

Article

Employing Genome-Wide SNP Discovery to Characterize the Genetic Diversity in *Cinnamomum camphora* Using Genotyping by Sequencing

Xue Gong^{1,2}, Aihong Yang¹, Zhaoxiang Wu¹, Caihui Chen¹, Huihu Li¹, Qiaoli Liu¹, Faxin Yu¹ and Yongda Zhong^{1,*}

¹ Key Laboratory of Horticultural Plant Genetic and Improvement of Jiangxi Province, Institute of Biological Resources, Jiangxi Academy of Sciences, No. 7777, Changdong Road, Gaoxin District, Nanchang 330096, China; gx13864477189@163.com (X.G.); yanglai2hong3@163.com (A.Y.); wuzhaoxiang@jxas.ac.cn (Z.W.); chenhui_1988@126.com (C.C.); lhh26521@163.com (H.L.); 15270996528@163.com (Q.L.); yufaxin@jxas.ac.cn (F.Y.)

² College of Forestry, Jiangxi Agricultural University, Nanchang 330045, China

* Correspondence: zhongyongda@jxas.ac.cn; Tel.: +86-791-8817-5720

Abstract: *Cinnamomum camphora* (L.) J.Presl is a representative tree species of evergreen broad-leaved forests in East Asia and has exceptionally high economic, ornamental, and ecological value. However, the excessive exploitation and utilization of *C. camphora* trees have resulted in the shrinking of wild population sizes and rare germplasm resources. In this study, we characterized 171 *C. camphora* trees from 39 natural populations distributed throughout the whole of China and one Japanese population. We investigated genetic diversity and population structure using genome-wide single-nucleotide polymorphism (SNP) identified by genotyping by sequencing (GBS) technology. The results showed the genetic diversity of the *C. camphora* populations from western China > central China > eastern China. Moreover, the Japanese population showed the highest diversity among all populations. The molecular variance analysis showed 92.03% of the genetic variation within populations. The average pairwise F_{ST} was 0.099, and gene flow N_m was 2.718, suggesting a low genetic differentiation among populations. Based on the genetic clustering analysis, the 40 *C. camphora* populations clustered into three major groups: Western China, Central China, and Eastern China + Japan. Eastern China's population had the closest genetic relationship with the Japanese population, suggesting possible gene exchange between the two adjacent areas. This study furthers our understanding of the genetic diversity and genetic structure of *C. camphora* in East Asia and provides genetic tools for developing strategies of *C. camphora* germplasm utilization.

Keywords: *Cinnamomum camphora*; genotyping by sequencing (GBS); genetic diversity; population structure; single nucleotide polymorphism (SNP)



Citation: Gong, X.; Yang, A.; Wu, Z.; Chen, C.; Li, H.; Liu, Q.; Yu, F.; Zhong, Y. Employing Genome-Wide SNP Discovery to Characterize the Genetic Diversity in *Cinnamomum camphora* Using Genotyping by Sequencing. *Forests* **2021**, *12*, 1511. <https://doi.org/10.3390/f12111511>

Academic Editors: Panteleimon Xofis and Konstantinos Poirazidis

Received: 12 October 2021

Accepted: 31 October 2021

Published: 1 November 2021

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



Copyright: © 2021 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

Cinnamomum camphora is a representative tree species of tropical and subtropical evergreen broad-leaved forests of the genus *Cinnamomum* in the Lauraceae family. It is one of the “bulk” trees with three major uses, including a source of spice, timber, and gardening. *C. camphora* is widely distributed in subtropical China and central and southern Japan [1,2]. Subtropical China is the most abundant wild resource and the main area of production of natural *C. camphora* essential oil globally, accounting for more than 80% of the global production [3]. However, since the 1950s, the destructive harvesting of essential oil by digging the trees and grubbing their roots, and the increasing demand for urban vegetation has led to severe reductions of the resources. At present, the natural distribution of *C. camphora* is scattered, and its mature natural forests are rare, which results in endangered high-quality essential oil containing precious chemical types. Therefore, it is urgent to collect and preserve the germplasm resources of *C. camphora*.

A previous study by Kameyama et al. showed rich genetic diversity in *C. camphora* populations from China and Japan using simple sequence repeat (SSR) markers. The genetic diversity of *C. camphora* populations in China was slightly higher than that in Japan [2]. We previously used transcriptome-based expressed sequence tag–derived SSR (EST-SSR) markers to investigate the genetic diversity and genetic structure of *C. camphora* in southern China. This study showed a moderate genetic diversity, prominent population structure, and genetic differentiation existing mainly within a population [4,5]. The above results revealed the significant genetic differentiation among and within the populations of *C. camphora*. However, due to the large differences in the populations' sources and sizes, previous genetic diversity studies vary significantly. Additionally, the shortcomings of SSR markers, including the unclarity of homozygosity and heterozygosity and invalid alleles, can lead to misidentification of SSR genotypes [6].

Due to their low mutation frequency, fewer alleles, high genetic stability, and distribution throughout the genome, single nucleotide polymorphisms (SNP) have been favoured in plant genetic diversity and genomic analysis [7]. The rapid development of next-generation sequencing (NGS) technology has reduced the cost and improved the efficiency of SNP discovery, which provides great potential for high-throughput SNP genotyping in plant breeding [8]. Wu et al. firstly used the genotyping by sequencing (GBS) technique to identify and genetically analyze plant species [9]. GBS is an NGS-based SNP genotyping technology suitable for species with high diversity and large genomes [10,11]. Due to the lack of need for the reference genome, low cost, simple procedures, and the SNP carrying rich genome-wide information, GBS is widely used in the analysis of plant genetic diversity [12,13], the study of systematic evolution [14], molecular marker localization [15,16], and genetic map construction [17].

The whole genome of *C. camphora* has not been sequenced yet, and the genetic diversity and structure of the *C. camphora* populations across its entire natural distribution area are still unclear. This study characterized 171 *C. camphora* trees from 39 natural populations from the whole distribution region in China and one population in Japan using genome-wide SNP molecular markers obtained by the GBS technique. We aimed to reveal the genetic diversity and structure of *C. camphora* in different regions, understand the sources and patterns of genetic variation, and provide a theoretical basis for protection strategies and genetic improvement of *C. camphora* germplasm resources.

2. Materials and Methods

2.1. Plant Material and DNA Isolation

Leaf samples from 171 *C. camphora* trees were collected in this study from 39 natural populations in the whole range of distribution of *C. camphora* in China and one population in Japan. Sample sources were grouped into four regions based on geographical origin. There were three populations with a total of 14 samples from the western region in China, six populations with a total of 25 samples from the central region, 30 populations with 128 samples from the eastern region, and one population with four samples from Japan (Figure 1 and Table 1). Trees with a diameter at breast height greater than 60 cm were randomly selected, and the distance between them was greater than 50 m. Five healthy leaves were collected from each tree, dried with silica gel. Details on sampling were summarized in Table S1. DNA was extracted using the ionic detergent cetyltrimethylammonium bromide, and the purity and integrity of the DNA were detected by 1% agarose gel electrophoresis. The DNA quality and quantity were determined by a NanoDrop 2000 (Thermo Fisher Scientific Inc., Waltham, MA, USA) and a Qubit[®] 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA), respectively.

