

## Article

# Genetic Diversity and Population Structure of Chinese Chestnut (*Castanea mollissima* Blume) Cultivars Revealed by GBS Resequencing

Xibing Jiang <sup>1,2</sup> , Zhou Fang <sup>1</sup>, Junsheng Lai <sup>3</sup>, Qiang Wu <sup>3</sup>, Jian Wu <sup>3</sup>, Bangchu Gong <sup>1,\*</sup> and Yanpeng Wang <sup>1,\*</sup> <sup>1</sup> Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, China<sup>2</sup> State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing 100091, China<sup>3</sup> Qingyuan Bureau of Natural Resources and Planning, Lishui 323800, China

\* Correspondence: gongbc@126.com (B.G.); yanpengwang@caf.ac.cn (Y.W.); Tel.: +86-571-63131182 (B.G.); Fax: +86-571-6334130 (B.G.)

**Abstract:** Chinese chestnut (*Castanea mollissima* Bl.) is one of the earliest domesticated and cultivated fruit trees, and it is widely distributed in China. Because of the high quality of its nuts and its high resistance to abiotic and biotic stresses, Chinese chestnut could be used to improve edible chestnut varieties worldwide. However, the unclear domestication history and highly complex genetic background of Chinese chestnut have prevented the efficiency of breeding efforts. To explore the genetic diversity and structure of Chinese chestnut populations and generate new insights that could aid chestnut breeding, heterozygosity statistics, molecular variance analysis, ADMIXTURE analysis, principal component analysis, and phylogenetic analysis were conducted to analyze single nucleotide polymorphism data from 185 Chinese chestnut landraces from five geographical regions in China via genotyping by sequencing. Results showed that the genetic diversity level of the five populations from different regions was relatively high, with an observed heterozygosity of 0.2796–0.3427. The genetic diversity level of the population in the mid-western regions was the highest, while the population north of the Yellow River was the lowest. Molecular variance analysis showed that the variation among different populations was only 2.07%, while the intra-group variation reached 97.93%. The Chinese chestnut samples could be divided into two groups: a northern and southern population, separated by the Yellow River; however, some samples from the southern population were genetically closer to samples from the northern population. We speculate that this might be related to the migration of humans during the Han dynasty due to the frequent wars that took place during this period, which might have led to the introduction of chestnut to southern regions. Some samples from Shandong Province and Beijing City were outliers that did not cluster with their respective groups, and this might be caused by the special geographical, political, and economic significance of these two regions. The findings of our study showed the complex genetic relationships among Chinese chestnut landraces and the high genetic diversity of these resources.

**Keywords:** Chinese chestnut; SNP-based genetic diversity; population structure; geographical division

**Citation:** Jiang, X.; Fang, Z.; Lai, J.; Wu, Q.; Wu, J.; Gong, B.; Wang, Y. Genetic Diversity and Population Structure of Chinese Chestnut (*Castanea mollissima* Blume) Cultivars Revealed by GBS Resequencing. *Plants* **2022**, *11*, 3524. <https://doi.org/10.3390/plants11243524>

Academic Editor: Luigi De Masi

Received: 26 September 2022

Accepted: 10 December 2022

Published: 14 December 2022

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



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

## 1. Introduction

*Castanea mollissima* Blume, which is commonly referred to as the Chinese chestnut, is a member of the family Fagaceae. Seven species of *Castanea* are widely distributed in the temperate zone of the northern hemisphere. Four species are distributed in Asia: *C. mollissima* Blume, *C. seguinii* Dode, and *C. henryi* (Skan) Rehd. et Wils. in China and *C. crenata* Sieb. et Zucc. (Japanese chestnut) in Japan and the Korean Peninsula. Two species are distributed in North America, *C. dentata* Borkh. (American chestnut) and *C. pumila* Mill. (American chinquapin), and one species, *C. sativa* Mill. (European or sweet chestnut), is distributed in Europe. Chinese chestnut, European chestnut, and Japanese chestnut are the

main cultivated species, and they are widely planted because of their high yields of edible chestnuts and economic value. These species also have abundant germplasm resources, can grow under diverse environmental conditions, and bear nutritious fruits; Chinese chestnut is also an important source of genes for the improvement of edible varieties [1,2]. Asian chestnuts are generally much more resistant to biotic and abiotic stresses, especially Chinese chestnut. The production of chestnuts in Europe has declined gradually due to chestnut blight and ink disease since the 20th century, and American chestnut has nearly disappeared because of chestnut blight [3]. Given that Chinese chestnut shows high resistance to both diseases, it has been introduced to other regions to breed disease-resistant varieties [4]. Chinese chestnut thus plays an important role in ensuring the sustainable utilization of chestnut resources worldwide [5].

