problems such as a small number of *Sassafras* samples selected and the failure to include all three species of *Sassafras*. This also led to fewer studies on the phylogenetic relationships among the three species within *Sassafras*, and there was still some uncertainty as to which group, *C. sect. Cinnamomum* or *C. sect. Camphora*, *Sassafras* was more closely related to.

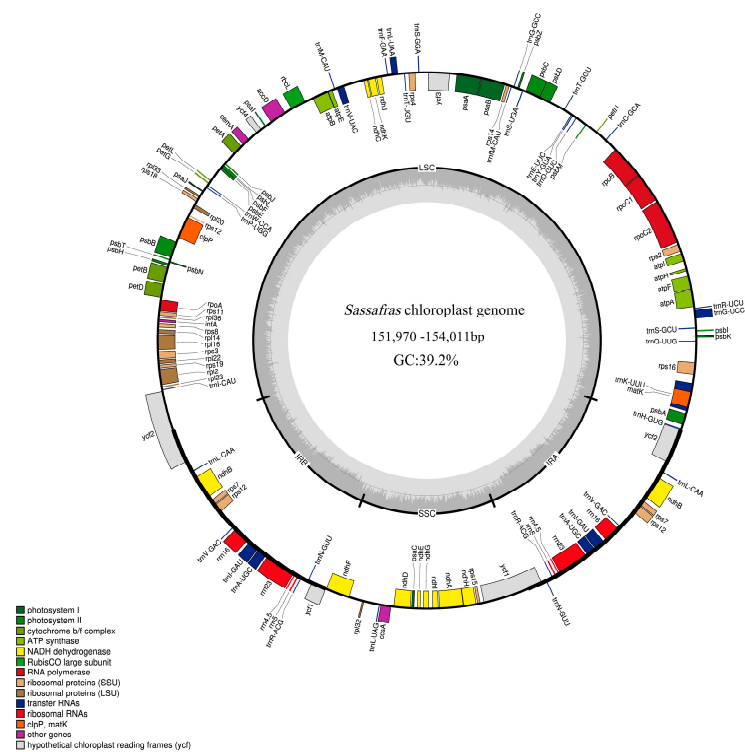
In addition, estimation of species divergence times is a research hotspot in phylogenetics and biogeography, which not only help us to estimate important parameters such as species differentiation and evolution rates, but also contribute to exploring the influence of geological history, paleoecology and other factors in the process of species evolution [24]. However, research on the divergence time of *Sassafras* is relatively lacking. So far, merely Nie et al. [14] estimated the divergence time of *Sassafras* using ITS and multiple cpDNA markers (*psbA-trnH*, *rpl16*, and *trnL-F*); the results showed that *S. albidum* diverged from its two East Asian siblings at  $13.80 \pm 2.29$ – $16.69 \pm 2.52$  Mya (mid-Miocene), and the divergence time between *S. tzumu* in mainland China and *S. randaiense* was approximately  $0.61 \pm 0.75$ – $2.23 \pm 0.76$  Mya (Pleistocene).

In this study, we first sequenced and assembled the complete chloroplast genomes of ten individuals of *Sassafras* (five individuals of *S. albidum*, three individuals of *S. tzumu*, and two individuals of *S. randaiense*). We aimed to: (1) investigate the mechanism of chloroplast genome evolution of three *Sassafras* species and develop genetic markers such as microsatellites and DNA barcodes for the genus; (2) explore the phylogenetic position of *Sassafras* in the Lauraceae family, the relationship between *Sassafras* and *Cinnamomum*, and the internal phylogenetic relationships of *Sassafras*; and (3) estimate the divergence times of *Sassafras* and infer their biogeography history combined with fossil evidence.

## 2. Results

### 2.1. Characteristics of *Sassafras* Chloroplast Genomes

The chloroplast (cp) genomes of ten *Sassafras* individuals all displayed a quadripartite structure, consisting of a pair of inverted repeat regions (IRa and IRb), large and small single-copy regions (LSC and SSC) (Figure 1), and ranging from 151,970 bp (*S. randaiense*: TW10) to 154,011 bp (*S. albidum*: B22) (Table 1). The size of the LSC regions ranged from 92,740 bp (*S. tzumu*: BYS) to 93,634 bp (*S. albidum*: B22), the SSC lengths ranged from 18,756 bp (*S. randaiense*: TW10) to 18,885 bp (*S. albidum*: B18), and the IR regions ranged from 20,054 (*S. albidum*: B1) to 20,809 bp (*S. tzumu*: TW3) (Table 1).



**Figure 1.** Circular map of chloroplast genomes of *Sassafra* with annotated genes. Genes shown inside and outside of the circle are transcribed in clockwise and counter-clockwise directions respectively. Genes belonging to different functional groups are color-coded. The GC and AT content are denoted by the dark grey and light grey color in the inner circle, respectively. LSC, SSC, and IR are large single-copy region, small single-copy region, and inverted repeat region, respectively.

**Table 1.** Comparison of complete plastid genomes of 10 *Sassafra* individuals.

| Species              | Collection Number | Genbank ID | Whole Length (bp) | Length of LSC (bp) | Length of IR (bp) | Length of SSC (bp) | Total GC Content (%) | Number of CDS | Number of tRNA | Number of rRNA |
|----------------------|-------------------|------------|-------------------|--------------------|-------------------|--------------------|----------------------|---------------|----------------|----------------|
| <i>S. albidum</i>    | B1                | MW683126   | 152,562           | 93,587             | 20,054            | 18,867             | 39.2                 | 82            | 36             | 8              |
| <i>S. albidum</i>    | B12               | MW696794   | 152,621           | 93,600             | 20,077            | 18,867             | 39.2                 | 82            | 36             | 8              |
| <i>S. albidum</i>    | B17               | MW696797   | 153,910           | 93,562             | 20,732            | 18,884             | 39.2                 | 82            | 36             | 8              |
| <i>S. albidum</i>    | B18               | MW696798   | 152,567           | 93,564             | 20,059            | 18,885             | 39.2                 | 82            | 36             | 8              |
| <i>S. albidum</i>    | B22               | MW696799   | 154,001           | 93,634             | 20,750            | 18,867             | 39.2                 | 82            | 36             | 8              |
| <i>S. tzumu</i>      | BYS               | MW696800   | 153,137           | 92,740             | 20,792            | 18,813             | 39.2                 | 82            | 36             | 8              |
| <i>S. tzumu</i>      | HNS               | MW696801   | 151,797           | 92,751             | 20,096            | 18,854             | 39.2                 | 82            | 36             | 8              |
| <i>S. tzumu</i>      | LF                | MW696802   | 153,154           | 92,752             | 20,774            | 18,854             | 39.2                 | 82            | 36             | 8              |
| <i>S. randaiense</i> | TW3               | MW696808   | 153,146           | 92,772             | 20,809            | 18,756             | 39.2                 | 82            | 36             | 8              |
| <i>S. randaiense</i> | TW10              | MW696807   | 151,790           | 92,772             | 20,131            | 18,756             | 39.2                 | 82            | 36             | 8              |

Note: LSC, SSC, and IR are large single-copy region, small single-copy region, and inverted repeat region, respectively.

The overall GC content of the ten *Sassafra* cp genomes was 39.2% (Table 1). A total of 128 genes were contained in the ten *Sassafra* cp genomes, including 82 protein-coding genes (CDS), 36 transfer RNA (tRNA) genes, 8 ribosomal RNA (rRNA) genes, and two pseudogenes, respectively (Table 2). Among these genes, 20 genes contained a single intron, while two protein-coding genes possessed two introns (Table 2). The gene *rps12* was trans-spliced; the exon at the 5' end located in the LSC region, whereas the 3' exon and intron located in the IR regions. Moreover, the  $\psi ycf1$  and  $\psi ycf2$  were identified as pseudogenes because of the partial duplication (Table 2). The cp genomes of ten *Sassafra* individuals were deposited in the GenBank (NCBI). The accession numbers are recorded in Table 1.



