forests

MDPI

Article

Genetic Diversity and Population Structure of *Camellia drupifera* (Theaceae) and Its Related Species Evaluated by SSR Markers

Yufen Xu [1], Yanju Liu [1], Zhaoyan Yu [1], Kunlin Wu [1], Qiulin Cao [1], Han Gong [1], Yaodong Yang [1], Jianqiu Ye [2] and Xiaocheng Jia [1,*]

[1] Hainan Key Laboratory of Tropical Oil Crops Biology, Coconut Research Institute, Chinese Academy of Tropical Agricultural Sciences, Wenchang 571339, China; yufen_xu@163.com (Y.X.); yzs_liuyj@163.com (Y.L.); yuzy@catas.cn (Z.Y.); wkk1225@163.com (K.W.); 15908130211@163.com (Q.C.); yyang@catas.cn (Y.Y.)

[2] Institute of Scientific and Technical Information, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China; yejianqiu2006@126.com

* Correspondence: xcjia@catas.cn

Abstract: Oil–tea camellia has a long history of cultivation and utilization, with a history of more than 2000 years. In China, it is the main woody oil crop with high economic value and a national resource with unique characteristics. Concurrently, it is also known as one of the four major woody oil crops in the world. However, the genetic background of *Camellia drupifera* Lour. on Hainan Island in China is still unclear, and there is a great lack of systematic genetic characterization, which seriously hinders the development and utilization of oil–tea camellia germplasm resources and genetic improvement on Hainan Island. To analyze the genetic diversity and kinship between *C. drupifera* and its related species, this study utilized SSR molecular marker technology to genotype 160 individuals from 23 populations. A total of 137 alleles were amplified from the 14 polymorphic primers, with an average of 9.786. The average number of effective alleles and that of Shannon's information index for each locus were 1.865 and 0.633, respectively, suggesting that the screened SSR markers presented a moderately high level of polymorphism. Additionally, the mean observed heterozygosity (0.915) was greater than the mean expected heterozygosity (0.450), indicating an excess of heterozygotes in the tested population. The results of the principal component analysis (PCA), molecular variation analysis of variance (AMOVA) and population structure analysis were generally consistent; specifically, there was a high degree of individual heterozygosity within the population, and genetic variation occurred primarily among individuals within the population (90%) but rarely among groups (10%). Additionally, the UPGMA clustering divided the 160 germplasm resources into four major clades, and *C. drupifera* was principally grouped in two distinct branches; meanwhile, *Camellia gauchowensis* was also mainly clustered in these two clades. *Camellia oleifera* individuals were chiefly concentrated in other independent branches. It can be speculated that *C. drupifera* is genetically close to *C. gauchowensis*, but genetically distant from *C. oleifera*. This study can provide the scientific basis for the identification, collection, preservation, evaluation, and innovative utilization of oil–tea camellia.

Keywords: *Camellia drupifera*; theaceae; genetic diversity; population structure; SSR

Citation: Xu, Y.; Liu, Y.; Yu, Z.; Wu, K.; Cao, Q.; Gong, H.; Yang, Y.; Ye, J.; Jia, X. Genetic Diversity and Population Structure of *Camellia drupifera* (Theaceae) and Its Related Species Evaluated by SSR Markers. *Forests* **2024**, *15*, 2066. https://doi.org/10.3390/f15122066

Academic Editor: Reiner Finkeldey

Received: 18 October 2024
Revised: 18 November 2024
Accepted: 19 November 2024
Published: 22 November 2024

1. Introduction

Oil–tea camellia refers to the oil tree species in the genus *Camellia* L. from the Theaceae Mirb. family with high oil contents in their seeds and is a significant source of healthy and high-quality edible oil [1]. They are native to southern China and Southeast Asia. As one of the four major woody oil crops in China, it is not only the Chinese 'National Characteristic Germplasm Resource', but it is also known as one of the four main global woody edible oilseed plants together with oil palm, olive and coconut, as well as the worldwide meaningful woody oilseed resources [2]. Additionally, they have a stable

basic number of chromosomes (X = 15), but diverse ploidy levels, ranging from diploid (2n = 2X = 30) to decaploid (2n = 10X = 150), where some species are polyploid complexes with intraspecific ploidy variation. In China, oil–tea camellia has been cultivated and utilized for more than 2000 years, and is now widely cultivated from the Yangtze River Basin in China to South China. Currently, the camellia plant is also widely planted in other countries and regions, such as the Philippines, India, Brazil, and South Korea [3]. Camellia seed oil is rich in unsaturated fatty acids, about 90%, specifically, and monounsaturated fatty acids are predominant, the content of which can attain over 76% [2]. Its fatty acid composition and physicochemical properties are similar to those of olive oil, which enjoys the reputation of 'Oriental Olive Oil' with high nutritional and medicinal values [4], and a wide range of application prospects in the functional food, cosmetics, and pharmaceutical industries [5].

As recorded in Flora Reipublicae Popularis Sinicae, the genus *Camellia* owns about 280 species of plants distributed in East Asia on both sides of the Tropic of Cancer. Of these, China is the most widely spread with 238 species, and the remaining species are found in the Indo–China Peninsula and Japan. The cultivation and distribution areas of oil–tea camellia are vast in China, and the ecological conditions vary greatly. After a long period of natural selection and artificial breeding, extremely abundant germplasm resources have formed; specifically, approximately 50 species of seed oil crop [6], which is a prominent material basis for the germplasm innovation and industry improvement of oil–tea camellia, including exceedingly generous genetic variation. Currently, there are more than ten oil–tea taxa cultivated on a large scale in China, including *Camellia oleifera* Abel, *Camellia meiocarpa* Hu, *Camellia drupifera* Lour, *Camellia gauchowensis* Hung T. Chang, and *Camellia osmantha* Ye CX, Ma JL et Ye H [7]. Among them, the cultivated area of *C. oleifera* accounts for more than 90% of the total area in China. Unfortunately, the development of the oil–tea camellia industry in China still faces the problems of low utilization of germplasm resources and severe degradation of cultivars. Therefore, research on oil–tea camellia germplasm resources is urgently needed to explore the excellent germplasm and to breed improved varieties.

Hainan Province is located in the southernmost part of the oil–tea resource distribution in China with a tropical climate and is separated from mainland China by the Qiongzhou Strait. Moreover, oil–tea is generally a cross-pollinated crop with self-incompatibility in terms of its dual role in the environment and genetic control; Hainan Island oil–tea camellia resources with distinctive characteristics, outstanding advantages, and multifarious variety were consequently born [8–10]. According to the classification of the Flora of China, the local characteristic oil–tea camellia species in Hainan belongs to *C. drupifera*. This species is a special oil–tea camellia in tropical and southern subtropical regions and primarily distributes in Hainan, Guangxi, Guangdong, and other southern provinces in China, with enormous breeding potential [11]. Tea-seed oil extracted from the seeds of native *C. drupifera* in Hainan has the characteristics of a strong aroma, high consistency, and mellow taste [12]. Furthermore, its taste and quality are diverse from those of mainland China, and its price is more than three times greater than that of mainland camellia oil [13]. Overall, *C. drupifera* in Hainan provides an excellent genetic material basis suitable for planting in tropical areas to breed high-yield improved varieties with local characteristics, and to develop the indigenous oil–tea camellia industry for Hainan [14].

Carrying out research related to the genetic diversity of germplasm resources can help predict the genetic variation of economically important traits, and fully explore and utilize inter- or intrapopulation genetic resources. Moreover, resolving the origin history and evolutionary potential of species can be profitable and is greatly significant to the identification, protection, development and utilization of germplasm resources, and molecular-assisted breeding [15]. However, the genetic background of *C. drupifera* in Hainan is still ambiguous, and systematic research on its genetic characteristics is particularly lacking, which seriously hinders the development, utilization, and genetic improvement of oil–tea camellia germplasm resources on Hainan Island. To date, scholars have carried out

several studies on the genetic diversity of local *C. drupifera* in Hainan Province by applying morphological markers [11,16], cytological markers [17,18], biochemical markers [4,19], and molecular markers [20,21]. In practice, incomplete sample collection or limited reliable markers have led to inconsistent speculations and conclusions from different studies, often accompanied by controversies and discussions, and research progress is still relatively slow. Previous studies have shown that it may be a variety of *C. gauchowensis* or its hybrid offspring with other *Camellia* taxa [9]. Some researchers have also speculated that the oil–tea camellia in Hainan may be an ecotype of *C. gauchowensis* adapted to the Hainan environment [10]. On the other hand, some experts have proposed that the characteristic oil–tea camellia of Hainan Island is a new species [22]. Hence, research on the genetic diversity of *C. drupifera* and its relatives is necessary, which is conducive not only to the utilization and innovation of oil–tea camellia germplasm resources but also to its further collection and preservation.

Molecular markers are a direct reflection of genetic variation at the DNA level and have the advantages of strong stability and high polymorphism. They are irrespective not only of the limits from the stage of growth and development but also of environmental factors, which makes this technology an ideal tool for germplasm evaluation and genetic breeding [23]. Among them, SSR molecular markers are one of the most effective and widely exploited marker types, and are characterized by the characteristics of co-occurrence, extensive marker content, powerful genetic stability, and good reproducibility [24]. Moreover, these methods have been widely used in research on germplasm resource identification, genetic diversity analysis, kinship assessment, and molecularly assisted breeding in crops [25]. This molecular marker technology also has extensive application in some oil–tea camellia species, such as *C. oleifera* [26] and *Camellia chekiangoleosa* [27], but it has been rarely applied in *C. drupifera*. In this study, the SSR molecular markers were utilized to assess genetic diversity and resolve the genetic structure of representative germplasm resources of *C. drupifera* and its relatives, including *C. drupifera*, *C. gauchowensis*, and *C. oleifera* from the Guangdong, Guangxi, and Fujian Provinces in China. The main objectives of this study were as follows: (1) to select effective SSR markers for the identification and classification of *C. drupifera* in Hainan Island; (2) to evaluate the genetic diversity and genetic structure of the population from *C. drupifera* and its relatives; and (3) to investigate the genetic differentiation and relatedness of *C. drupifera*, *C. gauchowensis*, and *C. oleifera*. This study provides a scientific basis for the identification of Hainan oil–tea camellia germplasm resources and the excavation of excellent parental material. In addition, it can contribute to the history of introduction, cultivation, and molecular marker-assisted breeding.