The origin of chestnut species and the center of genetic diversity of chestnuts is in mainland China [6]. The secondary origin and center of genetic diversity of chestnuts is thought to be in Turkey [7]. Chinese chestnut, one of the earliest cultivated fruit trees in China, was first depicted in the “Book of Songs” and has been cultivated for at least 3000 years [8]. Chinese chestnut is widely distributed in China, especially in the Dabie Mountains and Yanshan Mountains, and it is cultivated in at least 22 provinces. The genetic structure and domestication history of Chinese chestnut remain unclear.

The development of next-generation sequencing technology has greatly increased sequencing throughput and reduced sequencing costs [9]. The genomes of an increasing number of plant species have been sequenced, and these new data have enhanced our understanding of the genomic attributes of plant populations.

The factors affecting variation in the genome and genetic variation among populations, such as natural selection, mutation, gene drift, and gene flow, can be inferred through the study of polymorphic sites based on high-coverage whole-genome data [10]. Population genomics studies of plants have been a major focus of research in the life sciences in recent years, and such studies have greatly enhanced our understanding of the roles of genetic recombination, linkage disequilibrium, and selection on the genomes of target populations. Population genomics technology has also contributed to our ability to explore population genetic structure, the origin of cultivated populations, and the molecular mechanisms underlying complex traits [7–11]. The distribution of nucleotide diversity also provides a useful tool for inferring population history and genetic diversity [12]. This has led to an increase in the number of studies examining non-model plants with high economic value [13].

Single nucleotide polymorphisms (SNPs) are single-base differences in the genome of different individuals of a species, and they are considered the most common type of genetic variation [14,15]. SNPs are third-generation molecular markers with several advantages: they are dimorphic, high density, genetically stable, easy to detect, and present in all DNA sequences in the genome [16,17]. With the development of high-throughput technology and bioinformatics, large-scale automated detection methods have been developed, and SNPs as an important marker was used for genetic map construction, biodiversity detection, and association analysis of linkage disequilibrium [18]. However, most genetic studies of chestnuts have been conducted using simple sequence repeats (SSRs) and other second-generation molecular markers; by contrast, few genetic studies of Chinese chestnut based on SNPs have been conducted, while more recent research utilized SNPs in European chestnut [10,19,20].

Exploration of the origin and evolution of plants is important for enhancing the development and utilization of plant resources and genetically improving existing cultivars. The traditional classification of chestnut varieties was based on their morphological characteristics (e.g., leaf, fruit, and stem), biological characteristics, and economic traits [21]. However, accurately identifying chestnut varieties based on the traditional classification method of using morphological or biological characteristics is a major challenge due to the long seed-bearing period, which is not beneficial to the production and improvement of woody plant varieties.

Genotyping by sequencing (GBS) is a particularly promising genomic approach that has proven to be an effective tool for genetic studies of plants, such as barley, switchgrass, yellow mustard, olive, and Norway spruce [22–24]. GBS technology was used to conduct genome-wide association studies and characterize patterns of genomic diversity. Few studies have used GBS technology to study Chinese chestnuts to date. GBS has been used to construct a high-density genetic map and identify QTLs related to the size and ripening period of Chinese chestnut fruit [25], but few studies of the genetic structure of Chinese chestnut populations have been conducted using GBS technology. With the release of the whole genome of Chinese chestnut, GBS technology could provide new insights into the origin and evolution of Chinese chestnut, and this could aid future efforts to genetically improve chestnut varieties.

Here, GBS was performed on 185 Chinese chestnut landraces from five regions (North of the Yellow River region, Eastern Coastal region, Yangtze River Basin region, South Central region, and Midwest region) to obtain genome-wide SNP data used to characterize genetic diversity and population structure of Chinese chestnut. ADMIXTURE analysis, principal component analysis (PCA), and phylogenetic analysis were performed to clarify the genetic relationships among Chinese chestnut landraces from different regions. We also aimed to clarify the role that humans have played in the diffusion of Chinese chestnut populations.