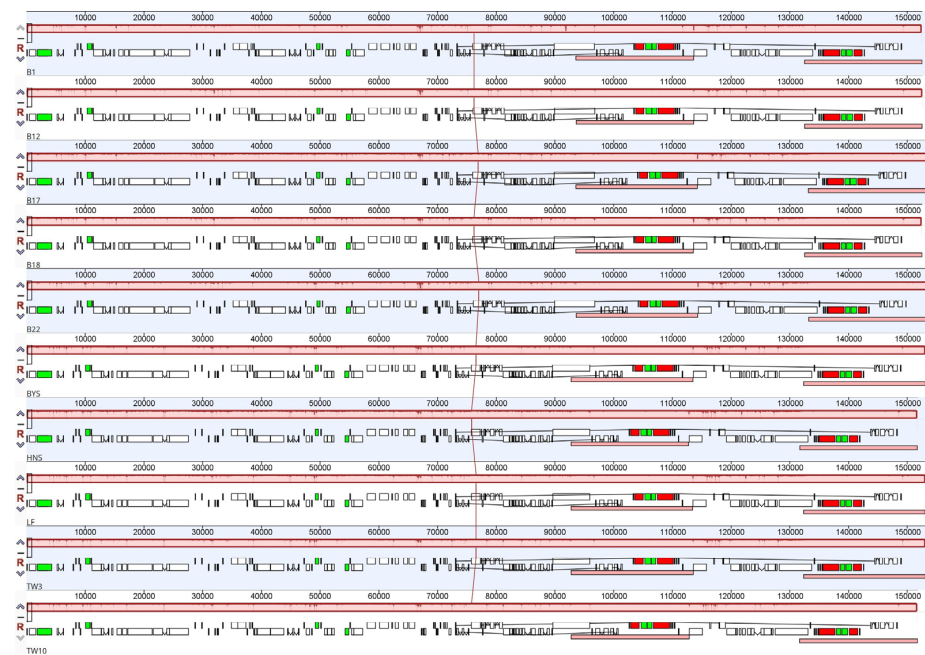
**Table 2.** List of genes in chloroplast genome of *Sassafras*.

| Groups of Genes                                   | Names of Genes                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|---------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ribosomal RNAs                                    | <i>rrn4.5</i> (×2), <i>rrn5</i> (×2), <i>rrn16</i> (×2), <i>rrn23</i> (×2)<br>* <i>trnA</i> -UGC (×2), <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC,<br><i>trnF</i> -GAA, <i>trnG</i> -GCC, * <i>trnG</i> -UCC, <i>trnH</i> -GUG, <i>trnI</i> -CAU,<br>* <i>trnI</i> -GAU (×2), * <i>trnK</i> -UUU, <i>trnL</i> -CAA (×2), *<br><i>trnL</i> -UAA, <i>trnL</i> -UAG, <i>trnM</i> -CAU, <i>trnM</i> -CAU,<br><i>trnN</i> -GUU (×2), <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG (×2),<br><i>trnR</i> -UCU, <i>trnS</i> -GCU, <i>trnS</i> -GGA,<br><i>trnS</i> -UGA, <i>trnT</i> -GGU, <i>trnT</i> -UGU, <i>trnV</i> -GAC (×2), *<br><i>trnV</i> -UAC, <i>trnW</i> -CCA, <i>trnY</i> -GUA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Transfer RNAs                                     | <i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i><br><i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> ,<br><i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i><br><i>petA</i> , * <i>petB</i> , * <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i><br><i>atpA</i> , <i>atpB</i> , <i>atpE</i> , * <i>atpF</i> , <i>atpH</i> , <i>atpI</i><br><i>rbcL</i><br>* <i>ndhA</i> , * <i>ndhB</i> (×2), <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> ,<br><i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i><br>** <i>clpP</i><br><i>cemA</i><br>* <i>rpl2</i> , <i>rpl14</i> , * <i>rpl16</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i><br><i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i> , <i>rps11</i> , * <i>rps12</i> (×2), <i>rps14</i> ,<br><i>rps15</i> , * <i>rps16</i> , <i>rps18</i> , <i>rps19</i><br><i>rpoA</i> , <i>rpoB</i> , * <i>rpoC1</i> , <i>rpoC2</i><br><i>infA</i><br><i>matK</i> , <i>accD</i> , <i>ccsA</i><br>** <i>ycf3</i> , <i>ycf4</i> , <i>ycf1</i> , <i>ycf2</i><br>Ψ <i>ycf1</i> , Ψ <i>ycf2</i> |
| Photosystem I                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Photosystem II                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Cytochrome                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| ATP synthase                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Rubisco                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| NADH dehydrogenase                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| ATP-dependent protease subunit P                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Chloroplast envelop membrane protein              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Large units                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Small units                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| RNA polymerase                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Translational initiation factor                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Miscellaneous proteins                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Hypothetical proteins and conserved reading frame |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Pseudogene                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |

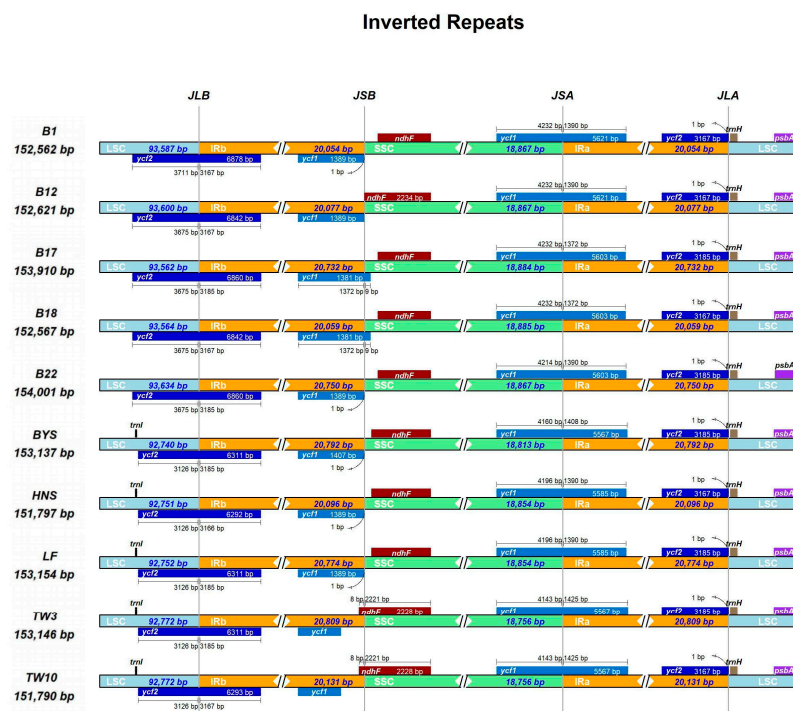
Note: Asterisks (\*) before gene names indicate one intron containing genes, and double asterisks (\*\*) indicate two introns in the gene. (×2) indicates genes duplicated in inverted repeat regions (IR). Pseudogene is represented by (Ψ).

The cp genomes comparisons of the ten *Sassafras* in mVISTA revealed a high sequence similarity, and the IR regions and coding regions were more conserved than the LSC region, SSC region, and non-coding regions (Figure S1). In addition, the MAUVE alignment with the algorithm of progressive MAUVE based on the ten *Sassafras* cp genomes showed only one locally collinear block between all analyzed cp genomes and all of the genes exhibited the same and consistent sequence order and no gene re-arrangements or inversion events were detected in these genomes (Figure 2).

The connection sites of various regions of the 10 cp genomes of *Sassafras* were visualized using the IR-Scope online tool (Figure 3). The results showed that there were certain differences in the lengths of the LSC and SSC regions of the chloroplast genomes of 10 individuals from the three *Sassafras* species. The length of the LSC region was 93,534–93,600 bp in *S. albidum*, 92,740–92,753 bp in *S. tzumu*, and 92,772 bp in *S. randaiense*. The length of the SSC region was 18,867–18,885 bp in *S. albidum*, 18,854–18,813 bp in *S. tzumu*, and 18,756 bp in *S. randaiense*. Correspondingly, the length of the IR region was 20,054–20,750 bp in *S. albidum*, 20,096–20,774 bp in *S. tzumu*, and 20,131–20,809 bp in *S. randaiense*.



**Figure 2.** Alignment of ten *Sassafras* chloroplast genomes. Genome of *Sassafras albidum* (B1) is shown at the top as the reference genome. Within each of the alignments, local collinear blocks are represented by blocks of the same color connected by lines.



**Figure 3.** Comparison of the borders of the IR, SSC and LSC regions among ten chloroplast genomes of *Sassafras*. JLB, JSB, JSA, and JLA represent the junctions of LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC, respectively.