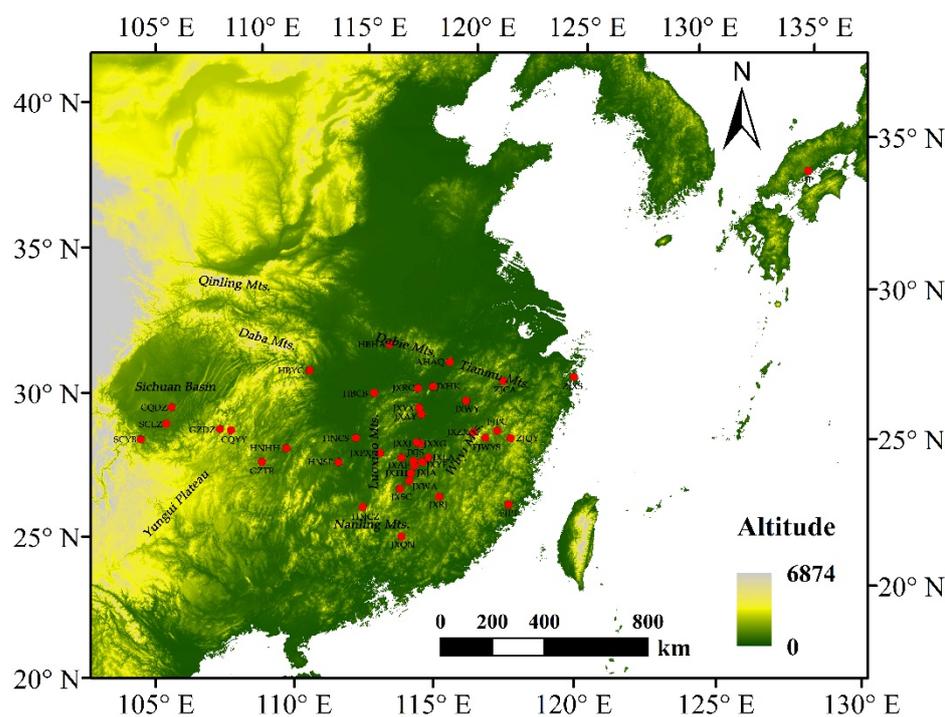


Figure 1. Geographical distribution of the plant materials. Solid dots represent populations collected in natural distribution areas in China and Japan.

2.2. GBS Library Construction and Population SNP Identification

Genomic DNA (0.1–11 μ g) was digested with the *ApeKI* restriction endonuclease, and the adaptor and barcode were ligated at both ends by T4 DNA ligase. Then each DNA sample was amplified by PCR. After mixing the samples, the DNA fragments (350–400 bp) were recovered and purified by agarose gel electrophoresis, and the required fragments were selected for the construction of the GBS library. Finally, paired-end sequencing at 150 bp per read was performed on the Illumina HiSeq 2500 sequencing platform (Illumina, San Diego, CA, USA).

Stacks 1.44 [18] was used to cluster the high-quality data of each individual based on sequence similarity. During clustering, low-depth stacks were filtered out, leaving high-depth stacks as the clustering results. The mismatched bases were corrected in each stack to eliminate the impact of sequencing errors and avoid false-positive SNP. Clean data of sample ZJCA-157 and JXAF-115 were selected to be combined and assembled as the reference genome under the parameter (ustacks -t gzfasta -i 1 -m 1 -M 6 -N 6 -p 6). BWA (Burrows-Wheeler-Aligner) was used to compare the high-quality and effective sequencing data against the reference genome (using the assembly results of samples ZJCA-157 and JXAF-115 as the reference genome) [19]. The SNP in the population were detected with SAMTOOLS software. The SNP in the population to be tested were detected using the Bayesian model. The obtained SNP were filtered with the SAMTOOLS [20] settings main parameter -minDP4-max miss0.1-maf0.01. The number of bases support for each SNP was not less than 1. Due to repeat regions in the genome, the number of bases support for each SNP should not be greater than 7000 to reduce the number of false-positive SNP caused by repeat region alignment errors. The genotype quality value was ≥ 5 .

2.3. Statistical Analysis of Data

The distance matrix was calculated using TreeBest-1.9.2 software (<http://treesoft.sourceforge.net/treebest.shtml>, accessed on 28 October 2021). The phylogenetic tree was constructed through 1000 calculations using the neighbour-joining method [21]. The eigenvectors and eigenvalues were calculated using GCTA (<http://cnsgenomics.com/software/gcta/pca.html>, accessed on 28 October 2021) software, and the principal component anal-

ysis (PCA) distribution map was plotted using R software [22]. The population genetic structure and individual allocation information were constructed using frappe software [23]. The maximum iteration of the expectation-maximization algorithm was set to 10,000 in the Frappe analysis. We predefined the number of genetic clusters from $K = 2$ to $K = 4$. Based on the filtered SNP loci, genetic diversity indicators including the observed heterozygosity (H_o), expected heterozygosity (H_e), interpopulation genetic differentiation coefficient (F_{ST}), and interindividual inbreeding coefficient (F_{IS}) of each population were calculated by Arlequin software [24]. Analysis of molecular variance (AMOVA) was implemented in GenALEx v6.5 [25]. Analysis of population nucleotide diversity (π) was performed with Stacks 1.44 [18].

Table 1. Geographical locations, sample size and sample numbers of 40 *C. camphora* populations in China and Japan.

Population Code	Sample Size	Locations	Latitude (°N)	Longitude (°E)
Western China	14			
SCLZ	5	Luzhou, Sichuan	28.967	105.433
SCYB	5	Yibing, Sichuan	28.417	104.417
CQDZ	4	Dazu, Chongqing	29.533	105.650
Central China	25			
CQYY	3	Youyang, Chongqing	28.700	108.000
GZDZ	3	Daozhen, Guizhou	28.750	107.567
GZTR	4	Tongren, Guizhou	27.567	109.183
HBYS	5	Yichang, Hubei	30.650	111.283
HNHH	5	Huaihua, Hunan	28.000	110.167
Eastern China	128			
HBCB	4	Chibi, Hubei	29.700	113.817
HBHA	5	Hongan, Hubei	31.333	114.633
AHAQ	5	Anqing, Anhui	30.500	117.000
HNCZ	4	Chenzhou, Hunan	25.783	112.983
HNCS	5	Changsha, Hunan	28.200	112.933
JYX	3	Yongxiu, Jiangxi	29.033	115.550
JXAY	4	Anyi, Jiangxi	28.800	115.617
JXHK	4	Hukou, Jiangxi	29.717	116.200
JXWY	4	Wuyuan, Jiangxi	29.083	117.450
JXRC	3	Ruichang, Jiangxi	29.717	115.600
JXPX	4	Pingxiang, Jiangxi	27.600	113.833
JXLA	5	Lean, Jiangxi	27.283	115.700
JXZX	4	Zixi, Jiangxi	27.983	117.583
JXQN	3	Quannan, Jiangxi	24.633	114.367
JXRJ	5	Ruijin, Jiangxi	25.883	115.950
JXSC	4	Suichuan, Jiangxi	26.300	114.483
JXTH	3	Taihe, Jiangxi	26.800	114.967
JXWA	5	Wanan, Jiangxi	26.550	114.867
JXAF	4	Anfu, Jiangxi	27.367	114.667
JXXJ	4	Xiajiang, Jiangxi	27.853	115.316
JXXG	5	Xingan, Jiangxi	27.800	115.450
JXYF	5	Yongfeng, Jiangxi	27.150	115.450
JXJA	5	Jian, Jiangxi	27.067	115.133
JXJS	4	Jishui, Jiangxi	27.217	115.117
FJPT	5	Putian, Fujian	25.317	118.567
FJPC	5	Pucheng, Fujian	27.917	118.517
FJWYS	4	Wuyishan, Fujian	27.733	118.017
ZJCA	4	Chunan, Zhejiang	29.600	119.033
ZJQY	4	Qingyuan, Zhejiang	27.600	119.000
ZJXS	5	Xiangshan, Zhejiang	29.367	121.867
Japan	4			
JP	4	Osaka, Japan	34.493	133.246
Total	171			