## 2. Material and Methods

### 2.1. Plant Material

A total of 185 different landraces of Chinese chestnut (*C. mollissima*) from five regions were collected, including 39 samples from the North of the Yellow River region, 90 samples from the Eastern Coastal region, 22 samples from the Yangtze River Basin region, 24 samples from the South Central region, and 10 samples from the Midwest region. These regions span 14 provinces in China, including Beijing City, Hebei Province, Anhui Province, Jiangsu Province, Shandong Province, Zhejiang Province, Shanghai City, Hubei Province, Hunan Province, Guangxi Province, Guangdong Province, Yunnan Province, Henan Province, and Shaanxi Province (Table S1). With the exception of the hybrid clone of RISF-72 (*C. mollissima* ‘Hongli’ × *Wild Chinese chestnut*), all other accessions are locally grown *C. mollissima* cultivars.

### 2.2. Sequencing and Mapping

Leaf tissues of Chinese chestnut samples were used for DNA extraction. Young chestnut leaves were harvested, immediately frozen in liquid nitrogen, and then transferred to  $-80\text{ }^{\circ}\text{C}$ . The DNA was extracted according to Jiang et al., 2017 [11]. The GBS library was constructed via the following steps: (1) the genomic DNA was digested using restriction enzymes according to Elshire 2011 [26], and each sample was amplified and mixed after ligating the barcoded adapters; (2) required fragments were selected for library construction, and paired-end 150 sequencing was conducted using the Illumina HiSeq sequencing platform; and (3) Qubit 2.0 fluorometer (Invitrogen, Waltham, MA, USA) was used to preliminarily quantify the concentration of DNA after library construction, and the library was diluted to  $1\text{ ng}/\mu\text{L}$ . Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) was used to detect the insert size of the library. When the insert size met expectations, the effective concentration of the library was quantified using qPCR (effective library concentration  $> 2\text{ nM}$ ) to ensure the quality of the library. The different libraries were then pooled according to the effective concentration and target offline data volume and sequenced using Illumina HiSeq PE150 platform (Invitrogen, USA).

The raw data obtained from sequencing were filtered by quality control using the following criteria: (1) reads containing adapters were filtered; (2) when the N content (undetected base) in single-end sequencing reads exceeded 10% of the length of the read, the paired reads were removed; and (3) when the number of low-quality ( $\leq 5$ ) bases contained in single-end sequencing reads exceeded 50% of the length of the read, the paired

reads were removed. The products of the GBS digestion were statistically analyzed to evaluate enzymatic digestion efficiency. The number of reads with two ends generated by MseI and the ratio of the number of captured reads to the number of high-quality reads (enzymeCatchRatio) were determined. After obtaining the high-quality sequencing data, they were mapped to the reference genome (<http://gigadb.org/dataset/view/id/100643>, accessed on 15 August 2019) using the Burrows-Wheeler Alignment tool (BWA) [27]. The parameters (mem -t 4 -k 32 -M) were used; that is, local alignment was performed using four threads, the minimum seed length was 32, and shorter split hits were marked as secondary.

### 2.3. SNP Detection and Annotation

SAMTOOLS was used to detect population data in our study [28]. A Bayesian model was used to detect polymorphic loci in populations. To obtain high-quality data, SNPs with a read depth (dp) < 2, missing rate (Miss) > 0.2, and minor allele frequency (MAF) < 0.01 were removed. The obtained high-quality data were annotated using ANNOVAR software [29].

### 2.4. Population Stratification Analysis

Population genetic structure refers to the non-random distribution of genetic variation in a species or population. The subpopulations based on their geographical distributions or other criteria are usually geographically isolated individuals or populations. Different individuals within the same subpopulations are closely related to each other, and individuals of different subpopulations are more distantly related. ADMIXTURE is a program for the maximum likelihood estimation of individual ancestries based on large SNP genotype datasets. Here, it was utilized to analyze the population structure of Chinese chestnut. After generating the PLINK input Ped file, ADMIXTURE analysis was performed to characterize population genetic structure and identify population lineages.

PCA is a dimensionality reduction method that can transform the initial data set into a set of linearly uncorrelated variables. It depends on the data set and divides into two or three major axes of variation for visualization. PCA can also be used for cluster analysis. Eigenvectors and eigenvalues were calculated using GCTA software, and 2D and 3D PCA graphs were plotted using R. Individuals were clustered into different subgroups based on principal components according to SNP differences in individual genomes. We used this approach to explore the genetic structure of Chinese chestnut subpopulations.

TreeBeST software (1.9.2\_i386, SourceForge Headquarters, San Diego, CA, USA) (<http://treesoft.sourceforge.net/treebest.shtml>, accessed on 19 October 2007) was used to generate the distance matrix. The neighbor-joining method was used to construct the phylogenetic tree of Chinese chestnut with 1000 bootstrap replicates.