2. Materials and Methods

2.1. Plant Material Collection

We collected 166 individuals from 23 populations of *C. drupifera* and its relatives (Figure 1 and Table S1) in the main distribution areas of China, including 100 germplasms from 12 *C. drupifera* populations, 15 accessions from two *C. gauchowensis* populations, 31 materials from six *C. oleifera* populations, and 10 samples from two *Camellia meiocarpa* populations. These materials were collected from the Hainan, Guangdong, Guangxi, and Fujian Provinces. Additionally, one outgroup was acquired in Guangxi (MH2), which was composed of ten individuals from nine other oil–tea camellia tree species. For each individual, four to six fresh leaves, free of pests and diseases, were sampled, dried in silica gel and preserved. However, six of the 166 samples were excluded from the genetic analysis due to missing massive loci information containing GZYN08, GZYN12, and GZYN13 from *C. gauchowensis* of Guangdong (GZ2), and LCYN04, LCYN05, and LCYN13 from *C. drupifera* of Guangxi (LC1).

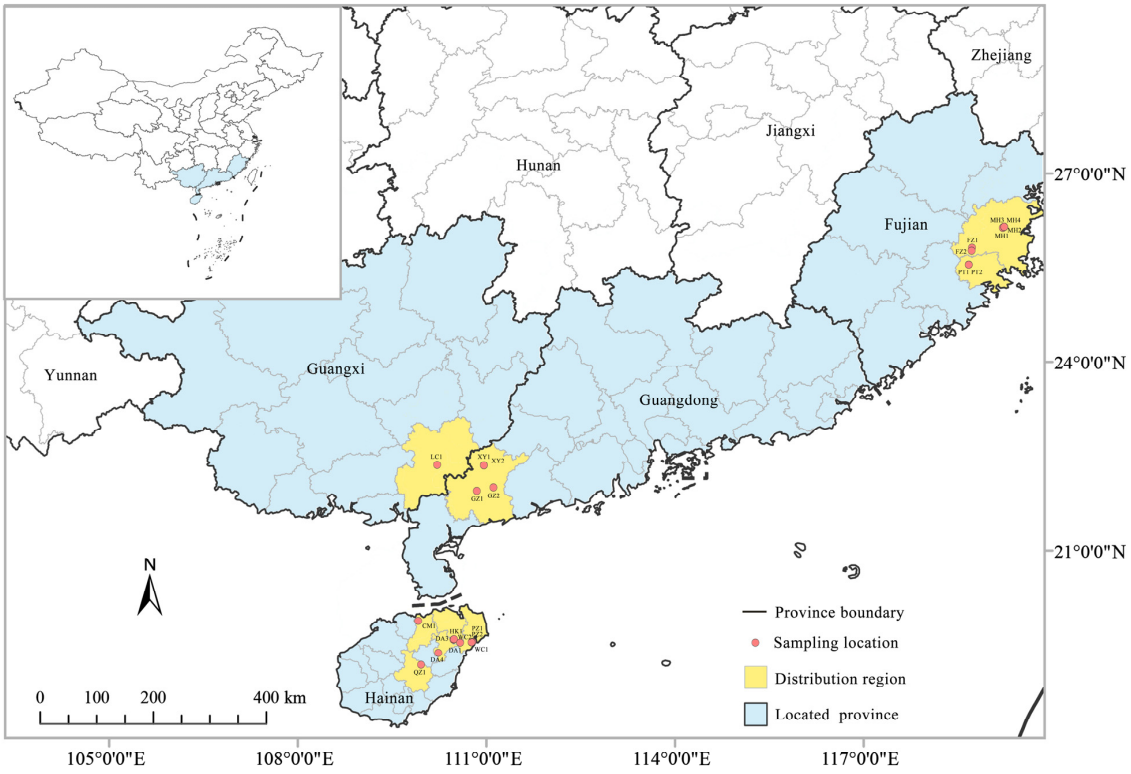

Figure 1. Population distribution of *Camellia drupifera* and its related species germplasms in this study. The map is produced based on the standard map with drawing approval number GS (2024) 0650 downloaded from the National Platform for Common GeoSpatial Information Services of the Ministry of Natural Resources of the People's Republic of China, with no modifications to the base map.

2.2. Genomic DNA Extraction and Detection

DNA samples of leaf tissues from 166 materials of *C. drupifera* and its relatives were extracted by the magnetic bead method plant tissue genomic DNA extraction kit produced by Wuhan Tianyi Huayu Gene Technology Co., Ltd. (Wuhan, China) A 2 µL DNA stock solution was added to 2 µL bromophenol blue for detection via a NanoDrop 8000 (Thermo Scientific, Waltham, MA, USA), and then 1% agarose gel electrophoresis was performed, and the samples were stored at 4 °C after the quality test.

2.3. Polymorphic SSR Primer Screening and Microsatellite Genotyping

In this study, 98 pairs of SSR primers (Table S2) developed by Liao [28], Wen et al. [29], Li et al. [30], and other scholars were used to screen out clear bands in the target segment, which had good repeatability. Moreover, primer screening experiments were carried out using 15 individuals randomly selected from different populations. Consequently, 14 pairs of target SSR markers selected with polymorphisms (Table 1) were applied for further population genotyping detection of 160 individuals in 23 populations.

Table 1. Characteristics of the 14 pairs of polymorphic SSR primer markers used for genetic diversity analysis.

Locus	Repeat Motif	Primer Sequence (5′–3′)	Amplified Fragments Size (bp)	Annealing Temperature (°C)
C191 [28]	(AAAT)6	F: CATCGACACAAATCCTAACAACA R: CCTTCCCTTCCTTATCCTTACAG	157	52
C46 [28]	(TTG)6	F: AATCGGATCTGAGGGTTGTCTAT R: TTGTAAATGCTTCAGAAATGCCT	160	52

Table 1. *Cont.*

Locus	Repeat Motif	Primer Sequence (5′–3′)	Amplified Fragments Size (bp)	Annealing Temperature (°C)
C96 [28]	(TG)11	F: ACAAAGAAACACAACCTCACGAT R: ACCCAAAAGATGAATTGTGCTAA	145	52
Ck89 [29]	(CCGATG)5	F: TGCCTTTGACCAACTCTA R: TTCCGACCTCCAACACTC	252	52
Co81 [30]	(GAT)n	F: GGTCAAAACGAAGAAGAAGATCA R: GGGATTCCCAATAGAGAGCC	146–161	52
CoA011 [31]	(CTT)5	F: TGGGTGGCTCAATATCATCA R: ACCGGCCATTTATATGGGTT	200	52
CoA032 [31]	(GCG)5	F: TTATTCTTCGGGAACAACGG R: ACACATGAAACAACGGCAAA	170	52
CoA038 [31]	(GTG)7	F: GAGATCGGCCAGAGTTTGAG R: CATCAAAGCCACACTCGCTA	202	52
CoA046 [31]	(TAAC)4	F: AACCAGAGGAACATCCAACG R: TATCCTTGCCGCTTTGAATC	196	52
CoA069 [31]	(TGC)6	F: CATGGCTTGGCTTCAATCTT R: CAATGTTCCCAAGCGATTCT	224	52
CoSSR68 [32]	(TGA)8	F: TTCAGGAGGGCTCGACGATAAT R: GTTGGGGATTCAGGGGCGATTT	234	52
CoUg3402 [33]	(GAT)8	F: ACTCTTGTGGGTGAATGTTG R: GCTGGTAGGTTGGTTATGTT	205	52
SJMCoa003 [34]	(CAA)7	F: ACGAAACATGTCGGACGTGA R: GGGAATGGACGAGACTTGGG	120	52
SJMCoa090 [34]	(TCA)9	F: ACAGAAGGCGTTTGAGTCAA R: GGCTTCTTCTTCGGAACCCA	165	52

Note: F, forward primer sequence; R, reverse primer sequence.

2.4. Fluorescence PCR Amplification

SSR fluorescent primers were synthesized by Wuhan Tianyi Huayu Gene Technology Co., Ltd. via the connector method [35]; specifically, the primers were synthesized by connecting a 21 bp FAM fluorescent connector sequence (5′-GAAGGTGACCAAGTTCATGCT-3′) at the 5′ end of the forward primer sequence. Polymorphic primer screening amplification and population typing amplification reactions were performed on a Veriti 384well PCR instrument (Applied Biosystems, Waltham, MA, USA). The PCR amplification program was set as follows: pre-denaturation at 95 °C for 5 min; denaturation at 95 °C for 30 s, gradient annealing from 62 to 52 °C for 30 s, and extension at 72 °C for 30 s, running 10 cycles (each cycle decreased by 1 °C); denaturation at 95 °C for 30 s, annealing at 52 °C for 30 s, and extension at 72 °C for 30 s, running 25 cycles; extension at 72 °C for 20 min. The final PCR product was stored at 4 °C. After the PCR, the amplification products were analyzed on an ABI3730xL fully automated sequencer (Applied Biosystems, Waltham, MA, USA) with GeneScan™ 500LIZ as the internal size standard.