3. Results

3.1. GBS Analysis and SNP Identification

Total 138.24 Gb bases were generated by GBS sequencing from the 171 *C. camphora* trees (Table S2). After cleaning out low-quality sequences, 138.23 Gb of high-quality sequence data ($Q20 \geq 93.34\%$, $Q30 \geq 85.00\%$) remained, for an average of 0.81 Gb per sample. The average GC content was 38.24%. A total of 960 million clean reads were obtained,

ranging from 2,878,540 (HBCB-40) to 13,528,004 (HNCZ-55), with an average of 5,613,668 per sample. An average of 5,389,736 reads (96.18%) was aligned to a reference genome, with the highest (97.18%) alignment rate for GZDZ-18 and the lowest (73.50%) rate for ZJCA-157. The reference genome coverage was in an average of 11.82% $1\times$ genome (between 8.2% and 16.1%), and an average of 5.5% $4\times$ genome (between 4.12% and 7.41%). A total of 130,450 SNP loci were obtained based on the comparison and detection by SAMTOOLS software. After filtering, 14,718 high-quality SNP loci were used for further analysis.

3.2. Population Genetic Structure

To reveal the population genetic structure of *C. camphora*, we divided the 40 populations into four groups, eastern China, central China, western China, and Japan. The neighbour-joining tree showed that, except for the HNCZ-54 trees from the east and HNHH-31 from the west regions, the other 169 *C. camphora* trees were clustered into three clades (Figure 2). Cluster I was composed of 18 *C. camphora* trees, including 16 from the central region and two from the QDZ population in the western region (11, 12). Cluster II included 15 trees mainly from the west region, in which two trees from the central region (GZDZ-19 and GZDZ-20) and one from the eastern region (HNCZ-54). Cluster III contained the most samples (137), including trees from east China, and a few from the central region (HNSF-35, HNSF-38, HNSF-39, HBYC-26, HBYC-28, and HBYC-29), and all of the trees from Japan.

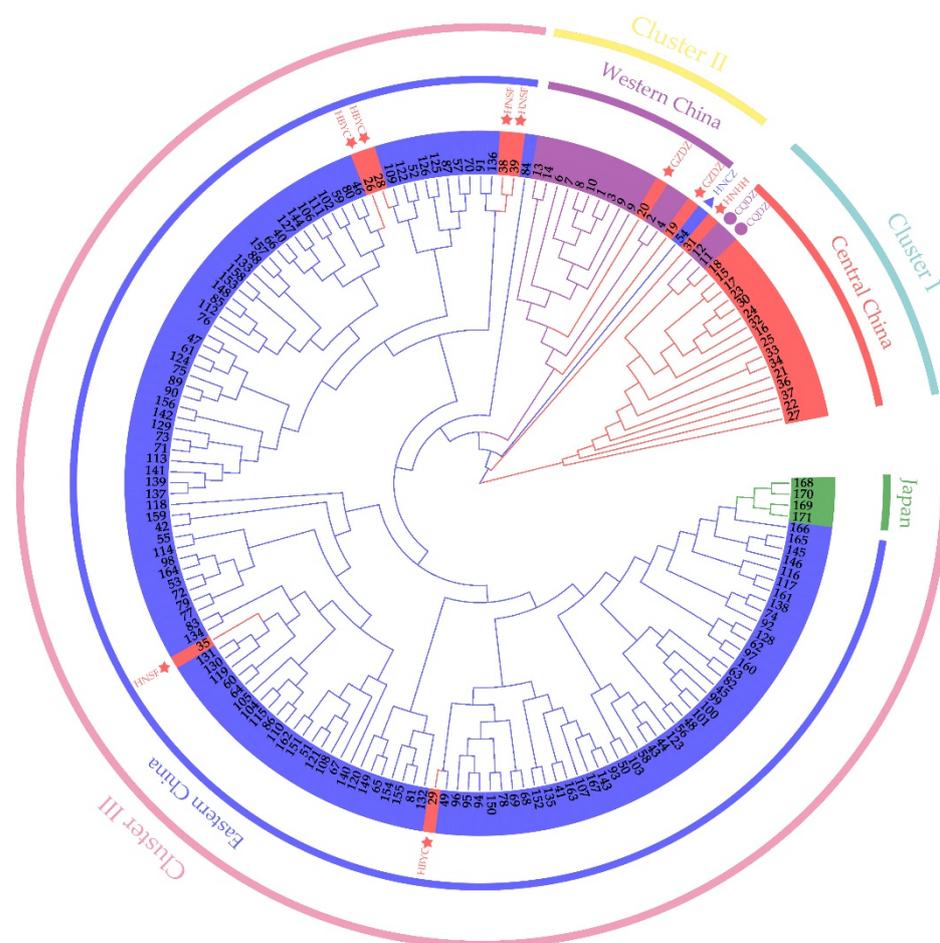


Figure 2. Neighbour-joining phylogenetic tree of *C. camphora* based on all single nucleotide polymorphisms (SNPs), with the evolutionary distances measured by p-distances with TreeBest. The populations from different geographic locations were coloured red (Central China), purple (Western China), blue (Eastern China), and green (Japan). ★, Individuals in Central China; ●, Individuals in Western China; ▲, Individuals in Eastern China.

To detect the highest population structure level, we performed the analysis when the parameter K was between 2 and 4 (Figure 3). The results showed that when $K = 2$, the trees from central China and the HNCZ-54, CQDZ-11, and CQDZ-12 from the western region were first separated from the others and formed subgroup I. When $K = 3$, trees from the west region were separated from subgroup II and formed subgroup III. When $K = 4$, subgroup IV was further divided into the populations from Japan and eastern China, and the Japanese *C. camphora* population was finally clustered into its own subgroup. These results are consistent with the above phylogenetic tree analysis.