The boundaries of LSC/IRb (JLB), SSC/IRa (JSA), and IRa/LSC (JLA) were highly conserved, with only minor changes. The boundary of IRb/SSC (JSB) changed significantly. The *ndhF* gene was only 8 bp located in the IRb region in *S. randiaense*, and was only located in the SSC region in *S. albidum* and *S. tzumu*. In addition, the *psbA* and *trnH* genes were located in the LSC region, and the distance from the *trnH* gene to JLA was 1 bp in all cases. The cp genomes of *Sassafras* all had complete *ycf1* and pseudo *ycf1*<sup>(ψycf1)</sup> genes. The length

of the  $\Psi_{ycf1}$  gene located in the IRb region was 1382 bp (B12) to 1408 bp (BYS), and only the  $\Psi_{ycf1}$  in *S. randaiense* was relatively far from the JSB boundary.

The comparison results of codon preference indexed by the value of the relative synonymous codon usage (RSCU) showed that the sequences of *Sassafras* plants had 64 shared codons (Figure S2). Among them, twenty-one codons had RSCU > 1.3 (nine ended with A and twelve ended with T), six codons had RSCU between 1.2 and 1.3 (four ended with T, one ended with A, and one ended with G), and four codons had RSCU between 1.0 and 1.2 (three ended with A and one ended with C). The RSCU values of two codons were equal to 1, which were ATG and TCG. The RSCU values of 31 codons were less than 1, and 28 of them ended with G or C (Table S3). These results indicated that codons ending with T/A had a relatively high usage frequency in the chloroplast genomes of *Sassafras* plants.

## 2.2. Enrichment of Chloroplast DNA Genetic Resources of *Sassafras*

A total of 500 dispersed repeat sequences in the cp genomes of *Sassafras* (Figure S3), including 169 forward repeat sequences (f) (33.8%), 108 reverse repeat sequences (r) (21.6%), 21 complementary repeat sequences (c) (4.2%), and 202 palindromic repeat sequences (p) (40.4%). A total of 672 chloroplast SSRs were detected in the chloroplast genomes of *Sassafras* by MISA (Table S4). The number of SSRs in a single cp genome ranged from 66 (*S. randaiense*: TW3) to 69 (*S. albidum*: B17). Among them, the nucleotides of mononucleotide SSRs (P1, 60.27%) were mainly composed of short A/T repeat sequences, and only a few G (2)/C (4) tandem repeat sequences were present in *S. albidum*. The dinucleotide (P2, 9.52%) types were AT, TA, GA, and TC, and only *S. albidum* had no TC. The trinucleotide (P3, 1.34%) types were TAT, AAT, and AAT only appeared 1 time in B1, and *S. tzumu* had no trinucleotides. The tetranucleotide (P4, 5.65%) types were TTTA, AAAT, AATG, and TTTC, and *S. randaiense* lacked one TTTC. The pentanucleotide (P5) only appeared 1 time in B12, with the type AATAA. No hexanucleotides were found. There were 155 complex nucleotide repeat sequences Pc, accounting for 23.07% of the total number of repeats. The chloroplast SSRs of *Sassafras* were mainly located in the LSC region (79.3%), and the proportions in the SSC (14.7%) and IR (5.95%) regions were relatively small.

Among the 10 individuals of *Sassafras*, the total average Pi value of 106 gene fragments was 0.002763. The average value of the gene-coding region was 0.00159, the average value of the gene intron region was 0.00222, and the average value of the intergenic region was 0.004678. Only four fragments were in the intergenic region, *trnH-psbA* and *petA-psbJ* were located in the LSC region, and *ccsA-ndhD* and *rps15-ycf1* were located in the SSC region. Among them, the *petA-psbJ* fragment had the highest Pi value among all fragments (Pi = 0.03081) (Figure 4 and Table S5).

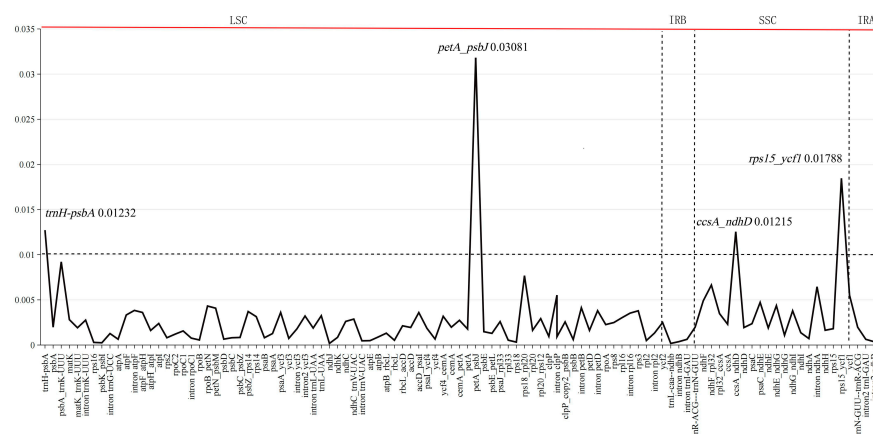


Figure 4. Nucleotide variability (Pi) values of ten individuals of *Sassafras* chloroplast genomes, respectively.

### 2.3. Phylogenetic Relationships

We used 106 cp genomes of 99 species from 30 genera in the Lauraceae family, as well as the cp genomes of *Calycanthus* (*C. floridus* and *C. chinensis*) as the outgroups. Phylogenetic trees were constructed based on two datasets (the whole cp genomes and CDS sequences) using the maximum likelihood method (ML) and the Bayesian method (BI). Results showed that the topological structures of the phylogenetic trees constructed with the complete cp genome sequences (Figure S4) and the CDS datasets (Figure S5) were basically the same. The Lauraceae family can be divided into nine branches: the Laurus-Neolitsea branch, the Cinnamomum-Ocotea branch, the Machilus-Persea branch, the Mezilaurus branch, the Caryodaphnopsis branch, the Neocinnamomum branch, the Cassytha branch, the Beilschmiedi-Cryptocarya branch, and the Hypodaphnideae branch.