2.5. Genotype Data Acquisition

The raw data in '*.fsa' format were transferred from the ABI 3730xL sequencer, classified and archived by the detection loci. The data were subsequently imported into GeneMarker V3.0.0 for fragment analysis, and the positions of the internal molecular weight markers in each lane were compared with the positions of the peaks in each sample for genotypic data reading, and Excel genotype raw data and PDF peak map files were exported according to the locus name. To conduct joint analysis on samples with diverse ploidy levels (4X, 6X, 8X, and 10X, etc.) in this study, the genotypic data were masked by the 0/1 assignment method [36]; i.e., the presence of amplified bands at the same migration position was recorded as 1, and the absence of amplified bands was recorded as 0. A binary data matrix of 0/1 was constructed, and the data served as data for the dominant markers for the genetic analyses.

2.6. Genetic Analysis

In order to assess the level of genetic diversity of each SSR locus in this study, several genetic indexes of SSR loci were calculated in GenAlEx v6.501 [37], including the number of alleles (Na), the number of effective alleles (Ne), Shannon's information index (I), observed heterozygosity (Ho), and expected heterozygosity (He).

To measure the degree of genetic variation among the 23 populations, Nei's genetic distance and genetic identity among the populations were evaluated via GenAlEx v6.501 software. Hotspot maps of Nei's genetic distance and genetic identity results were plotted using TBtools v2.124 [38]. Molecular variation analysis of variance (AMOVA) was performed to estimate the variation, divergence, and significance tests between and within populations. Moreover, principal component analysis (PCA) can assess the similarity or dissimilarity in the composition of sample populations, which reflects the variability between different individuals or clusters by visually comparing the straight-line distances between samples in the axes. The PCA was evaluated using the OmicStudio tools (https://www.omicstudio.cn/tool, accessed on 30 September 2024).

To explore the genetic linkages and differences among dissimilar germplasm resources, the population structure of 160 samples was analyzed by STRUCTURE v2.3.4 [39]. With the range of values of the number of groupings (K) set from 1 to 20, the length of the burn-in period was 10,000, the Markov Chain Monte Carlo (MCMC) iteration was set as 100,000 times, each K value was repeated 20 times, and other parameters were set by default. The ΔK values and four estimators (MedMed, MedMean, MaxMed, and MaxMean) were estimated in the online tool StructureSelector [40] (https://lmme.ac.cn/StructureSelector/index.html, accessed on 16 November 2024) based on the Evanno method [41] and the Puechmaille method [42]. The estimation results were used to determine the most probable appropriate size of the group, and the population structure results were integrated by Clumpak [43].

To identify the kinship of *C. drupifera* and its close relatives, this study converted the genotypes of 160 samples into a 0/1 matrix and applied the algorithm of Nei's method to assess the genetic distances between the accessions. The individual clustering tree was subsequently constructed by the unweighted arithmetic mean method (UPGMA). MEGA X [44] was utilized for systematic population clustering on the basis of Nei's genetic distance and the UPGMA method. Finally, the UPGMA genetic tree was visualized and decorated using iTOL v6.9.1 [45].

3. Results

3.1. Assessment of SSR Marker Diversity Levels

The 14 pairs of SSR primers screened were used to perform PCR amplification in 160 samples of *C. drupifera* and its related species, and a total of 137 alleles were detected (Table 2). The Na per locus ranged from 6 (C191 and C46) to 16 (Co81), with a mean of 9.786. The Ne per primer pair varied between 1.733 (Ck89) and 1.980 (C96), with an average of 1.865. The mean Shannon's information index per marker was 0.633, with Ck89 having the lowest value (0.542) and C96 having the highest value (0.688). The expected heterozygosity (range: 0.847–0.952) was lower than the observed heterozygosity (range: 0.383–0.495) sites in 14 pairs of SSR markers, which means that there was an excess of heterozygotes in 23 populations of *C. drupifera* and its relatives.

3.2. Population Genetic Evaluation

Genetic distance is a momentous index for quantifying the degree of genetic differentiation among biological individuals or populations, among which Nei's genetic distance is the most common calculation method. In this study, we compared the genetic distances of 23 populations of *C. drupifera* and its relatives (Figure 2, Tables S3 and S4), and observed that the distribution of genetic distances among the 23 populations was in the range of 0.025–0.151, with genetic concordance ranging between 0.860 and 0.976. As shown in Figure 2, the greatest genetic distance (0.151) was between the two *C. oleifera* populations

from distinct geographic positions (WC2 and PZ2) with the lowest degree of genetic identity (0.860), followed by the WC2 population and the *C. gauchowensis* population (GZ1), with a genetic distance of 0.145. However, the closest genetic distance (0.025) and the highest genetic identity (0.976) was between two *C. drupifera* populations in Hainan (CM1 and PZ1).

Table 2. Genetic diversity analysis of 14 pairs of SSR molecular markers.

Locus	Product Size (bp)	Na	Ne		I		Ho		He	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
C191	136–155	6	1.847	0.271	0.629	0.125	0.917	0.084	0.442	0.110
C46	152–170	6	1.968	0.047	0.685	0.013	0.914	0.083	0.492	0.013
C96	125–139	8	1.980	0.029	0.688	0.008	0.952	0.050	0.495	0.008
Ck89	243–288	13	1.733	0.376	0.542	0.252	0.925	0.083	0.383	0.185
Co81	134–186	16	1.748	0.273	0.587	0.166	0.847	0.235	0.409	0.123
CoA011	187–205	7	1.791	0.304	0.599	0.162	0.915	0.057	0.419	0.134
CoA032	161–179	7	1.855	0.205	0.643	0.079	0.931	0.066	0.453	0.073
CoA038	193–217	9	1.921	0.099	0.671	0.029	0.906	0.107	0.478	0.028
CoA046	188–216	8	1.762	0.313	0.573	0.220	0.881	0.129	0.405	0.157
CoA069	212–263	12	1.892	0.159	0.659	0.055	0.923	0.087	0.467	0.052
CoSSR68	221–266	14	1.883	0.257	0.631	0.176	0.936	0.081	0.452	0.127
CoUg3402	178–202	9	1.870	0.308	0.611	0.216	0.936	0.065	0.440	0.156
SJMCoa003	104–131	10	1.918	0.110	0.669	0.034	0.904	0.114	0.477	0.032
SJMCoa090	149–182	12	1.947	0.085	0.678	0.026	0.926	0.088	0.485	0.025
Mean	–	9.786	1.865	–	0.633	–	0.915	–	0.450	–

Note: Na, number of different alleles; Ne, number of effective alleles; I, Shannon's information index; Ho, observed heterozygosity; He, expected heterozygosity.

The PCA results (Figure 3A and Figure S1) of 160 samples from 23 populations revealed that two principal coordinate components (PCA1 and PCA2) contributed 38.89% and 4.17%, respectively, to explaining the variation in 160 germplasm resources. Different icons represent diverse species, and dissimilar colors represent distinct populations. The distribution among individuals within the 23 populations was more scattered, but there was no obvious differentiation between the populations. Except for a few individuals in groups XY1 and XY2 (XYYN03, XYYN04, and XYYN07), *C. drupifera* samples from different collection sites were almost clustered together, and there were some overlapping individuals. The *C. gauchowensis* samples were also divided into two parts scattered in groups formed by *C. drupifera* and *C. oleifera*. The MHPT01 of *C. oleifera* (MH3) was embedded in the *C. drupifera* group, and the remaining 30 resources were grouped into a group that was clearly distinguished from *C. drupifera*. However, massive individuals of *C. meiocarpa* were cross distributed with the *C. oleifera* group, and there was a small amount of overlap of individuals both within and between the two species. Moreover, the results of the AMOVA (Figure 3B and Table S5) showed that the genetic variation of *C. drupifera* and its relatives existed mainly between individuals within the populations, accounting for 90% of the total variation, whereas the interpopulation variation accounted for only 10%.

To reveal the genetic components of *C. drupifera* and its related germplasm resources, STRUCTURE divided 23 populations into taxa on the basis of a computational evaluation of the possible optimal number of taxa. The Evanno method revealed that, as K increased from 1 to 20, the ΔK value reached the highest peak at $K = 2$ (Figure S2A), and the Mean LnP(K) obtained the maximum value at $K = 5$. According to the Puechmaille alternative statistics, MedMed K, MedMean K, and MaxMed K presented the highest values when $K = 4$, and MaxMean K got the peak values when $K = 5$ (Figure S2B). Therefore, by combining the two methods, the 160 individuals were slipped into 3, 4, and 5 clusters for genetic structure mapping based on the Bayesian model, with diverse gene pool sources indicated by different colors (Figure 4). In the STRUCTURE population structure analysis, when the probability distribution is (Q value) ≥ 0.6 for a particular germplasm in a certain taxon, the germplasm is considered to possess a relatively pure lineage; otherwise, the origin of the germplasm is considered complex.

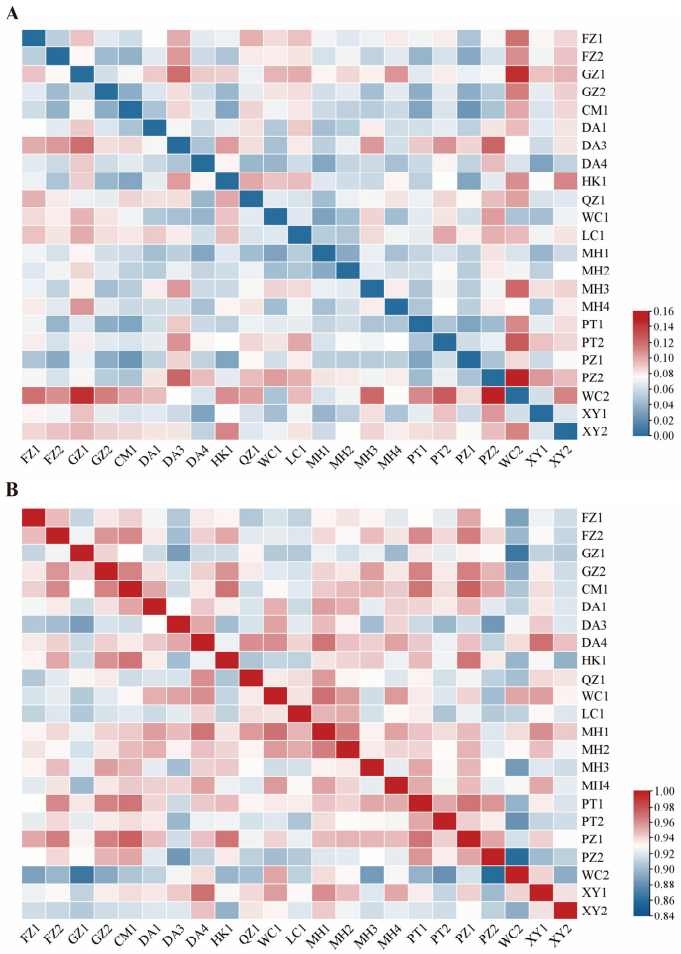

Figure 2. Genetic distance and genetic identity among 23 populations from *Camellia drupifera* and its related species. (**A**) Nei's genetic distance; (**B**) Nei's genetic identity.