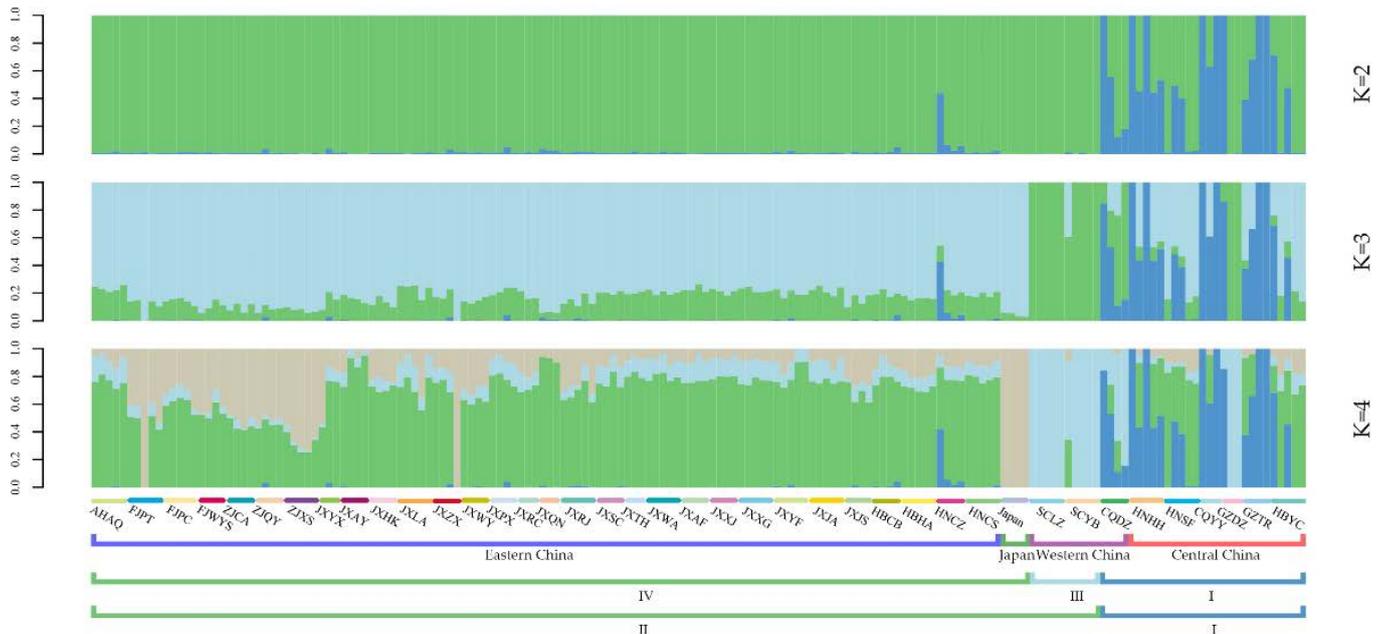


Figure 3. Population genetic structure of 171 *C. camphora* individuals. Each colour represents one cluster. Each individual is represented by a vertical bar, and the length of each coloured segment in each vertical bar represents the proportion contributed by ancestral populations.

PCA showed that the 171 *C. camphora* trees clustered three main groups (Figure 4). All trees can be significantly separated into two groups: central China and eastern China + Japan, according to the first principal component (PC1). In comparison, the western China populations showed a certain level of differentiation from the others two groups on the second principal component (PC2). This result is consistent with the clustering tree (Figure 2) and the population structure analysis (Figure 3). All these data indicated a clear geographical pattern of the genetic structure of the *C. camphora* populations.

3.3. Genetic Diversity in *C. camphora*

At the species level, the average values of H_o and H_e of *C. camphora* were 0.365 and 0.352, respectively (Table 2). The analysis of genetic diversity in each region showed that the population from Japan had the highest level of genetic heterozygosity ($H_o = 0.504$, $H_e = 0.442$), followed by Western China ($H_o = 0.333$, $H_e = 0.333$) and Central China ($H_o = 0.322$, $H_e = 0.327$), and the genetic diversity in eastern China was the lowest ($H_o = 0.302$, $H_e = 0.306$). The nucleotide diversity (π) of *C. camphora* in the different regions ranged from 0.230 (Japan) to 0.318 (Central China). The inbreeding coefficient F_{IS} ranged from -0.151 (Japan) to 0.020 (Eastern China), with an average of -0.037 .

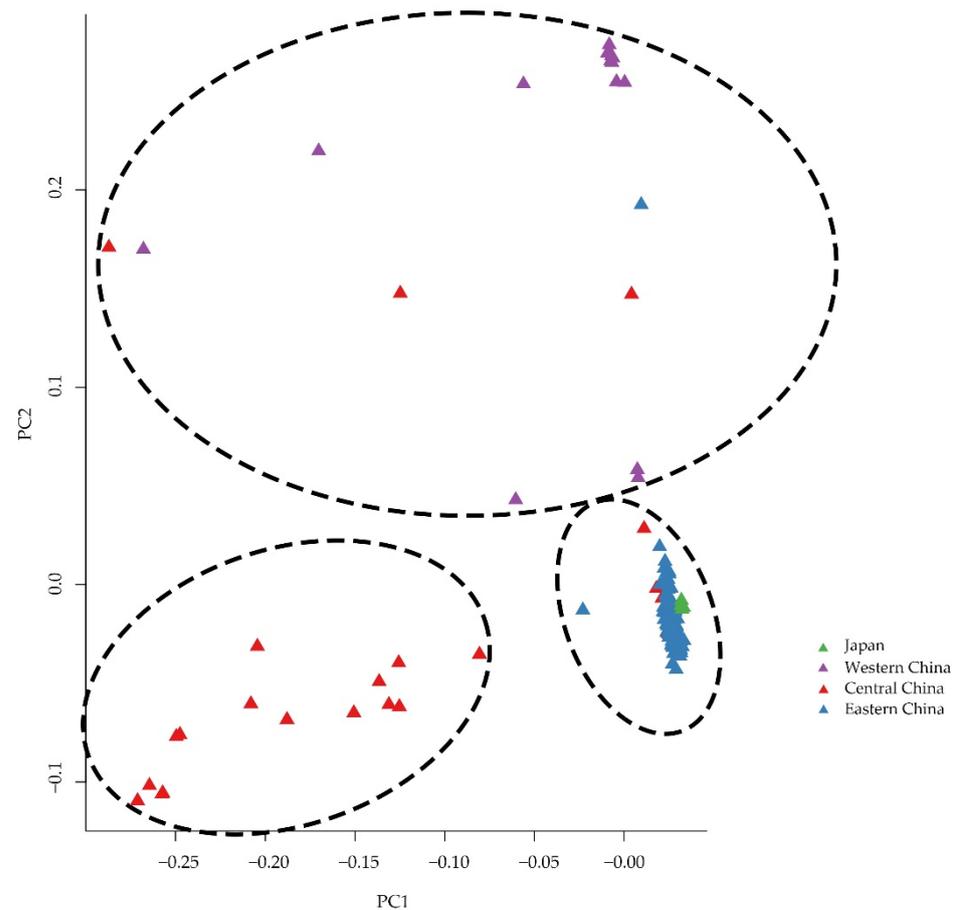


Figure 4. The principal coordinate analysis of *C. camphora* populations using all identified SNPs as markers. The populations from different geographic locations were coloured red (Central China), purple (Western China), blue (Eastern China), and green (Japan).

Table 2. Diversity statistics of four *C. camphora* regions.

Group	N	H_o	H_e	F_{IS}	π
Western China	14	0.333	0.333	0.011	0.288
Central China	25	0.322	0.327	−0.026	0.318
Eastern China	128	0.302	0.306	0.020	0.305
Japan	4	0.504	0.442	−0.151	0.230
Total	171	0.365	0.352	−0.037	0.285

Note: N , Number of individuals; H_o , Observed heterozygosity; H_e , Expected heterozygosity; F_{IS} , Fixation index; π , Total nucleotide diversity.