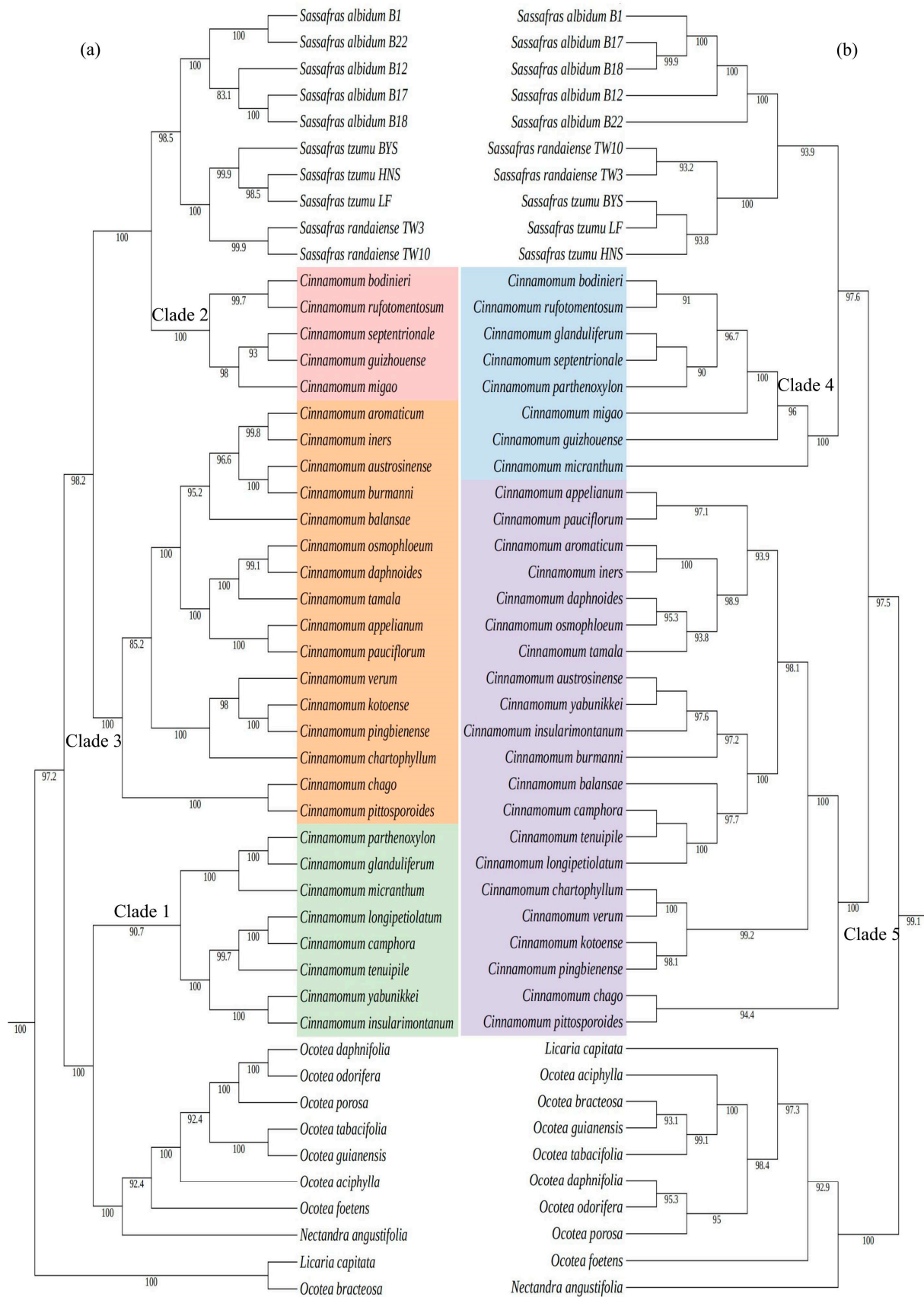
The minor differences between the complete cp genome-based tree and the CDS-based tree mainly focused on the topological structure of the Cinnamomeae tribe (including *Sassafras*). Specifically, the Cinnamomeae tribe in our phylogenetic trees consist of the genera *Sassafras*, *Cinnamomum*, *Ocotea*, *Nectandra angustifolia*, and *Licaria capitata*. Among them, *Sassafras* formed a monophyletic group, and the three species within *Sassafras* were also monophyletic taxa. *S. albidum* was at the base of *Sassafras* (BS = 100, PP = 1), *S. tzumu* and *S. randaiense* were sister species (BS = 100, PP = 1, Figure 5). The differences were as follows: in the phylogenetic tree constructed with the complete cp genome sequences (Figure 5a), *Licaria capitata* and *Ocotea bracteosa* formed the base of the Cinnamomeae tribe (BS = 100, PP = 1). The remaining seven *Ocotea* species formed a monophyletic group and were sister species to *Nectandra angustifolia* (BS = 100, PP = 1). *Cinnamomum* was a polyphyletic group and could be divided into three branches: Clade 1, Clade 2, and Clade 3 (Figure 5a). Clade 1 was a sister group to the *Ocotea* and *Nectandra angustifolia* complex (BS = 100, PP = 1). Clade 2 was a sister group to the *Sassafras* branch (BS = 100, PP = 1). Clade 3 was a sister group to the branch formed by Clade2 and *Sassafras* (BS = 100, PP = 1) (Figure 5a).

In contrast, in the cp CDS-based phylogenetic tree (Figure 5b), *Ocotea* complex (*Ocotea*, *Nectandra*, and *Licaria*) formed the base of the Cinnamomeae tribe and was sister species to the branch formed by *Cinnamomum* and *Sassafras* (BS = 100, PP = 1). *Nectandra* was at the base of the *Ocotea* complex. *Cinnamomum* was a paraphyletic group and could be divided into two branches: Clade 4 and Clade 5 (Figure 5b). Clade 4 was sister to the *Sassafras* (BS = 100, PP = 1). Clade 5 was a sister group to the branch formed by Clade 4 and *Sassafras* (BS = 100, PP = 1).

### 2.4. Divergence Time of *Sassafras*

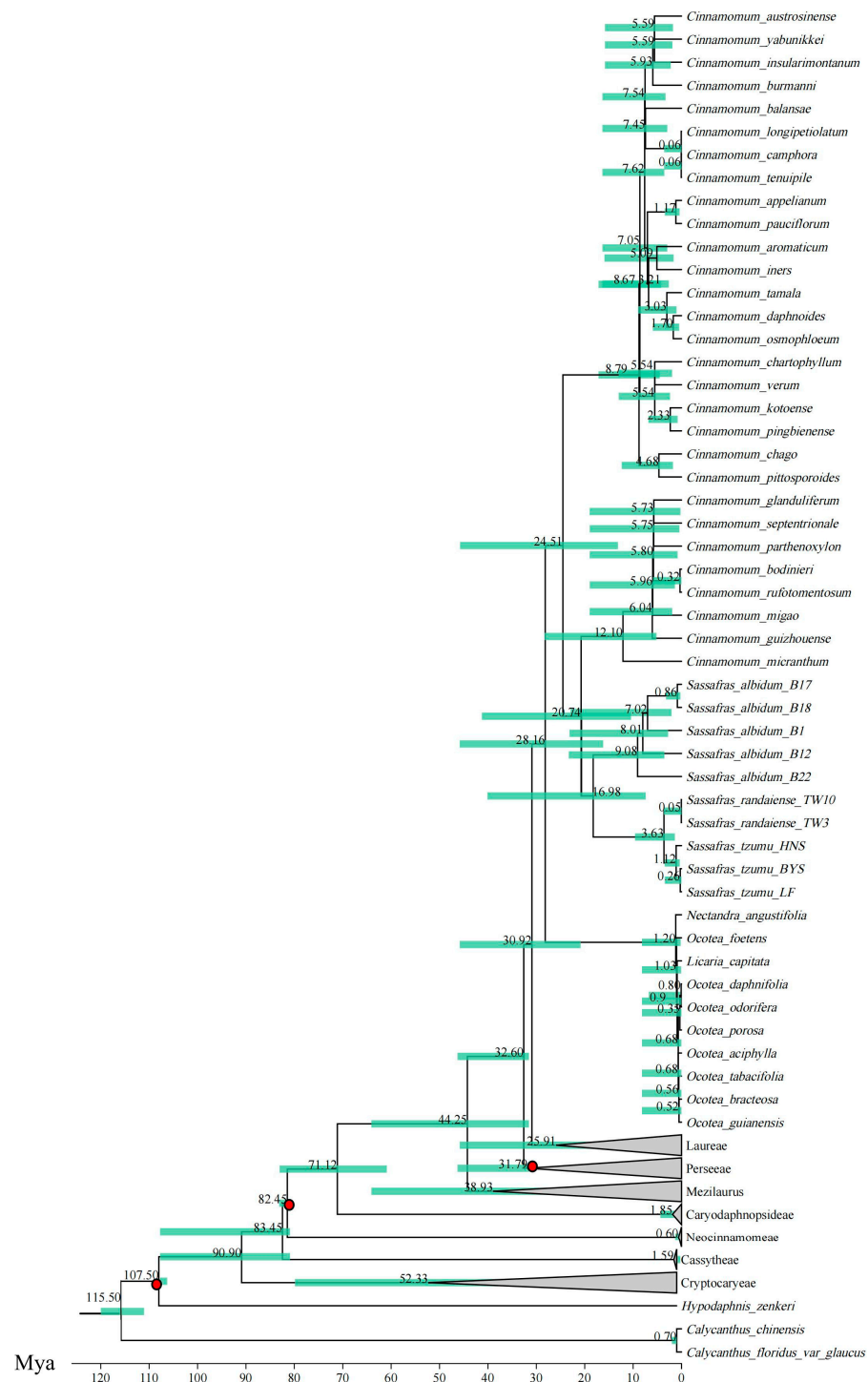
The results of divergence time estimation (Figure 6 and Figure S6) showed that the Cinnamomeae tribe originated at Early Oligocene (30.92 Mya; 95% HPD = 20.85–45.84 Mya) and diverged from the Laurus tribe at Middle Oligocene (28.16 Mya; 95% HPD = 16.19–45.84 Mya). The *C. sect. Cinnamomum* and the *C. sect. Camphora* + *Sassafras* complex separated at Late Oligocene (24.51 Mya; 95% HPD = 13.13–45.77 Mya). The divergence between *Sassafras* and the *C. sect. Camphora* group occurred at Early Miocene (20.74 Mya; 95% HPD = 10.43–41.27 Mya). *S. albidum* diverged from the two East Asian siblings at Middle Miocene (16.98 Mya; 95% HPD = 8.54–41.27 Mya), and the divergence time between *S. tzumu* and *S. randaiense* was at the Pleistocene epoch (3.63 Mya; 95% HPD = 1.37–9.61 Mya).