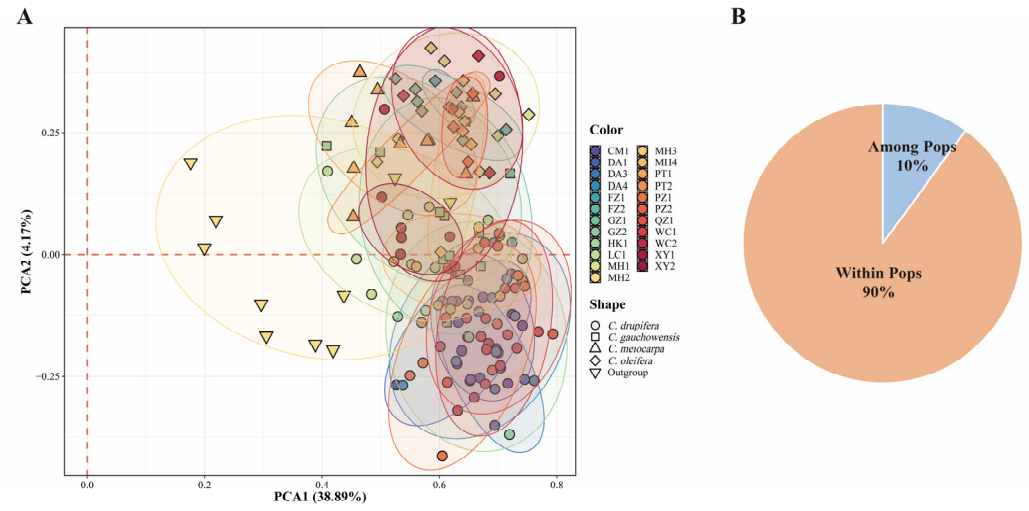

Figure 3. The PCA and AMOVA of 160 samples from 23 populations of *Camellia drupifera* and its related species. (**A**) Principal component analysis. The ellipses in the figure use the Khachiyan algorithm for the smallest possible ellipse enclosing all the points in each group. (**B**) Molecular variation analysis of variance.

3.3. Population Genetic Structure Analysis

When K = 2, for 97 *C. drupifera* germplasms, 76 individuals held Q values ≥ 0.6 in Cluster 1, and three germplasms (PZHK01, PZRY02, and XYYN03) all had a Q value greater than 0.6 in Cluster 2. However, the remaining 18 materials were distributed in Cluster 1 and Cluster 2, with Q values < 0.6. *C. gauchowensis* was assigned eight and one (GZYN03) individuals in Cluster 1 and Cluster 2, respectively, with a Q value > 0.6. The *C. oleifera* samples were located at seven and sixteen in Clusters 1 and 2, respectively, with a Q value ≥ 0.6.

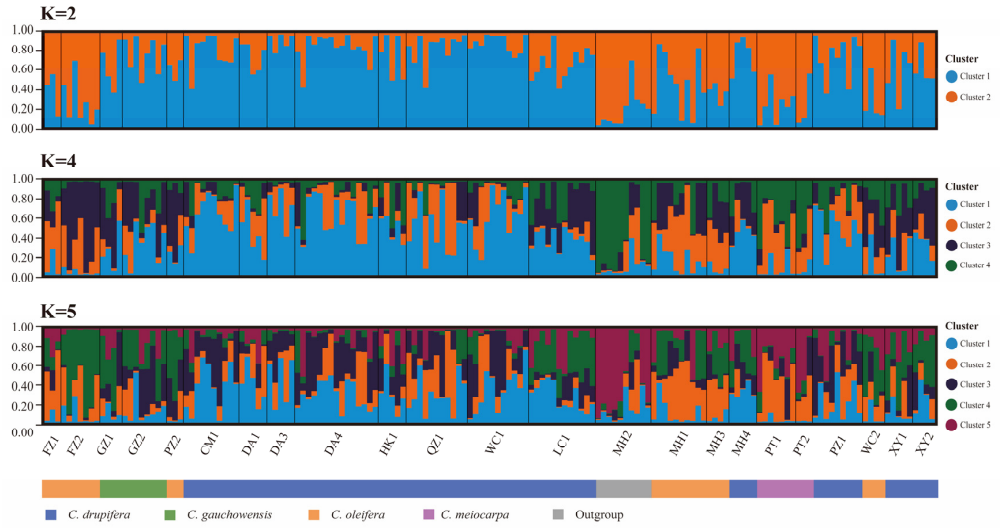

Figure 4. Genetic structure map of 23 populations from *Camellia drupifera* and its related species. The probability distribution is presented in the vertical axis. Each individual in this study is represented by a vertical bar, which is divided into several colors when individuals have mixed ancestral origins.

When K = 4, 31.96 % (31 individuals) of the *C. drupifera* germplasms had a Q value greater than 0.6 in Cluster 1, and the seven materials collected from the four Hainan populations (DA1, DA4, QZ1, and WC1) got Q values higher than 0.6 in Cluster 2. There were also samples LCYN15 and XYYN09 with Q values ≥ 0.6 in Cluster 3, and individual HNDD02 presented a Q value higher than 0.6 in Cluster 4, whereas 56 *C. drupifera* samples presented Q values < 0.6 in Clusters 1–4. For *C. gauchowensis*, sample GZYN03 had a Q value greater than 0.6 in Cluster 3, and the remaining 11 individuals were lower than 0.6 in Clusters 1–4. A total of 32.25% of the *C. oleifera* germplasms had a Q value higher than 0.6 in Clusters 1–4; specifically, three accessions (FZBD03, MHBD06, and MHBD07) were located in Cluster 2, and seven materials were assigned to Cluster 3 (FZPT02, FZPT03, FZPT05, FZPT06, HSPT01, WCPT01, and WCPT03).

When K = 5, for *C. drupifera*, 10.31% of its germplasm (10 individuals) had a Q value ≥ 0.6 in Cluster 1, 3.09% of it (HNQZ08, HNQZ09, and HNWC03) presented a Q value higher than 0.6 in Cluster 2, and 10.31% of its samples (10 individuals) were distributed in Cluster 3 with Q values ≥ 0.6. However, the Q value of 76.29% (74 individuals) of the *C. drupifera* materials was less than 0.6 in Clusters 1–5. There were only two accessions from *C. gauchowensis* with Q values ≥ 0.6 in Cluster 3 (GZYN11) and Cluster 4 (GZYN03), and the remaining 10 germplasms of *C. gauchowensis* were located in Clusters 1–5, with a Q value < 0.6. Nevertheless, 70.97% of the *C. oleifera* samples had a Q value lower than 0.6 in Clusters 1–5.

3.4. Genetic Relationship Identification

On the basis of Nei's genetic distance, individual clustering analysis and population clustering analysis of *C. drupifera* and its relatives were carried out via the UPGMA method. According to the individual clustering dendrogram (Figure 5), 160 accessions were classified

into four main branches (Clades I–IV), among which eight individuals in Clade I were from the outgroup (MH2). There were 34 individuals in Clade II, which mainly included 19 *C. drupifera* germplasms from the LC1, XY1, XY2, and MH4 populations in Guangxi, Guangdong, and Fujian, respectively, as well as *C. gauchowensis* (5), *C. meiocarpa* (7), *C. oleifera* (FZPT01 and MHBD01), and the outgroup (MHNR01). Clade III acquired 33 samples, which were dominated by *C. oleifera* (26), with a few *C. meiocarpa* (PTXG03, PTXG05, and PTZJ03), *C. gauchowensis* (GZYN01 and GZYN06), and *C. drupifera* (XYYN04) located as well. Additionally, 85 individuals composed Clade IV, of which *C. drupifera* accounted for 90.59% of the total individuals, including 71 samples from Hainan and six accessions collected in Guangxi, Guangdong and Fujian (LCYN01, MHYN02, MHYN03, XYYN05, XYYN08, and XYYN09). Clade IV also contained five *C. gauchowensis* samples from Guangdong (GZYN04, GZYN05, GZYN10, GZYN11, and GZYN15) and three *C. oleifera* accessions from Fujian (MHBD02 and MHPT01) and Hainan (HSPT01). In total, *C. drupifera* formed two relatively independent branches, and a small number of individuals from *C. gauchowensis*, *C. oleifera*, and *C. meiocarpa* were cross distributed within the two groups.