Analysis of molecular variance (AMOVA) of the *C. camphora* populations from the four regions (Table 3) showed that 7.02% of the genetic variation was between the regions. The genetic variation between the populations from the same region was only 0.95%. However, 92.03% of the genetic variation was among the individuals within each population. Among the four regions, the mean genetic differentiation coefficient F_{ST} was 0.099, gene flow (Nm) was 2.718, indicating a moderate genetic differentiation between populations. The genetic differentiation among eastern, western, and central China was relatively low ($F_{ST} = 0.052$ – 0.071 , 0.064 on average), less than the differentiation between China and Japan ($F_{ST} = 0.096$ – 0.152 , 0.133 on average). The genetic differentiation between the eastern China population and the Japanese population was low (0.096). In contrast, a relatively high level of genetic differentiation between western or central China and Japan was observed, 0.152 and 0.151, respectively. Gene exchange between different regions in China was relatively frequent ($Nm = 3.271$ – 4.558), but the gene flow between the Chinese population and the

Japanese population was limited ($Nm = 1.395\text{--}2.354$), indicating that geographical isolation has exacerbated the genetic differentiation between the *C. camphora* populations (Table 4).

Table 3. Analysis of molecular variance (AMOVA) for four regions *C. camphora*.

Source	Degree of Freedom	Sum of Squares	Variance Components	Percentage of Variation %
Among groups	3	27,901.615	154.380	7.02
Within groups among populations	167	345,029.070	20.920	0.95
Within populations	171	346,138.500	2024.202	92.03
Total	341	719,069.184	2199.502	

Table 4. Genetic differentiation of *C. camphora* across four regions (lower diagonal: pairwise F_{ST} ; upper diagonal: pairwise Nm).

	Western China	Central China	Eastern China	Japan
Western China	0	3.321	4.558	1.395
Central China	0.070	0	3.271	1.406
Eastern China	0.052	0.071	0	2.354
Japan	0.152	0.151	0.096	0

4. Discussion

4.1. Genotyping by Sequencing

Traditionally, molecular markers such as random amplified polymorphic DNA, amplified fragment length polymorphisms, and SSRs have been widely used to analyze plant genetic diversity. With the rapid development of NGS technology, a new method that is simple but powerful, GBS, has achieved high-efficiency and large-scale genome-wide SNP identification in some species [8]. In 2011, Elshire et al. used GBS to investigate maize (IBM) and barley (Oregon Wolfe Barley) populations. To establish a library with reduced genomic complexity, they identified approximately 200,000 and 25,000 sequence markers, respectively. The analysis showed the potential of this library for marker screening in large-genome species [10]. Pootakham et al. obtained 524,508,111 reads from oil palm (*Elaeis guineensis* Jacq.) using GBS technology, approximately 88% reads were aligned to the reference genome, and 21,471 SNP loci were identified, which was significantly lower than the alignment rate of reads on the reference genome in this study (96.18%) [26]. Gurcan et al. performed SNP detection on apricots by the GBS technique and generated approximately 28 Gb of high-quality data after filtering [27]. Wu et al. applied GBS technology to 75 samples of Lauraceae plants, including 19 samples of *C. kanehirae*, 31 of *C. camphora*, and 25 putative hybrids, for breed identification and genetic analysis. In total, 529,006 SNPs were obtained in the calling analysis, though after filtering, 840 high-quality SNPs were obtained for subsequent analysis. Among them, *C. camphora* groups in eastern and south-western Taiwan had 201 SNP loci, while the western and northern groups (including Vietnam and Japan) had 376 SNP loci [9]. In this study, by sequencing 171 *C. camphora* trees, we obtained 480 million high-quality reads and identified 130,450 SNPs. After filtering, 14,718 high-quality SNPs were used for subsequent genetic diversity and genetic structure analysis. The quality of the obtained reads and SNP loci was as high as the studies from Pootakham et al. and Gurcan et al.

4.2. Genetic Diversity of *C. camphora*

Genetic diversity refers to the degree of genetic polymorphism between different populations or individuals within a species. Rich genetic diversity is the basis for a species to respond to environmental changes and for people to carry out genetic improvement programs [28,29]. Based on 22 SSR markers, Kameyama evaluated the genetic polymorphism of 104 *C. camphora* samples from three Japanese populations: Meiji Jingu (Shinto Shrine), Kajiya Plantation, and Manazuru Peninsula. They found the average H_o , and H_e of each population was 0.53–0.60 and 0.55–0.68, respectively, suggesting a high level of genetic

diversity [1]. Kameyama et al. further performed genetic diversity analysis on 817 *C. camphora* individuals from six populations, including three populations (Fujian, Shanghai, and Taiwan) from China and three populations (Eastern and Western Kyushu regions) from Japan. This study showed that a high level of genetic diversity in the *C. camphora* population, in which the Chinese populations ($H_o = 0.728$, $H_e = 0.878$) showed significantly greater diversity than that of the Japanese populations ($H_o = 0.614$, $H_e = 0.714$) [2]. In comparison to the previous study, our previous studies showed a moderate level of genetic diversity ($H_o = 0.30$ – 0.61 , $H_e = 0.31$ – 0.48) of 41 Chinese *C. camphora* natural populations from the whole range of distribution of *C. camphora* in southern China using EST-SSR markers [4,5]. Interestingly, by using the GBS technique, Wu et al. observed a significantly lower genetic diversity in the population, including 75 individuals from Taiwan, 19 *C. kanehirae*, 31 *C. camphora*, and 25 putative hybrids. The population genetic diversity ($H_o = 0.108$, $H_e = 0.123$) of *C. camphora* in western and northern Taiwan was slightly higher than that in eastern and south-western Taiwan ($H_o = 0.085$, $H_e = 0.076$), and the H_o of *C. camphora* was greater than that of *C. kanehirae* ($H_o = 0.076$) [9]. In this study, we analyzed the genetic diversity of the Chinese and Japanese *C. camphora* populations through SNP markers generated by the GBS technique. We found that the genetic diversity of the Chinese *C. camphora* populations was relatively high ($H_o = 0.319$, $H_e = 0.322$), which is consistent with our previous findings based on SSR markers [4,5]. However, our populations showed higher genetic diversity than those found by Wu et al. [9]. One possible reason is that the geographical range of populations in Taiwan is small, and the cultivation history of *C. camphora* is long. During the process of the breeding program, a limited number of selected elite trees led to lower diversity [30]. Moreover, the asexual reproduction process using cuttings has further reduced the genetic diversity of *C. camphora*. In this study, a total of 11 Chinese *C. camphora* trees were clustered into different regions (Figure 2). And these trees were obtained from recent cultivations (Table S1). Kameyama et al. showed that human activity had significantly affected gene exchange between *C. camphora* populations [2]. The genetic diversity of a tree species is typically determined by its geographical distribution range and population size [31,32]. Although *C. camphora* is widely distributed in southern China, its origin is limited to East Asia and the Indochina Peninsula. The limited geographical distribution inhibits the gene exchange among populations and enhances genetic diversity within the population [33]. In addition, this study, together with the findings from Wu et al., showed lower genetic diversity of *C. camphora* in China than that of the Japanese population, which was not consistent with the results of Kameyama et al. [2]. This inconsistency may be due to different sampling strategies, such as inconsistencies in the source and size of the population used [32,34]. The difference between the GBS-SNPs and SSRs used in these studies could be another reason for the different findings.