**Figure 5.** Part of the Cinnamomeae tribe in maximum-likelihood and Bayesian analysis trees constructed based on 106 chloroplast genomes of 99 species from 30 genera in the Lauraceae family. Left (a) is the phylogenetic tree constructed with the complete chloroplast genomes, and right (b) is the phylogenetic tree constructed with the CDS genes. Numbers above the lines represent ML bootstrap.





**Figure 6.** Divergence time estimation using plastid protein-coding genes (PCGs). Green node bars indicate 95% highest posterior distributions; three red circles indicate fossil calibration points.

### 3. Discussion

#### 3.1. Chloroplast Evolution and Development of Genetic Resources for *Sassafras*

The comparative cp genomes of three *Sassafras* siblings revealed that their genome structures, overall gene arrangement, gene and GC content, and codon usage bias were highly conserved. The expansion/contraction of IR region (such as the size variation of the *ycf2* gene) is the primary cause of cp genome length variation in *Sassafras* (151,970–154,011 bp). This phenomenon is common in angiosperms. For example, Xiao

and Ge [25] found that the larger genome of *Cinnamomum chartophyllum* was caused by the expansion of the IR region.

In addition, abundant genetic resources were detected and developed from the ten cp genomes of *Sassafras* (Tables S4 and S5; Figures 1–4). Specifically, a total of 500 dispersed repeat sequences and 672 cpSSRs were enriched for *Sassafras*. These repeat sequences and cpSSR molecular markers developed in our study will be useful for population genetics and evolutionary studies of the genus *Sassafras* as well as the molecular marker-assisted selection, breeding and conservation of this and related genera in Lauraceae [26]. Chloroplast DNA molecular markers (divergence hotspot regions or DNA barcodes) have been extensively used for research on plant population genetics, phylogeny and phylogeography [26–28]. In this study, four hypervariable region fragments, *ccsA-ndhD*, *trnH-psbA*, *rps15-ycf1*, and *petA-psbJ* ( $Pi > 0.01$ ), were found in *Sassafras* as DNA barcodes.

### 3.2. Phylogenetic Insights

Both the complete cp genome-based and the CDS-based phylogenetic trees in our study strongly supported the following topological structures (Figures 5, S2 and S3): (1) nine major branches of Lauraceae: the Hypodaphnideae, the Cryptocaryeae, the Caryodaphnopsidae, the Neocinnamomeae, the Cassytheae, the Mezilaurus, the Perseeae, the Cinnamomeae, and the Laureae, which is consistent with the results of Song et al. [19]; (2) the classification of *Sassafras*, *Cinnamomum*, and the *Ocotea–Nectandra–Licaria* complex into the Cinnamomeae tribe, and the sister relationship between the tribes Cinnamomeae and Laureae; (3) *Sassafras* was a monophyletic group, and the three species within *Sassafras* were also monophyletic groups; *S. albidum* was at the base of *Sassafras*, and *S. tzumu* and *S. randaiense* were sister species ( $BS = 100$ ,  $PP = 1$ ), which was consistent with the results of Nie et al. [14] who constructed a phylogenetic tree of *Sassafras* using ITS data and multiple markers (*psbA-trnH*, *rpl16*, and *trnL-trnF*). However, some differences existed in the internal topological structure of the Cinnamomeae tribe between the complete cp genome-based and the CDS-based phylogenetic trees (Figure 5, Figures S2 and S3). The contradiction may result from the differences in evolutionary signals between the two types of datasets: the complete cp genome contains hypervariable non-coding regions, and their evolutionary rates are significantly higher than those of the CDS regions.

With respect to the phylogenetic position of *Sassafras*, in the complete cp genome-based phylogenetic tree, *Sassafras* was a sister group to Clade 2, while on the CDS-based tree, *Sassafras* was a sister group to Clade 4. According to Yang's research [29] on the classification of *Cinnamomum*, the *Cinnamomum* species in Clade 4 and Clade 2 belonged to the *C. sect. Camphora* group. Therefore, *Sassafras* and the *C. sect. Camphora* were sister groups, forming a monophyletic group ( $BS = 100$ ,  $PP = 1$ ). This is consistent with the result of Rohde et al. [15] that *Sassafras* appeared as the sister group of the *C. sect. Camphora* group in the phylogenetic tree established based on *psbA-trnH* and *trnG-trnS* sequences. *Sassafras* shares significant morphological similarities with *C. sect. Camphora*, supporting their close affinity, as evidenced by the following shared traits: both have prominent perulate terminal buds with well-developed, tightly wrapped scales, in contrast to the inconspicuous non-perulate buds of *C. sect. Cinnamomum*; their leaves are mostly alternate, clustered at branch apices, with pinnate or weakly tri-plinerved venation, differing from the usually opposite/subopposite leaves with typical tri-plinerved venation in *C. sect. Cinnamomum*; both are rich in oil cells in wood and leaves, accumulating volatile terpenoids; and their fruits are fleshy drupes borne on shallow cup-shaped receptacles with apically thickened pedicels, showing a higher degree of morphological matching in fruit and pedicel characteristics compared to other groups [29].

However, it is still inconsistent with the sister-group relationship between *Sassafras* and *C. sect. Cinnamomum* supported by the phylogenetic tree established based on ITS data. This phenomenon of nuclear-cytoplasmic inconsistency may be caused by ancient hybridization, gene introgression, or incomplete lineage sorting, which are common in plants [25].

### 3.3. Divergence Time and Biogeographic History of *Sassafras*

The estimation of divergence time based on fossil calibration showed that the divergence between *Sassafras* and the *C. sect. Camphora* group occurred at 20.74 Mya (95% HPD = 10.43–41.27 Mya), *S. albidum* diverged from its two East Asian species at 16.98 Mya (95% HPD = 8.54–41.27 Mya), and the divergence time between *S. tzumu* and *S. randaiense* was 3.63 Mya (95% HPD = 1.37–9.61 Mya). This is similar to the conclusion of Nie et al. [14], the divergence time between *S. albidum* and the two East Asian species was approximately  $13.80 \pm 2.29$ – $16.69 \pm 2.52$  Mya, and the divergence time between *S. tzumu* and *S. randaiense* was approximately  $0.61 \pm 0.75$ – $2.23 \pm 0.76$  Mya.

#### 3.3.1. Divergence Between *S. albidum* and Its East Asian Species

The divergence time between *S. albidum* and the East Asian clade was estimated at 16.98 Mya (Mid-Miocene), a timing that aligns with both the genus' long geological evolutionary history and critical transitions in the Northern Hemisphere temperate flora. Fossilized leaves of *S. hesperia*, a Miocene survivor from western North America, confirmed the continuous presence of North American populations in humid forest habitats until the Miocene [6]. This fossil record spatially and temporally corroborated the intercontinental divergence time inferred from molecular clock dating (16.98 Mya), indicating that drastic environmental changes during the Miocene were key drivers of the modern disjunct distribution pattern.

This divergence time (Mid-Miocene) shares significant commonalities with typical East Asia–eastern North America disjunct species, such as *Liriodendron* (14.15 Mya) and *Phryma leptostachys* (3.68–5.23 Mya) all concentrated within the Miocene (23–5 Mya) [30,31]. This pattern reflects the synergistic effects of the “periodic Bering Land Bridge openings and Climatic Optimum” during this interval. From the perspective of speciation theory, the divergence of *Sassafras* integrated mechanisms of allopatric speciation and niche differentiation: during the early Miocene (23–16 Mya), the decline in CO<sub>2</sub> concentration from 1500 to 700 ppm [32] triggered the contraction of pantropical flora, forcing temperate-adapted *Sassafras* to diverge ecologically from tropical relatives (e.g., *C. sect. Camphora*) by 20.74 Mya.