A cluster analyses of the genetic distances of the populations revealed (Figure 6) that 23 populations of *C. drupifera* and its relatives were clustered into four distinct branches (Clades I–IV). Among them, Clade I included two Guangdong populations (XY1 and XY2) of *C. drupifera*, two *C. oleifera* populations (FZ1 and FZ2), and GZ1 of *C. gauchowensis*. At the same time, Clade II clustered three *C. drupifera* populations from Hainan (QZ1 and WC1) and Guangxi (LC1) and gathered two Fujian populations (MH1 and MH3) of *C. oleifera* and the outgroup (MH2). However, Clade III had a more complex composition, which consisted of six populations from *C. drupifera*, *C. oleifera*, and *C. meiocarpa*. The distribution of Clade IV was dominated by *C. drupifera*, specifically five *C. drupifera* populations from Hainan (CM1, DA1, DA3, DA4, and HK1) and one *C. gauchowensis* population from Guangdong (GZ2). In general, most of the *C. drupifera* populations from Hainan formed a relatively independent branch and were more closely related to *C. gauchowensis*, whereas the *C. drupifera* populations from Guangxi, Guangdong, and Fujian were cross clustered on different branches with *C. oleifera* and *C. meiocarpa*.

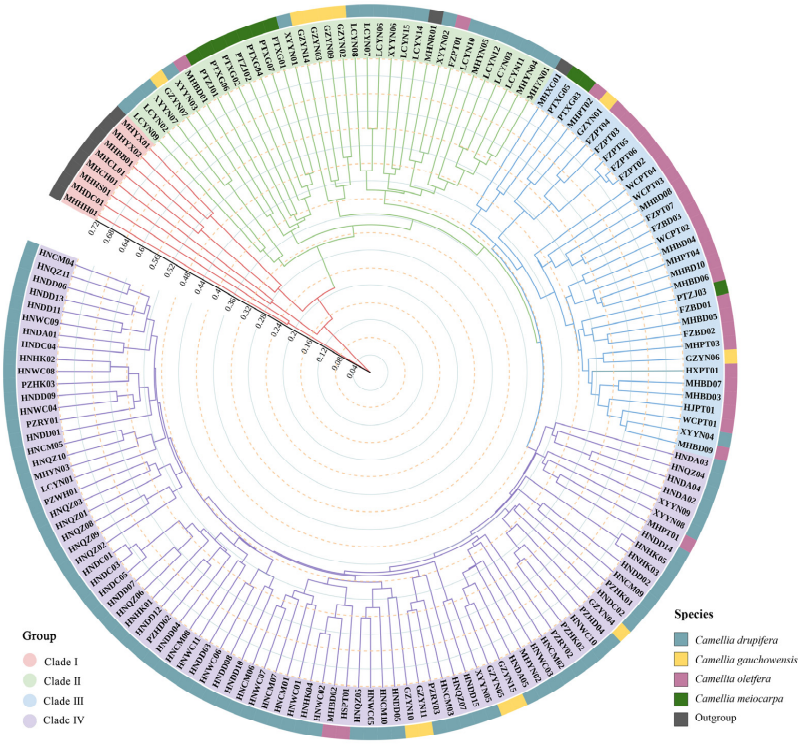

Figure 5. The UPGMA phylogenetic analysis of 160 samples from *Camellia drupifera* and its related species based on SSR data.

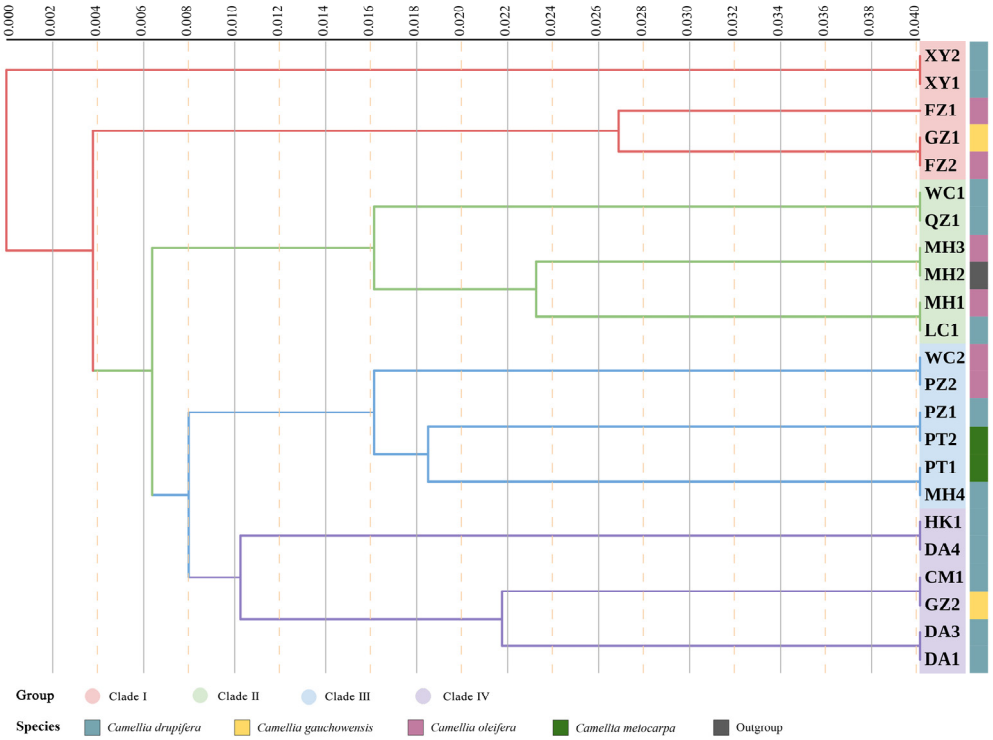

Figure 6. The UPGMA phylogenetic analysis of 23 populations from *Camellia drupifera* and its related species based on Nei's genetic distance.

4. Discussion

4.1. Development of Effective SSR Genotyping Markers for C. drupifera and Its Related Species

With the development and improvement of technology, DNA molecular markers have been rapidly developed and widely utilized in germplasm identification, genetic diversity assessment, the construction of molecular genetic maps, and genetic breeding [24]. Molecular marker technology has evolved from the first generation to the third generation, and various molecular marker techniques have been applied to study the genetic diversity of oil–tea camellia. Furthermore, dominant markers such as SRAP [46], ISSR [47], AFLP [48], and RAPD [49] have been used to detect genetic relationships among germplasms of diverse *Camellia* genera, but these dominant markers are unable to differentiate between allelic variations; in fact, they cannot accurately reflect genetic relationships among germplasms with different ploidy levels [50]. In contrast, codominant markers have greater advantages than dominant markers in revealing genetic diversity because they can detect multifarious variations in the chromosomal DNA of diploids or polyploids and identify homozygotes or heterozygotes. In recent years, codominant markers such as SSR [26], RFLP [51], and SNP [50] have also been exploited for many cases to evaluate genetic variation in populations of the *Camellia* genus, with SSR molecular markers having extensive applications.

In order to assess the population genetic diversity of *C. oleifera*, Xiao et al. [47] analyzed 135 samples using four ISSR primers and three SCoT primers, which amplified 51 and 49 bands, with 12.75 and 16.33 bands per primer, respectively. In addition, Yan et al. [50] screened and sequenced 20 pairs of SNP primers on 102 oil–tea camellia germplasms to generate 644 SNPs, and then explored the genetic relationships among germplasms and identified SNP loci related to oil content. Otherwise, in previous SSR analyses of genetic diversity in oil–tea camellia [26,31,33,52–57], the mean distributions of Na, Ne, I, Ho, and He at different loci ranged from 1.092 to 17.833, 1.168 to 8.999, 0.067 to 2.301, 0.225 to 0.965, and 0.602 to 0.850, respectively. These SSR loci have provided reliable genetic information for kinship identification, population genetic analysis, and the core germplasm screening of oil–tea camellia resources. Even the same SSR markers do not maintain the same

values of genetic diversity parameters across diverse research topics. Since these metrics depend largely on the genetic background of the experimental materials and the number of individuals, they can reflect their relative ability to assess genetic polymorphisms to a certain extent.

In this study, 14 pairs of polymorphic SSR primers were screened from 98 pairs of SSR makers in the *Camellia* genus developed by previous researchers, and a total of 137 alleles were amplified from 14 pairs of loci in 160 individuals of *C. drupifera* and its relatives (Table 2), which showed good generalizability among different oil–tea camellia species. Furthermore, the average values of Na, Ne, I, Ho, and He for each pair of primers were 9.786, 1.865, 0.633, 0.915 and 0.45, respectively, indicating a relatively high level of polymorphism. At the same time, by comparing the values of Ho and He, it is speculated that there is an excess of heterozygosity in the population as a whole. The occurrence of heterozygosity excess may also be that the population is small or subjected to artificial directional selection pressure, or it has suffered a certain degree of genetic drift or bottleneck effects [58]. In addition, the results of PCA and UPGMA further indicated that the 14 pairs of SSR loci could provide reliable genetic information for distinguishing the germplasm resources of *C. drupifera* and its relatives. Thus, the SSR molecular markers in this study presented medium to high polymorphism and are suitable tools for germplasm resource identification, genetic diversity analysis, and kinship detection in oil–tea camellia.

4.2. Population Structure of C. drupifera and Its Related Species

Genetic diversity, as a meaningful measure of the adaptability of a species, can be affected by a variety internal and external factors. Herein, internal factors include genetic drift (the bottleneck effect is a special case), natural selection, mutation, and gene flow, whereas external factors mainly refer to environmental changes and anthropogenic perturbations such as plate motion, human migration, and breeding activities [59]. In most cases, if the genetic identity between individuals or populations is relatively small and the genetic distances are relatively larger, they are more genetically differentiated from each other and thus more distantly related [60]. The genetic distances among the 23 populations obtained in this study were generally low (0.025–0.151), and the genetic identity was generally high (0.860–0.976) (Figure 2), which implied that there was a small degree of genetic differentiation among the populations of *C. drupifera* and its relatives, and that they are closely related. Previously, population genetic research has revealed that *C. drupifera* [20], *C. oleifera* [47], and *C. meiocarpa* [61] germplasms exhibited relatively little genetic variation. However, this narrower genetic base may be influenced by factors such as anthropogenic factors, animal migration, and variations in the number and type of germplasm resources and marker primers.