4.3. Genetic Differentiation and Population Genetic Structure of *C. camphora*

Studying the genetic structure of plant populations is critical to the understanding and maintenance of plant genetic diversity. The genetic structure of populations is affected by the breeding system, genetic drift, population size, seed dispersal, gene flow, evolutionary history, and natural selection [2,35]. As an essential indicator of the population's differentiation, the F_{ST} ranging from 0.05 to 0.15 indicates moderate genetic differentiation between populations [36,37]. In this study, the F_{ST} of *C. camphora* between the four areas was 0.052–0.152 (Table 3), indicating a moderate genetic differentiation of *C. camphora* groups. AMOVA showed 92.03% of the genetic variation was within the population, leaving the genetic variation between regions and populations at only 7.97%. These results are consistent with previous studies on *C. camphora* [2,4] and other tree species in southern China [38,39]. The N_m value also suggested frequent gene exchange between the populations in the region, indicating that the gene exchange offsets the effect of genetic drift to some extent [40]. Considering that *C. camphora* is an insect-pollinated plant, the impact of insects on the long-distance transmission of its pollen may be relatively small. The seeds of *C. camphora* are taken by various birds in the south, so the gene flow between

populations is likely to be carried out by long-distance transport of seeds. Hence, the low level of genetic differentiation between populations can be explained by the long-distance transmission and transport of pollen or seeds [41]. In addition, considering the riparian habitat of *C. camphora*, its seed dispersal is easily affected by the regional water system. On the other hand, the insect-pollinated plants with dominant outcrossing provide a new isolation method for the reproductive system, reduce pollen flow between populations, and thus increase genetic differentiation and erosion among populations [4].

The phylogenetic tree showed our 171 *C. camphora* trees from 40 populations clustered into three major groups, indicating an apparent geographical distribution pattern. Southwest China is considered one of the hot spots of species diversity, with rich species diversity, and is also the origin and differentiation centre of many plants and glacial refuges [2,42]. The western regional populations of the Yunnan–Guizhou Plateau constituted subgroup I, with the highest genetic diversity, including YB and LZ in Sichuan and DZ in Chongqing. 137 *C. camphora* trees from the Japanese population and eastern China (excluding only HNCZ-761) constituted the largest subgroup (Figure 2, Cluster III), with the lowest genetic diversity. These regional populations in China are in Jiangxi, Anhui, and Hunan Provinces. Their border areas with neighbouring provinces have the most extensive distribution and the most extant ancient *C. camphora* populations in China. The *C. camphora* population in central China was clustered into a single subgroup II, with two exceptionals of the Hubei YC and Hunan SF populations into subgroup III, indicating a certain degree of gene exchange within eastern China. Similar results were obtained from PCA and population structure analysis. Mountains play a crucial role in the extinction and preservation of subtropical Chinese species and the generation of new species and new population structures. Due to the geographical isolation caused by mountains, gene exchange across mountains is hindered, resulting in significant differentiation of genetic structure and increased genetic diversity [43–46]. It was speculated that the Japanese *C. camphora* originated from China [47]. Moreover, according to our study, the *C. camphora* species formation was relatively late, approximately in the late Tertiary period (unpublished data). Multiple transgressions and regressions in the East China Sea in the fourth century affected the segregation and connection of eastern China, southern Japan, and the Korean Peninsula, which significantly impacted the genetic structure of plants [48–51]. This genetic exchange led to the close genetic relationship between the *C. camphora* populations in eastern China and Japan.

4.4. Conservation Strategies for *C. camphora*

In situ conservation is an effective measure to protect wild germplasm resources [52]. The native habitat of *C. camphora* is widely distributed throughout East Asia and China in the region with the most abundant wild resources. In recent decades in southern China, many wild *C. camphora* has been cut down to extract essential oil by destructive methods, such as digging trees and grubbing roots. Meanwhile, many ancient *C. camphora* (defined as diameter at breast height exceeding 100 cm) were transplanted into the city during the urban vegetation process. The ageing, physiological function decline, and habitat deterioration have also aggravated the weakening and death of some ancient *C. camphora* trees. Altogether, these factors caused the shrinking of existing wild *C. camphora* resources. Gene flow is a crucial factor affecting the genetic diversity and genetic differentiation of species [53]. The breeders' exchange of seedlings and seeds has significantly affected gene flow between *C. camphora* populations in China and Japan [2]. Transplantation of large and ancient trees should be avoided to maintain the genetic structure of *C. camphora* in each region. At the same time, as an insect-pollinated plant, factors such as genetic erosion and genetic drift may also affect the level of genetic diversity of the offspring of *C. camphora*. There are frequent gene introgressions between *C. camphora* and other species of the same genus [9], so ex-situ conservation is another important way to protect wild species. We should include as many individuals with different haplotypes and evolutionarily significant units as possible for ex-situ introduction. Attention should also be given to the mutual

introduction of various populations to increase their gene exchange and diversity, to resist the negative effects of genetic drift, thereby reducing the risk of genetic diversity reduction in wild populations.

5. Conclusions

This is the first study to analyze the genetic diversity and genetic structure of *C. camphora* in East Asia, including 39 populations in its whole distribution range in China and one Japanese population, based on the GBS technique. The results showed that the genetic diversity of the *C. camphora* populations in China were significantly lower than that of the Japanese population. The AMOVA, pairwise F_{ST} , and N_m analysis indicated that the genetic differentiation between populations was quite low, as the genetic variation was mainly within each population. Phylogenetic tree, PCA, and population structure analysis showed that the 40 populations were clustered into three major groups, i.e., western China, central China, and eastern China + Japan, showing a clear geographical distribution pattern. The phylogenetic analysis showed the closest relationship between the populations in eastern China and Japan. There is significant gene exchange between adjacent regions, and due to geographical isolation, the gene exchange between Chinese and Japanese populations is relatively weak. Because the Japanese population sample size in this study was small, further studies will be needed with increased sample size from the Japanese population, and populations from Taiwan and Vietnam to further confirm the findings of existing studies. In any case, this study's results will help further understand the current status of genetic diversity and genetic structure of *C. camphora* in East Asia and provide a theoretical basis for forming appropriate conservation measures and genetic improvement strategies.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/f12111511/s1>, Table S1: Location information of the 171 sampled *C. camphora* individuals; Table S2: Sequencing quality of the 171 sampled *C. camphora* individuals.