#### 3.3.2. Divergence Between *S. tzumu* and *S. randaiense*

According to the theory of allopatric speciation, the divergence between *S. randaiense* and *S. tzumu* at the Pleistocene epoch (3.63 Mya) fundamentally represents a product of geographic isolation triggered by land bridge closure. While the main body of Taiwan Island formed 4–5 Mya, the critical driver of population genetic divergence was the Pliocene–Pleistocene transition (5.3–2.6 Mya), during which the East China Sea shelf emerged as a migration corridor during glacial sea level drops, allowing some *Sassafras* populations to colonize Taiwan. Interglacial sea level rises (e.g., periodic flooding of the land bridge after 3.63 Mya) severed gene flow between mainland and island populations. This isolation was amplified by Quaternary glacial-interglacial cycles (2.58–0.01 Mya), eventually leading to the formation of a Taiwanese endemic species through independent evolution, while mainland populations could not re-colonize Taiwan due to long-term geographic isolation [33–35]. Notably, compared to the divergence times of *Picea wilsonii* (mainland spruce) and *P. morrisonicola* (Taiwan spruce, 1–2 Mya), and *Acer oliverianum* (mainland five-lobed

maple) vs. *A. oliverianum* subsp. *formosanum* (Taiwan five-lobed maple, 2.91 Mya) [36,37], the divergence of *S. tzumu* and *S. randaiense* occurred during an earlier period of land bridge instability, implying more complete isolation and lower potential for gene flow recovery.

Taiwan's unique topographic and climatic gradients (e.g., high-elevation habitats in the Central Mountain Range, tropical-subtropical monsoon climate) also imposed niche filtering. Modern *S. randaiense* is restricted to humid forests at 1000–2500 m elevation in central Taiwan, with morphological adaptations to high humidity and foggy environments [38]. In contrast, mainland *S. tzumu* populations inhabit low mountainous in southeastern China, adapted to relatively dry habitats with distinct seasonal temperature variations [39]. Over time, the two lineages developed non-overlapping niches, eliminating the ecological basis for mainland *S. tzumu* to re-colonize Taiwan.

## 4. Materials and Methods

### 4.1. Plant Sampling and DNA Extraction of *Sassafras*

To extract DNA, fresh young leaflets from 10 *Sassafras* individuals (5 of *S. albidum*, 3 of *S. tzumu*, and 2 of *S. randaiense*) were collected and preserved in silica gel for drying. Specific sampling details and voucher specimen numbers are provided in Table S1. Genomic DNA was isolated from these silica-dried leaf samples using a modified CTAB protocol [40]. The integrity and quality of the extracted DNA were evaluated via agarose gel electrophoresis and further verified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Meanwhile, DNA concentration was quantified with a NanoDrop LITE spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).

### 4.2. Illumina Sequencing, Chloroplast Genome De Novo Assembly and Annotation of *Sassafras*

High-quality genomic DNA from each of the 10 *Sassafras* individuals was utilized to construct Illumina paired-end ( $2 \times 150$  bp) sequencing libraries. These libraries were sequenced on a HiSeq Xten platform (Illumina, San Diego, CA, USA) in a single lane, with the sequencing work conducted at the Beijing Genomics Institute (BGI, Shenzhen, China). For each individual, approximately 10 Gb of raw sequencing data were generated. Clean data were obtained using the NGS QC Tool Kit v.2.3.3, which involved filtering out adapters and low-quality reads with a Q-value  $\leq 20$ . The complete cp genomes of the 10 individuals were de novo assembled using GetOrganelle v.1.7.6.1 [41] with default parameters. Annotation of these cp genomes was performed via GeSeq v.2.0.0 and CPGAVAS2 v.2.1.0 [42,43]. The annotated cp genomes were then submitted to the NCBI GenBank database, with their respective accession numbers listed in Table 1. Circular physical maps of the *Sassafras* cp genomes were constructed using OrganellarGenomeDRAW v.1.3.1 [44], followed by manual adjustments for accuracy.

### 4.3. Comparative Chloroplast Genome Analysis of *Sassafras*

To determine the sequence divergence levels among *Sassafras* species, we compared the 10 cp genomes using the mVISTA program (<https://genome.lbl.gov/vista/index.shtml>, accessed on 1 July 2024) [45]. The analysis was performed under the Shuffle-LAGAN mode, with the annotated cp genome of *S. tzumu* (GenBank ID: NC045268) serving as the reference.

To explore genome-wide evolutionary dynamics and structural variations across the 10 *Sassafras* individuals, MAUVE v.1.1.1 (<https://darlinglab.org/mauve/mauve.html>, accessed on 3 July 2024) [46] was utilized to detect key evolutionary events in multiple sequence alignments, including gene loss, duplication, rearrangement, and translocation. Additionally, IRscope (<https://irscope.shinyapps.io/irapp/>, accessed on 3 July 2024) [47] was employed to visualize the overall cp genome structure and track size variations at the

boundary regions between inverted repeat (IR), small single copy (SSC), and large single copy (LSC) regions.

Codon usage patterns and relative synonymous codon usage (RSCU) values [48] were calculated for all protein-coding genes in the 10 cp genomes using CodonW v.1.4.2 (<http://codonw.sourceforge.net/>, accessed on 8 July 2024) [49]. For this analysis, genes shorter than 300 bp and duplicated genes were excluded. Prior to computation, two non-degenerate unique codons (AUG and UGG) and three stop codons (TAA, TAG, TGA) were removed to ensure accuracy.

#### 4.4. Identification of cp Microsatellites, Repeats, and DNA Barcodes of *Sassafras*

To detect chloroplast microsatellite markers (simple sequence repeats, SSRs) in the cp genomes of the 10 *Sassafras* individuals, we employed the MicroSatellite (MISA) perl script [50] with specific parameter settings: 10 repeat units for mononucleotide SSRs, 6 for dinucleotide SSRs, and 5 for tri-, tetra-, penta-, and hexa-nucleotide SSRs.

REPuter [51] was utilized to quantify four types of dispersed sequence repeats (complement, forward, reverse, and palindromic repeats) in the 10 cp genomes, with parameters set to a minimum repeat length of 50 bp and a Hamming distance of 8.

For identifying potential DNA barcodes of *Sassafras*, DnaSP v.6.0 [52] was used to calculate nucleotide variability ( $\pi$ ) in both coding regions and non-coding regions. Prior to this, sequences were aligned using MAFFT v.7 [53], and only those with an aligned length >200 bp and  $\geq 1$  mutation site were included. The resulting  $\pi$  values were visualized using R v.4.0.2.

#### 4.5. Phylogenetic Analysis

To clarify the phylogenetic placement of *Sassafras* within the Lauraceae family, we analyzed 10 *Sassafras* individuals alongside 96 published complete chloroplast (cp) genomes from 30 Lauraceae genera (retrieved from NCBI; detailed information in Table S2). *Calycanthus floridus* and *C. chinensis* were designated as outgroups, and sequence alignment was performed using MAFFT v.7 [54].

Phylogenetic trees were constructed via maximum likelihood (ML) and Bayesian inference (BI) methods, based on two datasets: the complete cp genome and concatenated sequences of 43 common protein-coding genes (CDS). These 43 genes were identified from the 108 cp genomes using PhyloSuite [55,56], which was also employed for multiple sequence alignment and sequence concatenation to generate the CDS dataset.

Model selection and parameterization were based on BIC values calculated by jModelTest v.2.1.10 [57,58]. The ML tree was built with IQ-TREE (<http://www.iqtree.org/>) under the GTR + R3 + F model, with 1000 ultrafast bootstrap replicates [59]. For the BI tree, MrBayes (<http://nbisweden.github.io/MrBayes/>, accessed on 11 July 2024) was used with the GTR + I + G + F model [60]; Markov Chain Monte Carlo (MCMC) runs lasted 1,000,000 generations (sampled every 1000 generations), with the first 25% of trees discarded as burn-in. The remaining trees were used to construct a 50% consensus tree, with posterior probabilities (PPs) estimated from the top 25% best-scoring trees. Convergence was confirmed when the average standard deviation of split frequencies fell below 0.01. For the ML analysis, two independent searches were conducted to verify topological consistency, with nodal support assessed via 1000 bootstrap (BS) replicates per run. Finally, the phylogenetic tree topology was visualized using the iTOL online tool (<https://itol.embl.de/>, accessed on 28 July 2024) [61].