The comprehensive results of the PCA (Figure 3A and S1) and STRUCTURE analysis (Figure 4) revealed a certain degree of genetic differentiation between *C. drupifera* and *C. oleifera* with a relatively low degree of gene flow. Nevertheless, no significant genetic differentiation was observed between *C. drupifera* and *C. gauchowensis*, or between *C. oleifera* and *C. meiocarpa* with relatively frequent gene exchange. These findings suggest that the genetic relationships between *C. drupifera* and *C. gauchowensis*, and *C. oleifera* and *C. meiocarpa* are relatively intimate; moreover, it is relatively distant from that between *C. oleifera* and *C. drupifera*. This result is analogous to the findings obtained via phenotypic traits and molecular markers by Yang et al. [10] and Huang et al. [62].

In addition, all the population genetic structure maps of *C. drupifera*, *C. gauchowensis*, *C. oleifera*, and *C. meiocarpa* revealed that the same species did not have pure gene pools individually. There was also a certain degree of genetic mixing among different species. Moreover, the PCA results showed that there was no obvious genetic differentiation among the populations, but the degree of heterozygosity of individuals within the populations was greater. These results are in accordance with the AMOVA results, which revealed that only 10% of the genetic variation occurred between populations, whereas 90% of the variation occurred between individuals within populations. In summary, it is spec-

ulated that human activities, such as anthropogenically mediated plant migration and artificially selected breeding, may importantly contribute to a greater degree of genetic mixing between populations, thus reducing genetic differences between geographically distinct populations [63].

4.3. Genetic Relationships Among C. drupifera and Its Related Species

The UPGMA cluster analysis divided 160 individuals of *C. drupifera* and its relatives into four clades, and the taxon profile of the sample materials was generally consistent with the results of PCA and STRUCTURE (K = 4) as a whole. Among these three distinct clustering methods, no particularly obvious relationship was observed between the dissimilar geographic origins of the most participating germplasm resources. This phenomenon may be caused by two potential factors. On the one hand, the population structure analysis revealed gene exchange signals within the population and the presence of gene mixing within and between species. Concurrently, the genetic background of the germplasm resources is highly heterogeneous, which makes it difficult to subdivide. On the other hand, the number of molecular markers is small, and the genetic information of the detected locus may fail to cover all the variations, which agrees with the speculation reported by Song et al. [64].

According to the results of the individual UPGMA dendrogram (Figure 5), the *C. drupifera* samples were mainly clustered into two different branches (Clade II and Clade IV), and only a single sample was distributed in Clade III. Meanwhile, *C. gauchowensis* was also mainly grouped in these two branches of *C. drupifera*, with a few cross distributed in Clade III. However, almost all the *C. oleifera* individuals were located in the Clade III branch, and a few samples were embedded in the Clade II and Clade IV branches. These results suggest that *C. drupifera* has an intimate genetic relationship with *C. gauchowensis* but is more distantly related to *C. oleifera*. A key morphological index that distinguishes between *C. drupifera* and *C. gauchowensis* is whether the twigs of the current year are hairy or not in traditional taxonomy [65]. Qi et al. [66] reported that most branchlets in populations of *C. drupifera* and *C. gauchowensis* were hairy, and that only a few populations had single trees with glabrous twigs. In the 2007 edition of the Flora of China, *C. gauchowensis* was subsumed as a synonym for the species *C. drupifera*. Thus, the results of this study support the traditional classification of *C. drupifera* and *C. gauchowensis* by FOC and provide molecular evidence for it.

The kinship of native oil–tea camellia resources in Hainan has attracted extensive attention and discussion from many scholars. Yuan et al. [9] conducted a resource survey of Hainan's oil camellia and reported that many of the morphological characteristics of local special oil–tea camellia lie between *C. gauchowensis* and *C. oleifera*. It is considered to be either a variety of *C. gauchowensis* or the progeny of a cross between *C. gauchowensis* and other *Camellia* species. Yang et al. [10] found that the native oil camellia of Hainan is very similar to *C. gauchowensis* in terms of fruit traits. Xu et al. [22] used the chloroplast gene *matK* to carry out the phylogenetic analyses of *Camellia* species, and reported that Hainan's local oil–tea camellia was particularly close to *C. drupifera* and *C. gauchowensis*. Nevertheless, there was some dissimilarity in pollen morphology among these three taxa, and the morphology of Hainan's characteristic oil–tea camellia was most similar to that of *C. oleifera*, but its fruits were larger and had brown pericarp. Therefore, they supposed that Hainan's native oil–tea camellia should be a new species in the oil–tea camellia group of the genus *Camellia* (sect. *Oleifera*), named *Camellia hainanica* Zhao et Shi.

In the present study, individual cluster analyses (Figure 5) revealed that *C. drupifera* individuals collected in Guangxi, Guangdong and Fujian were mainly distributed in Clade II. However, all the *C. drupifera* individuals acquired from Hainan formed a relatively independent branch (Clade IV). In addition, STRUCTURE results (Figure 4) showed that the content of the genetic component from Cluster I in most *C. drupifera* populations from Hainan was generally greater than that in its populations from mainland China. These findings imply that *C. drupifera* in Hainan is distinguished from this species in mainland

China with a certain degree of genetic variation. This study speculated that *C. drupifera* in Hainan is a variety of *C. drupifera*, and that the long-term isolation of the strait and differences in climatic conditions resulted in genetic differentiation between the two, which is analogous with the inference of Qi et al. [66].

In reality, oil–tea camellia is a self-incompatibility and cross-pollination plant with widespread interspecific hybridization and polyploidy, and present high genetic heterozygosity and strong phenotypic polymorphisms, resulting in unclear species boundaries and extremely challenging kinship identification. Owing to the limited number of molecular markers applied in this study, there are several limitations in the type, number, and geographic origin of the test germplasm resources. It is difficult to determine whether the *C. drupifera* in Hainan belongs to a variety of *C. drupifera*. Alternatively, is it an independent new species? Therefore, extensive and efficient molecular markers should be developed in the future, and a comprehensive evaluation and identification of characteristic germplasm resources from native oil–tea camellia in Hainan should be carried out in conjunction with morphology, cytology, palynology and molecular methods to construct an accurate evaluation system. These findings could further promote the selection and breeding process of Hainan's oil–tea camellia varieties.

5. Conclusions

This study screened 14 pairs of SSR primers with medium to high polymorphism, which can be utilized as molecular markers for germplasm resource identification, genetic diversity analysis, kinship studies, and DNA fingerprinting of oil–tea camellia. The population structure of *C. drupifera*, *C. gauchowensis*, *C. oleifera*, and *C. meiocarpa* were detected, all showing complex genetic compositions. In addition, the results of genetic differentiation and cluster analysis implied that *C. drupifera* and *C. gauchowensis* were genetically familiar but were distantly related to *C. oleifera*. Furthermore, the clustering dendrogram of individuals preliminarily confirmed that *C. drupifera* from Hainan formed a relatively independent group, which led to the speculation that the local *C. drupifera* in Hainan might belong to a variety of *C. drupifera*. In conclusion, this study conducted a genetic evaluation of *C. drupifera* and its close relatives and explored the kinship of *C. drupifera* resources in Hainan. This study could provide a scientific basis for further clarifying the resource characteristics of native *C. drupifera* in Hainan, knowledge of the genetic diversity and genetic structure of oil–tea camellia resources, and innovative utilization and conservation of germplasm resources.

Supplementary Materials: The following materials are available online at https://www.mdpi.com/article/10.3390/f15122066/s1, Figure S1: The principal component analysis of *Camellia drupifera* and its related species; Figure S2: The value of the optimum K determined by two methods. A, B: The mean LnP(K) and Δ*K* values calculated by the Evanno method. The optimal K value (X-axis) for spurious clusters are indicated by red lines. C~D: The MedMed K, MedMean K, MaxMed K and MaxMean K values assessed by the Puechmaille method. The optimal K (Y-axis) after removing spurious clusters are indicated by red lines; Table S1: Detailed information for *Camellia drupifera* and its related species germplasms in this study; Table S2: Detailed information on 98 pairs of SSR primers used for polymorphism screening [67–71]; Table S3: Nei's genetic distance among 23 populations from *Camellia drupifera* and its related species; Table S4: Nei's genetic identity among 23 populations from *Camellia drupifera* and its related species; Table S5: An analysis of molecular variance among 160 samples of 23 populations within *Camellia drupifera* and its related species.

Author Contributions: Conceptualization, X.J.; Data Curation, Y.L.; Formal Analysis, Y.X. and K.W.; Funding Acquisition, Y.Y. and J.Y.; Methodology, X.J.; Resources, Y.X. and H.G.; Software, Y.X., Z.Y., K.W. and Q.C.; Supervision, Y.Y. and X.J.; Visualization, Z.Y., Q.C. and H.G.; Writing—Original Draft, Y.X.; Writing—Review and Editing, X.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key R&D Program of China (No. 2023YFD2200701), the Central Public-interest Scientific Institution Basal Research Fund (No. 1630152024011 and No.

1630152024002), the Science and Technology special fund of Hainan Province (No. ZDYF2023XDNY054), and the Hainan Provincial Natural Science Foundation of China (No. 321QN0939).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article and Supplementary Materials.