Author Contributions: Conceptualization, Y.Z.; methodology, X.G., Y.Z. and A.Y.; software, A.Y., X.G. and Z.W.; validation, C.C., Q.L. and H.L.; formal analysis, X.G.; investigation, Y.Z., A.Y. and F.Y.; resources, A.Y., Z.W. and Y.Z.; data curation, C.C.; writing—original draft preparation, X.G. and Y.Z.; writing—review and editing, X.G., A.Y. and Y.Z.; supervision, F.Y.; project administration, Y.Z.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (no. 31860079 and 32060354), and Important S & T Specific Projects of Jiangxi Provincial Department of Science and Technology (20203ABC28W016).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We are grateful to Enying Liu for her help in revising the manuscript and for her constructive advice.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Kameyama, Y. Development of microsatellite markers for *Cinnamomum camphora* (Lauraceae). *Am. J. Bot.* **2012**, *99*, e1–e3. [[CrossRef](#)]
2. Kameyama, Y.; Furumichi, J.; Li, J.; Tseng, Y.-H. Natural genetic differentiation and human-mediated gene flow: The spatiotemporal tendency observed in a long-lived *Cinnamomum camphora* (Lauraceae) tree. *Tree Genet. Genomes* **2017**, *13*, 38. [[CrossRef](#)]
3. Shi, S.; Wu, Q.; Su, J.; Li, C.; Zhao, X.; Xie, J.; Gui, S.; Su, Z.; Zeng, H. Composition analysis of volatile oils from flowers, leaves and branches of *Cinnamomum camphora* chvar. Borneol in china. *J. Essent. Oil Res.* **2013**, *25*, 395–401. [[CrossRef](#)]
4. Zhong, Y.; Yang, A.; Li, Z.; Zhang, H.; Liu, L.; Wu, Z.; Li, Y.; Liu, T.; Xu, M.; Yu, F. Genetic Diversity and Population Genetic Structure of *Cinnamomum camphora* in South China Revealed by EST-SSR Markers. *Forests* **2019**, *10*, 1019. [[CrossRef](#)]
5. Li, Z.; Zhong, Y.; Yu, F.; Xu, M. Novel SSR marker development and genetic diversity analysis of *Cinnamomum camphora* based on transcriptome sequencing. *Plant Genet. Resour.-Charact. Util.* **2018**, *16*, 568–571. [[CrossRef](#)]
6. Wu, X.; Wang, L.; Zhang, D.; Wen, Y. Microsatellite null alleles affected population genetic analyses: A case study of Maire yew (*Taxus chinensis* var. *mairei*). *J. For. Res.* **2019**, *24*, 230–234. [[CrossRef](#)]

7. Mammadov, J.; Aggarwal, R.; Buyyarapu, R.; Kumpatla, S. SNP markers and their impact on plant breeding. *Int. J. Plant Genom.* **2012**, *2012*, 728398. [[CrossRef](#)]
8. Yang, T.Y.; Gao, T.X.; Meng, W.; Jiang, Y.-L. Genome-wide population structure and genetic diversity of Japanese whiting (*Sillago japonica*) inferred from genotyping-by-sequencing (GBS): Implications for fisheries management. *Fish. Res.* **2020**, *225*, 105501. [[CrossRef](#)]
9. Wu, C.C.; Chang, S.H.; Tung, C.W.; Ho, C.K.; Gogorcena, Y.; Chu, F.H. Identification of hybridization and introgression between *Cinnamomum kanehirae* Hayata and *C. camphora* (L.) Presl using genotyping-by-sequencing. *Sci. Rep.* **2020**, *10*, 15995. [[CrossRef](#)]
10. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **2011**, *6*, e19379. [[CrossRef](#)]
11. Poland, J.A.; Rife, T.W. Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome* **2012**, *5*, 92–102. [[CrossRef](#)]
12. Lee, K.J.; Lee, J.R.; Sebastin, R.; Shin, M.J.; Kim, S.H.; Cho, G.T.; Hyun, D.Y. Genetic Diversity Assessed by Genotyping by Sequencing (GBS) in Watermelon Germplasm. *Genes* **2019**, *10*, 822. [[CrossRef](#)] [[PubMed](#)]
13. Lee, H.Y.; Jang, S.; Yu, C.R.; Kang, B.C.; Chin, J.H.; Song, K. Population Structure and Genetic Diversity of *Cucurbita moschata* Based on Genome-Wide High-Quality SNPs. *Plants* **2021**, *10*, 56. [[CrossRef](#)]
14. Solomon, A.M.; Han, K.; Lee, J.H.; Lee, H.Y.; Jang, S.; Kang, B.C. Genetic diversity and population structure of Ethiopian *Capsicum* germplasms. *PLoS ONE* **2019**, *14*, e0216886. [[CrossRef](#)] [[PubMed](#)]
15. Nuñez-Lillo, G.; Balladares, C.; Pavez, C.; Urrea, C.; Sanhueza, D.; Vendramin, E.; Dettori, M.T.; Arús, P.; Verde, I.; Blanco-Herrera, F.; et al. High-density genetic map and QTL analysis of soluble solid content, maturity date, and mealiness in peach using genotyping by sequencing. *Sci. Hortic.* **2019**, *257*, 108734. [[CrossRef](#)]
16. Zhang, L.; Guo, D.; Guo, L.; Guo, Q.; Wang, H.; Hou, X. Construction of a high-density genetic map and QTLs mapping with GBS from the interspecific F₁ population of *P. ostii* 'Fengdan Bai' and *P. suffruticosa* 'Xin Riyuejin'. *Sci. Hortic.* **2019**, *246*, 190–200. [[CrossRef](#)]
17. Carrasco, B.; Gonzalez, M.; Gebauer, M.; Garcia-Gonzalez, R.; Maldonado, J.; Silva, H. Construction of a highly saturated linkage map in Japanese plum (*Prunus salicina* L.) using GBS for SNP marker calling. *PLoS ONE* **2018**, *13*, e0208032. [[CrossRef](#)]
18. Catchen, J.M.; Amores, A.; Hohenlohe, P.; Cresko, W.; Postlethwait, J.H. Stacks: Building and genotyping Loci de novo from short-read sequences. *G3-Genes Genomes Genet.* **2011**, *1*, 171–182. [[CrossRef](#)]
19. Jo, H.; Koh, G. Faster single-end alignment generation utilizing multi-thread for BWA. *Bio-Med. Mater. Eng.* **2015**, *26*, S1791–S1796. [[CrossRef](#)]
20. Etherington, G.J.; Ramirez-Gonzalez, R.H.; MacLean, D. Bio-samtools 2: A package for analysis and visualization of sequence and alignment data with SAMtools in Ruby. *Bioinformatics* **2015**, *31*, 2565–2567. [[CrossRef](#)]
21. Levy, D.; Yoshida, R.; Pachter, L. Beyond pairwise distances: Neighbor-joining with phylogenetic diversity estimates. *Mol. Biol. Evol.* **2006**, *23*, 491–498. [[CrossRef](#)]
22. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [[CrossRef](#)] [[PubMed](#)]
23. Tang, H.; Peng, J.; Wang, P.; Risch, N.J. Estimation of Individual Admixture: Analytical and Study Design Considerations. *Genet. Epidemiol.* **2005**, *28*, 289–301. [[CrossRef](#)]
24. Excoffier, L.; Lischer, H.E. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **2010**, *10*, 564–567. [[CrossRef](#)]
25. Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics* **2012**, *28*, 2537–2539. [[CrossRef](#)] [[PubMed](#)]
26. Pootakham, W.; Jomchai, N.; Ruang-Areerate, P.; Shearman, J.R.; Sonthirod, C.; Sangrakru, D.; Tragoonrung, S.; Tangphatsornruang, S. Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics* **2015**, *105*, 288–295. [[CrossRef](#)]
27. Gurcan, K.; Teber, S.; Ercisli, S.; Yilmaz, K.U. Genotyping by Sequencing (GBS) in Apricots and Genetic Diversity Assessment with GBS-Derived Single-Nucleotide Polymorphisms (SNPs). *Biochem. Genet.* **2016**, *54*, 854–885. [[CrossRef](#)]
28. Li, M.; Chen, S.; Shi, S.; Zhang, Z.; Liao, W.; Wu, W.; Zhou, R.; Fan, Q. High genetic diversity and weak population structure of *Rhododendron jinggangshanicum*, a threatened endemic species in Mount Jinggangshan of China. *Biochem. Syst. Ecol.* **2015**, *58*, 178–186. [[CrossRef](#)]
29. Wang, X.; Zhao, W.; Li, L.; You, J.; Ni, B.; Chen, X. Clonal plasticity and diversity facilitates the adaptation of *Rhododendron aureum* Georgi to alpine environment. *PLoS ONE* **2018**, *13*, e0197089. [[CrossRef](#)]
30. Quintana, J.; Contreras, A.; Merino, I.; Vinuesa, A.; Orozco, G.; Ovalle, F.; Gomez, L. Genetic characterization of chestnut (*Castanea sativa* Mill.) orchards and traditional nut varieties in El Bierzo, a glacial refuge and major cultivation site in northwestern Spain. *Tree Genet. Genomes* **2014**, *11*, 826. [[CrossRef](#)]
31. Li, X.; Li, M.; Hou, L.; Zhang, Z.; Pang, X.; Li, Y. De Novo Transcriptome Assembly and Population Genetic Analyses for an Endangered Chinese Endemic *Acer miaotaiense* (Aceraceae). *Genes* **2018**, *9*, 378. [[CrossRef](#)]
32. Ferrer, M.M.; Eguiarte, L.E.; Montana, C. Genetic structure and outcrossing rates in *Flourensia cernua* (Asteraceae) growing at different densities in the South-western Chihuahuan Desert. *Ann. Bot.* **2004**, *94*, 419–426. [[CrossRef](#)]
33. Shen, Y.; Cheng, Y.; Li, K.; Li, H. Integrating Phylogeographic Analysis and Geospatial Methods to Infer Historical Dispersal Routes and Glacial Refugia of *Liriodendron chinense*. *Forests* **2019**, *10*, 565. [[CrossRef](#)]