#### 4.6. Divergence Time Estimation

We used the BEAST v.2.6.3 to estimate the divergence time of *Sassafras* species based on the dataset composed of the concatenated sequences of 43 shared CDS [62]. For cali-



brating branch divergence times within the Cinnamomeae clade, we adopted the same fossil markers as reported by Xiao and Ge [25]: we used the fossil flower of *Virginianthus calycanthoides* to calibrate the crown age of the Lauraceae family to  $107.1 \pm 0.5$  Mya, the *Neusenien tetrasporangiata* Eklund fossil was used to calibrate the crown node age of the Neocinnamomum-Caryodaphnopsis core Lauraceae branch to  $83 \pm 1$  Mya, and the *Machilus maomingensis* fossil was used to calibrate the stem node age of *Machilus* to  $33.7 \pm 1$  Mya. The dataset file was imported into BEAUti and parameters were set as follows: GTR as the nucleotide substitution model, lognormal relaxed as the molecular clock model, Yule Model as the prior setting. The MCMC method (running 10,000,000 times, sampling every 1000 times) was used to estimate the divergence time, and an xml file was obtained. This file was run in BEAST v.2.6.3, and the resulting log file was checked for convergence using Tracer v.1.7.1 [63]; effective sample size (ESS)  $\geq 200$  indicated successful convergence, while ESS  $< 200$  required increasing the number of iterations. In Tree Annotator, the first 10% of samples were discarded as burn-in to produce a time-calibrated tree file, which was then visualized using FigTree v.1.4.0.

## 5. Conclusions

Our study first sequenced, assembled and conducted the comparative cp genomics of 10 individuals of three *Sassafras* siblings. We found the whole length of cp genomes of the 10 *Sassafras* had typical quadripartite structure, and conserved gene arrangements and contents. Four hypervariable intergenic regions (*ccsA-ndhD*, *trnH-psbA*, *rps15-ycf1*, and *petA-psbJ*) and 672 cpSSRs were identified for the enrichment of genetic resources of *Sassafras*. Phylogenetic analysis based on 106 cp genomes from 30 genera within the Lauraceae family demonstrated that *Sassafras* constituted a monophyletic clade and grouped a sister branch with the *C. sect. Camphora* within the tribe Cinnamomeae. Divergence time between *S. albidum* and its East Asian siblings was estimated at the Middle Miocene (16.98 Mya), *S. tzumu* diverged from *S. randaiense* at the Pleistocene epoch (3.63 Mya). Combined with fossil evidence, our results further revealed the crucial role of the Bering Land Bridge and glacial refugia in the speciation and differentiation of *Sassafras*. Overall, our study clarified the evolution pattern of *Sassafras* cp genomes and elucidated the phylogenetic position and divergence time framework of *Sassafras*.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijms26157357/s1>.

**Author Contributions:** Conceptualization, Z.L. and Z.W.; Methodology, Z.L. and Y.Z.; Software, Y.T., Y.Z., D.Y.P.T. and Z.L.; Validation, Y.Z., D.Y.P.T. and Z.W.; Formal Analysis, Z.W., Y.Z. and J.Z.; Investigation, Y.Z. and J.Z.; Resources, D.Y.P.T. and Z.W.; Data Curation, Y.T., Y.Z. and D.Y.P.T.; Writing—Original Draft Preparation, Z.L. and Y.Z.; Writing—Review and Editing, D.Y.P.T., Y.Z. and Z.W.; Visualization, Y.Z., Q.C., Z.L. and Y.W.; Supervision, Z.W. and Y.Z.; Project Administration, Y.Z. and Z.W.; Funding Acquisition, Z.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Jiangsu Funding Program for Excellent Postdoctoral Talent (2023ZB389).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available in the article and Supplementary Materials. The whole chloroplast genome data this study is openly available in GenBank of NCBI at <https://www.ncbi.nlm.nih.gov>. The accession numbers are recorded in Table 1.

**Acknowledgments:** The authors sincerely thank Pan Li for his great help in collecting plant materials and Shan Lu for his helpful instruction for preserving the plant specimen at the herbarium of Nanjing University.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Presl, J.S. *Sassafras*. In *O Prirozenosti Rostlin 2. k.w.*; Berchtold, F., Presl, J.S., Eds.; Endersa: Praha, Czech Republic, 1825; p. 30.
- Hemsley, W.B. *Sassafras* in China. (*Sassafras tzumu*, Hemsl.). *Bull. Misc. Inf. (R. Gard. Kew)* **1907**, 1907, 55–56. [[CrossRef](#)]
- Rehder, A. The American and Asiatic Species of *Sassafras*. *J. Arnold Arbor.* **1920**, 1, 242–245. [[CrossRef](#)]
- Wen, J. Evolution of Eastern Asian and Eastern North American Disjunct Distributions in Flowering Plants. *Annu. Rev. Ecol. Syst.* **1999**, 30, 421–455. [[CrossRef](#)]
- Wen, J. Evolution of Eastern Asian–North American Biogeographic Disjunctions: A Few Additional Issues. *Int. J. Plant Sci.* **2001**, 162, S117–S122. [[CrossRef](#)]
- Berry, E.W. *A Revision of the Flora of the Latah Formation*; US Government Printing Office: Washington, DC, USA, 1929.
- Poole, I.; Richter, H.G.; Francis, J.E. Evidence for Gondwanan Origins for *Sassafras* (Lauraceae)? Late Cretaceous Fossil Wood of Antarctica. *IAWA J.* **2000**, 21, 463–475. [[CrossRef](#)]
- Kostermans, A.J.G.H. Lauraceae. *Commun. (Pengumuman) For. Res. Inst.* **1957**, 57, 1–64.
- Van der Werff, H.; Richter, H.G. Toward an Improved Classification of Lauraceae. *Ann. Mo. Bot. Gard.* **1996**, 83, 409–418. [[CrossRef](#)]
- Rohwer, J.G. Toward a Phylogenetic Classification of the Lauraceae: Evidence from matK Sequences. *Syst. Bot.* **2000**, 25, 60–71. [[CrossRef](#)]
- 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](#)]
- 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.
- Li, J.; Christophel, D.C.; Conran, J.G.; Li, H.W. Phylogenetic Relationships within the ‘Core’ Laureae (*Litsea complex*, Lauraceae) Inferred from Sequences of the Chloroplast Gene matK and Nuclear Ribosomal DNA ITS Regions. *Plant Syst. Evol.* **2004**, 246, 193–214. [[CrossRef](#)]
- Nie, Z.L.; Wen, J.; Sun, H. Phylogeny and Biogeography of *Sassafras* (Lauraceae) Disjunct between Eastern Asia and Eastern North America. *Plant Syst. Evol.* **2007**, 267, 191–203. [[CrossRef](#)]
- Rohde, R.; Rudolph, B.; Ruthe, K.; Lorea-Hernández, F.G.; de Moraes, P.L.R.; Li, J.; Rohwer, J.G. Neither *Phoebe* nor *Cinnamomum*—The Tetrasporangiate Species of *Aiouea* (Lauraceae). *Taxon* **2017**, 66, 1085–1111. [[CrossRef](#)]
- Song, Y.; Yu, W.B.; Tan, Y.H.; Liu, B.; Yao, X.; Jin, J.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](#)]
- Zhao, M.L.; Song, Y.; Ni, J.; Yao, X.; Tan, Y.H.; Xu, Z.F. Comparative Chloroplast Genomics and Phylogenetics of Nine *Lindera* Species (Lauraceae). *Sci. Rep.* **2018**, 8, 8844. [[CrossRef](#)] [[PubMed](#)]
- Jo, S.; Kim, Y.K.; Cheon, S.H.; Fan, Q.; Kim, K.J. Characterization of 20 Complete Plastomes from the Tribe Laureae (Lauraceae) and Distribution of Small Inversions. *PLoS ONE* **2019**, 14, e0224622. [[CrossRef](#)] [[PubMed](#)]
- 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](#)]
- 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.* **2020**, 58, 423–439. [[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](#)]
- Song, Y.; Yu, Q.-F.; Zhang, D.; Chen, L.-G.; Tan, Y.-H.; Zhu, W.; Su, H.-L.; Yao, X.; Liu, C.; Corlett, R.T. New Insights into the Phylogenetic Relationships within the Lauraceae from Mitogenomes. *BMC Biol.* **2024**, 22, 241. [[CrossRef](#)] [[PubMed](#)]
- Li, L.; Liu, B.; Song, Y.; Meng, H.-H.; Ci, X.-Q.; Conran, J.G.; de Kok, R.P.J.; de Moraes, P.L.R.; Ye, J.-W.; Tan, Y.-H.; et al. Global Advances in Phylogeny, Taxonomy and Biogeography of Lauraceae. *Plant Divers.* **2025**, 47, 341–364. [[CrossRef](#)] [[PubMed](#)]
- Luo, A.; Zhang, C.; Zhou, Q.-S.; Ho, S.Y.W.; Zhu, C.-D.; Ree, R. Impacts of Taxon-Sampling Schemes on Bayesian Tip Dating Under the Fossilized Birth-Death Process. *Syst. Biol.* **2023**, 72, 781–801. [[CrossRef](#)] [[PubMed](#)]
- Xiao, T.W.; Ge, X.J. Plastome Structure, Phylogenomics, and Divergence Times of Tribe *Cinnamomeae* (Lauraceae). *BMC Genom.* **2022**, 23, 642. [[CrossRef](#)]