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

References

1. Liang, H.; Qi, H.; Wang, Y.; Sun, X.; Wang, C.; Xia, T.; Chen, J.; Ye, H.; Feng, X.; Xie, S.; et al. Comparative chloroplast genome analysis of *Camellia oleifera* and *C. meiocarpa*: Phylogenetic relationships, sequence variation and polymorphic markers. *Trop. Plants* **2024**, *3*, e023. [CrossRef]
2. Sun, J.; Huang, D.; Xia, S.; Zhang, Y.; Tao, J. Research progress of woody oil crops in China: A review. *Seed Biol.* **2024**, *3*, e008. [CrossRef]
3. Lin, P.; Wang, K.; Wang, Y.; Hu, Z.; Yan, C.; Huang, H.; Ma, X.; Cao, Y.; Long, W.; Liu, W.; et al. The genome of oil-Camellia and population genomics analysis provide insights into seed oil domestication. *Genome Biol.* **2022**, *23*, 14. [CrossRef] [PubMed]
4. Dou, X.; Wang, X.; Ma, F.; Yu, L.; Mao, J.; Jiang, J.; Zhang, L.; Li, P. Geographical origin identification of camellia oil based on fatty acid profiles combined with one-class classification. *Food Chem.* **2024**, *433*, 137306. [CrossRef] [PubMed]
5. Luan, F.; Zeng, J.; Yang, Y.; He, X.; Wang, B.; Gao, Y.; Zeng, N. Recent advances in *Camellia oleifera* Abel: A review of nutritional constituents, biofunctional properties, and potential industrial applications. *J. Funct. Foods* **2020**, *75*, 104242. [CrossRef]
6. Chen, J.; Guo, Y.; Hu, X.; Zhou, K. Comparison of the Chloroplast Genome Sequences of 13 Oil-Tea Camellia Samples and Identification of an Undetermined Oil-Tea Camellia Species From Hainan Province. *Front. Plant Sci.* **2021**, *12*, 798581. [CrossRef]
7. Yao, X.; Ren, H.; Ma, J.; Wang, K. *Oil-Tea Camellia Genetic Resource in China*; Science Press: Beijing, China, 2020.
8. Ye, Z.; Wu, Y.; Muhammad, Z.U.H.; Yan, W.; Yu, J.; Zhang, J.; Yao, G.; Hu, X. Complementary transcriptome and proteome profiling in the mature seeds of *Camellia oleifera* from Hainan Island. *PLoS ONE* **2020**, *15*, e0226888. [CrossRef]
9. Yuan, J.; Han, Z.; He, Y.; Huang, L.; Zhou, N. Investigation and Cluster Analysis of Main Morphological and Economical Characters for Oiltea Resource in Hainan Province. *J. Plant Genet. Resour.* **2014**, *15*, 1380–1384. [CrossRef]
10. Yang, L.; Zhang, Z.; Chen, J.; Chen, X.; Ji, Q.; Zheng, D. The quantitative characters and diversity of oiltea fruit in Hainan province. *Nonwood For. Res.* **2018**, *36*, 69–76. [CrossRef]
11. Guo, Y.; Wang, P.; Zhou, K.; Chen, J. Effects of producing regions on the cultivation traits differences of *Camellia vietnamensis* in Hainan Province. *J. For. Environ.* **2019**, *39*, 431–437. [CrossRef]
12. Xia, T.; Xiong, Z.; Wang, C.; Sun, X.; Chen, Y.; Chen, J.; Qi, H.; Liang, H.; Zhang, L.; Zheng, D. Comprehensive analysis of the effects of the traditional stir-fry process on the dynamic changes of volatile metabolites in Hainan camellia oil. *Food Chem. X* **2024**, *23*, 101747. [CrossRef] [PubMed]
13. Jia, X.; Liu, Y.; Xu, Y.; Yu, Z. Metabonomic Analysis of Mature Grains of *Camellia hainanica* and *Camellia oleifera*. *Mol. Plant Breed.* **2022**, *20*, 8255–8263. [CrossRef]
14. Zheng, D.; Pan, X.; Zhang, D.; Xi, L.; Zeng, J.; Zhang, Z.; Ye, H. Survey and Analysis on Tea-oil Camellia Resource in Hainan. *J. Northwest For. Univ.* **2016**, *31*, 130–135, 169. [CrossRef]
15. Wan, T. Genetic Diversity and Phylogeography of Wild Prunus tomentosa in China. Ph.D. Thesis, Northwest A&F University, Xianyang, China, 2023.
16. Jia, X.; Chen, L.; Yu, F.; Zhang, N. Preliminary Report on Genetic and Economic Traits of Tea-oil Camellia in Hainan Province. *Chin. J. Trop. Agric.* **2018**, *38*, 56–60.
17. Gong, Q. A Comparative Study on Resource Characteristics and Affiliation of Hainan Oil-Tea. Master's Thesis, Central South University of Forestry & Technology, Changsha, China, 2018.
18. Ye, T.; Yuan, D.; Li, Y.; Xiao, S.; Gong, S.; Zhang, J.; Li, S.; Luo, J. Ploidy Identification of *Camellia hainanica*. *Sci. Silvae Sin.* **2021**, *57*, 61–69. [CrossRef]
19. Xia, T.; Xiong, Z.; Sun, X.; Chen, J.; Wang, C.; Chen, Y.; Zheng, D. Metabolomic profiles and health-promoting functions of *Camellia drupifera* mature-seeds were revealed relate to their geographical origins using comparative metabolomic analysis and network pharmacology approach. *Food Chem.* **2023**, *426*, 136619. [CrossRef]
20. Qi, H.; Sun, X.; Wang, C.; Chen, X.; Yan, W.; Chen, J.; Xia, T.; Ye, H.; Yu, J.; Dai, J.; et al. Geographic isolation causes low genetic diversity and significant pedigree differentiation in populations of *Camellia drupifera*, a woody oil plant native to China. *Ind. Crop. Prod.* **2023**, *192*, 116026. [CrossRef]
21. Chen, X.; Yun, Y.; Wu, Y.; Qi, H.; Yang, L.; Chen, J.; Zheng, D. Genetic Diversity Analysis of *Camellia oleifera* Resources Based on SRAP Markers in Hainan Island. *J. Trop. Subtrop. Bot.* **2019**, *27*, 659–668.
22. Xu, Z.; Yuan, D.; Tang, Y.; Wu, L.; Zhao, Y. *Camellia hainanica* (Theaceae) a new species from Hainan, supported from morphological characters and phylogenetic analysis. *Pak. J. Bot.* **2020**, *52*, 1025–1032. [CrossRef]

23. Rani, R.; Raza, G.; Tung, M.H.; Rizwan, M.; Ashfaq, H.; Shimelis, H.; Razzaq, M.K.; Arif, M. Genetic diversity and population structure analysis in cultivated soybean (*Glycine max* [L.] Merr.) Using SSR and EST-SSR markers. *PLoS ONE* **2023**, *18*, e0286099. [CrossRef]
24. Yang, W.; Bai, Z.; Wang, F.; Zou, M.; Wang, X.; Xie, J.; Zhang, F. Analysis of the genetic diversity and population structure of *Monochasma savatieri* Franch. Ex Maxim using novel EST-SSR markers. *BMC Genom.* **2022**, *23*, 597. [CrossRef] [PubMed]
25. Kapoor, M.; Mawal, P.; Sharma, V.; Gupta, R.C. Analysis of genetic diversity and population structure in *Asparagus* species using SSR markers. *J. Genet. Eng. Biotechnol.* **2020**, *18*, 50. [CrossRef] [PubMed]
26. Cheng, L.; Cao, B.; Xie, S.; Yang, L.; Zhang, Y.; Cheng, Q.; Liu, J.; Yu, S.; Hu, D. Genetic diversity of wild *Camellia oleifera* in northern China revealed by simple sequence repeat markers. *Genet. Resour. Crop Evol.* **2024**, *71*, 2657–2672. [CrossRef]
27. Dong, L.; Tian, Q.; Huang, B.; Xu, L.; Wen, Q. Association Analysis of Economic Traits with SSR Markers in *Camellia chekiangoleosa*. *Mol. Plant Breed.* **2022**, *20*, 4710–4722. [CrossRef]
28. Liao, C. Development of EST-SSR Markers and Genetic Diversity Analysis of Tea-Oil Camellia in Southern Jiangxi. Master's Thesis, Guangzhou University, Guangzhou, China, 2023.
29. Wen, Q.; Xu, L.; Gu, Y.; Huang, M.; Xu, L. Development of polymorphic microsatellite markers in *Camellia chekiangoleosa* (Theaceae) using 454-ESTs. *Am. J. Bot.* **2012**, *99*, e203–e205. [CrossRef]
30. Li, H.; Ding, H.; Chen, Y.; Xu, L.; Li, N.; Chuanjiu, H. Identification of 12 Superior Cultivars of *Camellia oleifera* by Using Simple Sequence Repeat Feature Indexes. *J. Chin. Cereals Oils Assoc.* **2017**, *32*, 171–178. [CrossRef]
31. Yan, H.; Qi, H.; Li, Y.; Wu, Y.; Wang, Y.; Chen, J.; Yu, J. Assessment of the Genetic Relationship and Population Structure in Oil-Tea Camellia Species Using Simple Sequence Repeat (SSR) Markers. *Genes* **2022**, *13*, 2162. [CrossRef]
32. Dong, B.; Deng, Z.; Liu, W.; Rehman, F.; Yang, T.; Huang, Y.; Gong, H. Development of expressed sequence tag simple sequence repeat (EST-SSR) markers and genetic resource analysis of tea oil plants (*Camellia* spp.). *Conserv. Genet. Resour.* **2022**, *14*, 41–45. [CrossRef]
33. Jia, B.G.; Lin, Q.; Feng, Y.Z.; Hu, X.Y.; Tan, X.F.; Shao, F.G.; Zhang, L. Development and cross-species transferability of unigene-derived microsatellite markers in an edible oil woody plant, *Camellia oleifera* (Theaceae). *Genet. Mol. Res.* **2015**, *14*, 6906–6916. [CrossRef]
34. Song, J.; Li, X.; Zhang, S.; Lai, H.; Hu, X.; Zhou, Y.; Chen, M.; Wang, J.; Pang, Z. Development and Evaluation of SSR Markers Based on Transcriptome Sequencing of *Camellia* in Hainan Province. *Mol. Plant Breed.* **2022**, *20*, 6791–6801. [CrossRef]
35. Schuelke, M. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* **2000**, *18*, 233–234. [CrossRef] [PubMed]
36. Christelová, P.; De Langhe, E.; Hřibová, E.; ížková, J.; Sardos, J.; Hušáková, M.; Van den Houwe, I.; Sutanto, A.; Kepler, A.K.; Swennen, R.; et al. Molecular and cytological characterization of the global Musa germplasm collection provides insights into the treasure of banana diversity. *Biodivers. Conserv.* **2017**, *26*, 801–824. [CrossRef]
37. 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]
38. Chen, C.; Wu, Y.; Li, J.; Wang, X.; Zeng, Z.; Xu, J.; Liu, Y.; Feng, J.; Chen, H.; He, Y.; et al. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol. Plant* **2023**, *16*, 1733–1742. [CrossRef] [PubMed]
39. Pritchard, J.K.; Stephens, M.; Rosenberg, N.A.; Donnelly, P. Association Mapping in Structured Populations. *Am. J. Hum. Genet.* **2000**, *67*, 170–181. [CrossRef]
40. Li, Y.L.; Liu, J.X. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol. Ecol. Resour.* **2018**, *18*, 176–177. [CrossRef]
41. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef]
42. Puechmaille, S.J. Program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Mol. Ecol. Resour.* **2016**, *16*, 608–627. [CrossRef]
43. Kopelman, N.M.; Mayzel, J.; Jakobsson, M.; Rosenberg, N.A.; Mayrose, I. Clumpak: A program for identifying clustering modes and packaging population structure inferences across *K*. *Mol. Ecol. Resour.* **2015**, *15*, 1179–1191. [CrossRef]
44. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
45. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v6: Recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **2024**, *52*, W78–W82. [CrossRef] [PubMed]
46. Lin, P.; Yao, X.H.; Wang, K.L.; Zheng, T.T.; Teng, J.H. Identification and Genetic Analysis of *Camellia oleifera* Changlin Series Superior Clones by SRAP Molecular Marker. *J. Agric. Biotechnol.* **2010**, *18*, 272–279.
47. Xiao, P.; Liu, H.; Wang, D.; Tang, W.; Yang, H.; Wang, C.; He, Z.; Wang, R.; Wang, X.; Lu, X.; et al. Assessment of genetic diversity in *Camellia oleifera* Abel. Accessions using inter-simple sequence repeat (ISSR) and start codon targeted (SCoT) polymorphic markers. *Genet. Resour. Crop Evol.* **2020**, *67*, 1115–1124. [CrossRef]
48. Xie, Y.; Yao, X.; Li, Z.; Huang, Y. Analysis of Genetic Difference and Relationship of *Camellia meiocarpa* Native Varieties by Morphology and AFLP Markers. *Forest Res.* **2014**, *27*, 201–207.
49. Kao, A.; Wang, S.; Wang, Y.; Xu, Z.; Zhou, Y.; Ding, C. Analyses of genetic diversity among 65 wild *Camellia oleifera* based on ISSR and RAPD. *Guihaia* **2014**, *34*, 419–425.