34. Liu, F.; Hong, Z.; Xu, D.; Jia, H.; Zhang, N.; Liu, X.; Yang, Z.; Lu, M. Genetic Diversity of the Endangered *Dalbergia odorifera* Revealed by SSR Markers. *Forests* **2019**, *10*, 225. [[CrossRef](#)]
35. Kulshreshtha, V.P. Plant population Genetics, Breeding and Genetic Resources. *Sinauer Assoc.* **1990**, *10*, 43–63.
36. Wright, S. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **1965**, *19*, 395–420. [[CrossRef](#)]
37. Slatkin, M. Gene flow and the geographic structure of natural populations. *Science* **1987**, *236*, 787–792. [[CrossRef](#)]
38. Tan, J.; Zhao, Z.-G.; Guo, J.-J.; Wang, C.-S.; Zeng, J. Genetic Diversity and Population Genetic Structure of *Erythrophleum fordii* Oliv., an Endangered Rosewood Species in South China. *Forests* **2018**, *9*, 636. [[CrossRef](#)]
39. Hartvig, I.; So, T.; Changtragoon, S.; Tran, H.T.; Bouamanivong, S.; Theilade, I.; Kjær, E.D.; Nielsen, L.R. Population genetic structure of the endemic rosewoods *Dalbergia cochinchinensis* and *D. oliveri* at a regional scale reflects the Indochinese landscape and life-history traits. *Ecol. Evol.* **2017**, *8*, 530–545. [[CrossRef](#)] [[PubMed](#)]
40. Wright, S. Evolution in mendelian populations. *Genetics* **1931**, *16*, 97–159. [[CrossRef](#)] [[PubMed](#)]
41. Albaladejo, R.G.; Carrillo, L.F.; Aparicio, A.; Fernandez-Manjarres, J.F.; Gonzalez-Varo, J.P. Population genetic structure in *Myrtus communis* L. in a chronically fragmented landscape in the Mediterranean: Can gene flow counteract habitat perturbation? *Plant Biol.* **2009**, *11*, 442–453. [[CrossRef](#)]
42. Myers, N.; Mittermeier, R.A.; Mittermeier, C.G.; Fonseca, G.; Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **2000**, *403*, 853–858. [[CrossRef](#)]
43. Tian, S.; Kou, Y.; Zhang, Z.; Yuan, L.; Li, D.; Lopez-Pujol, J.; Fan, D.; Zhang, Z. Phylogeography of *Eomecon chionantha* in subtropical China: The dual roles of the Nanling Mountains as a glacial refugium and a dispersal corridor. *BMC Evol. Biol.* **2018**, *18*, 20. [[CrossRef](#)]
44. Sun, Y.; Hu, H.; Huang, H.; Vargas-Mendoza, C.F. Chloroplast diversity and population differentiation of *Castanopsis fargesii* (Fagaceae): A dominant tree species in evergreen broad-leaved forest of subtropical China. *Tree Genet. Genomes* **2014**, *10*, 1531–1539. [[CrossRef](#)]
45. 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. Phylogenetics Evol.* **2011**, *59*, 225–244. [[CrossRef](#)] [[PubMed](#)]
46. Zhang, Y.H.; Wang, I.J.; Comes, H.P.; Peng, H.; Qiu, Y.X. Contributions of historical and contemporary geographic and environmental factors to phylogeographic structure in a Tertiary relict species, *Emmenopterys henryi* (Rubiaceae). *Sci. Rep.* **2016**, *6*, 24041. [[CrossRef](#)] [[PubMed](#)]
47. Hammer, M.F.; Karafet, T.M.; Park, H.; Omoto, K.; Harihara, S.; Stoneking, M.; Horai, S. Dual origins of the Japanese: Common ground for hunter-gatherer and farmer Y chromosomes. *J. Hum. Genet.* **2006**, *51*, 47–58. [[CrossRef](#)] [[PubMed](#)]
48. Kimura, M. Paleogeography of the Ryukyu Islands. *Tropics* **2000**, *10*, 5–24. [[CrossRef](#)]
49. Haq, B.U.; Hardenbol, J.A.N.; Vail, P.R. Chronology of fluctuating sea levels since the triassic. *Science* **1987**, *235*, 1156–1167. [[CrossRef](#)]
50. Sakaguchi, S.; Qiu, Y.X.; Liu, Y.H.; Xin-Shuai, Q.I.; Kim, S.H.; Han, J.; Takeuchi, Y.; Worth, J.; Yamasaki, M.; Sakurai, S. Climate oscillation during the Quaternary associated with landscape heterogeneity promoted allopatric lineage divergence of a temperate tree *Kalopanax septemlobus* (Araliaceae) in East Asia. *Mol. Ecol.* **2012**, *21*, 3823–3838. [[CrossRef](#)]
51. Lambeck, K.; Chappell, J. Sea level change through the last glacial cycle. *Science* **2001**, *292*, 679–686. [[CrossRef](#)] [[PubMed](#)]
52. Shen, S.K.; Wang, Y.H.; Wang, B.Y.; Ma, H.Y.; Shen, G.Z.; Han, Z.W. Distribution, stand characteristics and habitat of a critically endangered plant *Euryodendron excelsum* H. T. Chang (Theaceae): Implications for conservation. *Plant Species Biol.* **2009**, *24*, 133–138. [[CrossRef](#)]
53. Setoguchi, H.; Mitsui, Y.; Ikeda, H.; Nomura, N.; Tamura, A. Genetic structure of the critically endangered plant *Tricyrtis ishiana* (Convallariaceae) in relict populations of Japan. *Conserv. Genet.* **2010**, *12*, 491–501. [[CrossRef](#)]