26. Qin, Z.; Zheng, Y.J.; Gui, L.J.; Xie, G.A.; Wu, Y.F. Codon Usage Bias Analysis of Chloroplast Genome of Camphora Tree (*Cinnamomum camphora*). *Guizhou Sci.* **2018**, *38*, 1346–1355. [\[CrossRef\]](#)
27. Li, H.W.; Liu, B.; Davis, C.C.; Yang, Y. Plastome Phylogenomics, Systematics, and Divergence Time Estimation of the *Beilschmiedia* Group (Lauraceae). *Mol. Phylogenet. Evol.* **2020**, *151*, 106901. [\[CrossRef\]](#)
28. Li, H.-T.; Luo, Y.; Gan, L.; Ma, P.-F.; Gao, L.-M.; Yang, J.-B.; Cai, J.; Gitzendanner, M.A.; Fritsch, P.W.; Zhang, T.; et al. Plastid Phylogenomic Insights into Relationships of All Flowering Plant Families. *BMC Biol.* **2021**, *19*, 232. [\[CrossRef\]](#)
29. Yang, Z.; Liu, B.; Yang, Y.; Ferguson, D.K. Phylogeny and Taxonomy of *Cinnamomum* (Lauraceae). *Ecol. Evol.* **2022**, *12*, e9378. [\[CrossRef\]](#)
30. Nie, Z.-L.; Wen, J.; Azuma, H.; Qiu, Y.-L.; Sun, H.; Meng, Y.; Sun, W.-B.; Zimmer, E.A. Phylogenetic and Biogeographic Complexity of Magnoliaceae in the Northern Hemisphere Inferred from Three Nuclear Data Sets. *Mol. Phylogenet. Evol.* **2008**, *48*, 1027–1040. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Nie, Z.-L.; Sun, H.; Beardsley, P.M.; Olmstead, R.G.; Wen, J. Evolution of Biogeographic Disjunction between Eastern Asia and Eastern North America in *Phryma* (Phrymaceae). *Am. J. Bot.* **2006**, *93*, 1343–1356. [\[CrossRef\]](#)
32. McElwain, J.C.; Yiotis, C.; Lawson, T. Using Modern Plant Trait Relationships between Observed and Theoretical Maximum Stomatal Conductance and Vein Density to Examine Patterns of Plant Macroevolution. *New Phytol.* **2016**, *209*, 94–103. [\[CrossRef\]](#)
33. Chou, Y.-W.; Thomas, P.I.; Ge, X.-J.; LePage, B.A.; Wang, C.-N. Refugia and Phylogeography of *Taiwania* in East Asia. *J. Biogeogr.* **2011**, *38*, 1992–2005. [\[CrossRef\]](#)
34. Qiu, Y.-X.; Fu, C.-X.; Comes, H.P. Plant Molecular Phylogeography in China and Adjacent Regions: Tracing the Genetic Imprints of Quaternary Climate and Environmental Change in the World's Most Diverse Temperate Flora. *Mol. Phylogenet. Evol.* **2011**, *59*, 225–244. [\[CrossRef\]](#)
35. Zhisheng, A.; Kutzbach, J.E.; Prell, W.L.; Porter, S.C. Evolution of Asian Monsoons and Phased Uplift of the Himalaya–Tibetan Plateau since Late Miocene Times. *Nature* **2001**, *411*, 62–66. [\[CrossRef\]](#)
36. Bodare, S.; Stocks, M.; Yang, J.-C.; Lascoux, M. Origin and Demographic History of the Endemic Taiwan Spruce (*Picea morrisonicola*). *Ecol. Evol.* **2013**, *3*, 3320–3333. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Gao, J.; Yu, T.; Li, J.Q. Phylogenetic and Biogeographic Study of *Acer* L. section *Palmata* Pax (Sapindaceae) Based on Three Chloroplast DNA Fragment Sequences. *Acta Ecol. Sin.* **2020**, *40*, 5992–6000. [\[CrossRef\]](#)
38. Chung, K.F.; van der Werff, H.; Peng, C.I. Observations on the Floral Morphology of *Sassafras randaiense* (Lauraceae). *Ann. Mo. Bot. Gard.* **2010**, *97*, 1–10. [\[CrossRef\]](#)
39. Guan, B.; Liu, Q.; Liu, X.; Gong, X. Environment Influences the Genetic Structure and Genetic Differentiation of *Sassafras tzumu* (Lauraceae). *BMC Ecol. Evol.* **2024**, *24*, 80. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Doyle, J.J.; Doyle, J.L. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
41. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
42. 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. [\[CrossRef\]](#)
43. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq—Versatile and Accurate Annotation of Organelle Genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Lohse, M.; Drechsel, O.; Bock, R. OrganellarGenomeDRAW (OGDRAW): A Tool for the Easy Generation of High-Quality Custom Graphical Maps of Plastid and Mitochondrial Genomes. *Curr. Genet.* **2007**, *52*, 267–274. [\[CrossRef\]](#)
45. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational Tools for Comparative Genomics. *Nucleic Acids Res.* **2004**, *32* (Suppl. S2), W273–W279. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Darling, A.C.E.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [\[CrossRef\]](#)
47. Amiryousefi, A.; Hyvönen, J.; Pocai, P. IRscope: An Online Program to Visualize the Junction Sites of Chloroplast Genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [\[CrossRef\]](#)
48. Wright, F. The 'Effective Number of Codons' Used in a Gene. *Gene* **1990**, *87*, 23–29. [\[CrossRef\]](#)
49. Sharp, P.M.; Li, W.H. The Codon Adaptation Index—A Measure of Directional Synonymous Codon Usage Bias, and Its Potential Applications. *Nucleic Acids Res.* **1987**, *15*, 1281–1295. [\[CrossRef\]](#)
50. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A Web Server for Microsatellite Prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [\[CrossRef\]](#) [\[PubMed\]](#)
51. 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. [\[CrossRef\]](#)
52. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [\[CrossRef\]](#)

53. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [\[CrossRef\]](#)
54. Nakamura, T.; Yamada, K.D.; Tomii, K.; Katoh, K. Parallelization of MAFFT for Large-Scale Multiple Sequence Alignments. *Bioinformatics* **2018**, *34*, 2490–2492. [\[CrossRef\]](#)
55. Zhang, D.; Gao, F.; Jakovlić, I.; Zhou, 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. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Xiang, C.; Gao, F.; Jakovlić, I.; Lei, H.; Hu, Y.; Zhang, H.; Zou, H.; Wang, G.; Zhang, D. Using PhyloSuite for Molecular Phylogeny and Tree-Based Analyses. *iMeta* **2023**, *2*, e87. [\[CrossRef\]](#)
57. Guindon, S.; Gascuel, O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.* **2003**, *52*, 696–704. [\[CrossRef\]](#)
58. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest2: More Models, New Heuristics and Parallel Computing. *Nat. Methods* **2012**, *9*, 772. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [\[CrossRef\]](#)
60. Ronquist, F.; Teslenko, M.; Van Der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, 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. [\[CrossRef\]](#)
61. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v4: Recent Updates and New Developments. *Nucleic Acids Res.* **2019**, *47*, 256–259. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Bouckaert, R.; Heled, J.; Kühnert, D.; Vaughan, T.; Wu, C.-H.; Xie, D.; Suchard, M.A.; Rambaut, A.; Drummond, A.J.; Prlic, A. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comp. Biol.* **2014**, *10*, e1003537. [\[CrossRef\]](#)
63. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. [\[CrossRef\]](#) [\[PubMed\]](#)

**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.