50. Yan, R.; Ruan, C.; Zhao, S.; Ding, J.; Du, W.; Wang, H.; Han, P. SNP discovery of *Camellia oleifera* based on RNA-seq and its application for identification of genetic relationships and locus for oil content among different cultivars. *J. Hortic. Sci. Biotechnol.* **2020**, *95*, 687–702. [CrossRef]
51. Devarumath, R.M.; Nandy, S.; Rani, V.; Marimuthu, S.; Muraleedharan, N.; Raina, S.N. RAPD, ISSR and RFLP fingerprints as useful markers to evaluate genetic integrity of micropropagated plants of three diploid and triploid elite tea clones representing *Camellia sinensis* (China type) and *C. assamica* ssp. Assamica (Assam-India type). *Plant Cell Rep.* **2002**, *21*, 166–173. [CrossRef]
52. Tian, Q.; Huang, B.; Huang, J.; Wang, B.; Dong, L.; Yin, X.; Gong, C.; Wen, Q. Microsatellite analysis and polymorphic marker development based on the full-length transcriptome of *Camellia chekiangoleosa*. *Sci. Rep.* **2022**, *12*, 18906. [CrossRef]
53. Wang, P.; Su, J.; Wu, H.; Chen, Y.; Xie, Y.; Wang, H.; He, G.; Chen, N.; Wei, C.; Yang, L.; et al. Analysis of germplasm genetic diversity and construction of a core collection in *Camellia oleifera* C. Abel by integrating novel simple sequence repeat markers. *Genet. Resour. Crop Evol.* **2023**, *70*, 1517–1530. [CrossRef]
54. Wu, S.; Ye, H.; Chen, Y.; Deng, J.; Su, J.; Xie, Y.; Xie, Q.; Zhang, Z.; Qin, Z.; Xiao, Y.; et al. Characterization and cross-species transferability of a novel set of microsatellites derived from root transcriptomes of *Camellia oleifera*. *Plant Genet. Resour. Charact. Util.* **2019**, *17*, 371–374. [CrossRef]
55. Jia, B.; Lin, Q.; Zhang, L.; Tan, X.; Lei, X.; Hu, X.; Shao, F. Development of 15 genic-SSR markers in oil-tea tree (*Camellia oleifera*) based on transcriptome sequencing. *Genetika* **2014**, *46*, 789–797. [CrossRef]
56. Yin, X.; Li, T.; Tian, Q.Q.; Dong, L.; Xu, L.A.; Wen, Q. Development of Novel Polymorphic Microsatellite Markers and their Application for Closely Related *Camellia* (Theaceae) Species. *Russ. J. Genet.* **2022**, *58*, 404–412. [CrossRef]
57. Zhu, Y.; Liang, D.; Song, Z.; Tan, Y.; Guo, X.; Wang, D. Genetic Diversity Analysis and Core Germplasm Collection Construction of *Camellia oleifera* Based on Fruit Phenotype and SSR Data. *Genes* **2022**, *13*, 2351. [CrossRef] [PubMed]
58. Balloux, F. Heterozygote excess in small populations and the heterozygote-excess effective population size. *Evol. Int. J. Org. Evol.* **2004**, *58*, 1891–1900. [CrossRef]
59. Zhou, Q.; Mu, K.; Ni, Z.; Liu, X.; Li, Y.; Xu, L. Analysis of genetic diversity of ancient Ginkgo populations using SSR markers. *Ind. Crop Prod.* **2020**, *145*, 111942. [CrossRef]
60. Zou, Y.; Ge, X.; Yan, C.; Zhong, Q.; Chen, D.; Chen, Z.; Yuan, Y.; Guo, H.; Zhou, Y.; Wang, J.; et al. Assessment of genetic diversity of *Camellia yuhsienensis* based on leaf structure and inter simple sequence repeat (ISSR) markers. *Genet. Resour. Crop Evol.* **2024**, *71*, 4749–4762. [CrossRef]
61. Yang, Y.; Hong, Y.; Huang, Y.; Yao, X.; Wang, K. Analysis of Genetic Diversity for *Camellia meiocarpa* Hu. Population in Southwest Area of China by SRAP. *Hunan Agric. Sci.* **2011**, *13*, 1–4. [CrossRef]
62. Huang, Y. Study on introgressive hybridization of sympatric populations of *Camellia meiocarpa* and *Camellia oleifera*. *Chin. J. Appl. Ecol.* **2013**, *24*, 2345–2352.
63. Ding, Y. Genetic Diversity Analysis and Molecular ID Card Construction of Ancient Chestnut Trees and Varieties (Lines) in Yanshan. Master's Thesis, Hebei Normal University of Science & Technology, Qinhuangdao, China, 2023.
64. Song, L.; Wang, X.; Liu, W.; Yang, H.; Fu, J.; Hu, X.; Gao, Y.; Li, J. Genetic Diversity Analysis of 504 Tomato Germplasm Resources Based on SNP Markers. *J. Nucl. Agric. Sci.* **2022**, *36*, 2366–2373. [CrossRef]
65. Qi, H. Molecular Identification and Genetic Diversity Evaluation of Tea-Oil *Camellia* resources of Hainan Island. Master's Thesis, Hainan University, Haikou, China, 2020.
66. Qi, H.; Sun, X.; Yan, W.; Ye, H.; Chen, J.; Yu, J.; Jun, D.; Wang, C.; Xia, T.; Chen, X.; et al. Genetic relationships and low diversity among the tea-oil *Camellia* species in Sect. *Oleifera*, a bulk woody oil crop in China. *Front. Plant Sci.* **2020**, *13*, 996731. [CrossRef]
67. Shi, J.; Dai, X.; Chen, Y.; Chen, J.; Shi, J.; Yin, T. Discovery and experimental analysis of microsatellites in an oil woody plant *Camellia chekiangoleosa*. *Plant Syst. Evol.* **2013**, *299*, 1387–1393. [CrossRef]
68. Sharma, R.K.; Bhardwaj, P.; Negi, R.; Mohapatra, T.; Ahuja, P.S. Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.). *BMC Plant Biol.* **2009**, *9*, 53. [CrossRef] [PubMed]
69. Huang, X. Genetie structure of hexaploid wild *Camellia oleifera* in mount Jinggang and Lu based on microsatellite markers. Master's Thesis, Nanchang University, Nanchang, China, 2016.
70. Dai, H.; Shen, T.; Shi, T.; Li, R. Genomic SSR Loci Mining and Genetic Diversity Analysis of *Camellia oleifera* Based on Genome Sequences. *Crops* **2024**, *40*, 23–31. [CrossRef]
71. Yan, R.; Ruan, C.; Du, W.; Ding, J.; Wu, B.; Liu, L. Development of SSR Markers for Target-Genes Derived from *Camellia oleifera* by RNA-seq Technology. *Mol. Plant Breed.* **2018**, *16*, 2540–2548. [CrossRef]