## 3. Results

### 3.1. Sequencing Data Analysis

A total of 185 Chinese chestnut samples were used for sequencing analysis. The total sequencing data volume was 76.24665 Gb, with an average of 412.144 Mb per sample. Results showed that the sequencing quality was high (Q20 ≥ 88.27%, Q30 ≥ 81.43%) with a normal GC distribution (Table 1).

**Table 1.** Statistics of sequencing data.

Parameter	Raw Base (bp)	Clean Base (bp)	Effective Rate (%)	Error Rate (%)	Q20 (%)	Q30 (%)	GC Content (%)
Minimum	89,598,816	89,598,816	100	0.02	88.27	81.43	33.47
Maximum	631,849,824	631,849,824	100	0.04	95.97	93.05	39.27
Mean	411,594,134	411,591,572	100	0.03	92.88	89.15	36.58
Total	76,246,654,464	76,246,175,232					

No samples were contaminated by adapter sequences, indicating a successful library construction. Then, sequencing data of 185 chestnuts were mapped to the reference genome

(724,001,627 bp). The average mapping rate of the population samples was 91–98.13%, and the average sequencing depth of the genome was 6.51–14.75 $\times$  with a 1 $\times$  coverage rate (at least one base coverage) for over 5.08% of the samples (Table 2).

**Table 2.** Statistics of sequencing depth and coverage.

Parameter	Clean Reads	Mapped Reads	Mapping Rate (%)	Average Depth (X)	Coverage at Least 1X (%)	Coverage at Least 4X (%)
Minimum	622,214	601,909	91.00	6.51	5.08	0.80
Maximum	4,387,846	4,227,193	98.13	14.75	19.20	6.13
Mean	2,858,275	2,769,234	96.91	9.74	11.80	4.32

Next, we detected a total of 3,962,053 SNP sites using SAMTOOLS, which were filtered to obtain high-quality SNPs using the criteria of dp2, Miss0.2, Maf0.01, and the filtering criteria were performed according to the description of Lipka et al., 2012 [30]. A total of 299,015 SNPs were obtained for subsequent analysis (Table 3).

**Table 3.** SNP statistics and annotation results.

Category	Number of SNPs	
Total	299,015	
Upstream	4065	
Exonic	Stop gain	209
	Stop loss	9
	Non-synonymous	3800
	Synonymous	2241
Intronic	4724	
Splicing	47	
Downstream	4449	
Upstream/Downstream	119	
Intergenic	277,994	
Transitions (ts)	204,498	
Transversions (tv)	94,517	
ts/tv	2.163	

### 3.2. Genetic Diversity Analysis of Population

We found that the five populations have high heterozygosity with obvious differences (Table 4). Results showed that the values of observed heterozygosity and expected heterozygosity of the Midwest population were the highest, reflecting the highest level of genetic diversity. Furthermore, the South Central population's genetic diversity level is close to that of the population of the Yangtze River Basin and the Eastern Coastal populations. However, the genetic diversity level of the population in the North of the Yellow River region was the lowest. Results of gene flow demonstrated higher chestnut genetic diversity in the Yangtze River Basin region, and relatively lower diversity in the Midwest region (Table 4).

**Table 4.** Genetic diversity level of five populations.

Population	Observed Heterozygosity ( $H_o$ )	Expected Heterozygosity ( $H_e$ )	Gene Flow ( $Nm$ )
North of the Yellow River	0.27963	0.30531	1.7655
Eastern Coastal region	0.30334	0.30612	1.9817
Yangtze River Basin region	0.29995	0.30800	3.2154
South Central region	0.31340	0.30824	2.2518
Midwest region	0.34265	0.33958	1.4372

Molecular variance analysis is an analysis of molecular differences. It defines different genetic structures and carries out statistical tests by classifying and dividing the studied

populations at different levels, so as to estimate the proportion of differences within populations and among individuals in the total variation among populations. The variance component among the five Chinese chestnut populations was 8.759, and the variation percentage was 2.07%. Moreover, the variance component within the population was 414.740, and the variation percentage was 97.93%, which indicates that the genetic variation of the Chinese chestnut population mainly came from within the five populations (Table 5). These results were consistent with the results of phenotypic genetic variation of Chinese chestnut populations in our previous studies.

**Table 5.** Analysis of molecular variance of population.

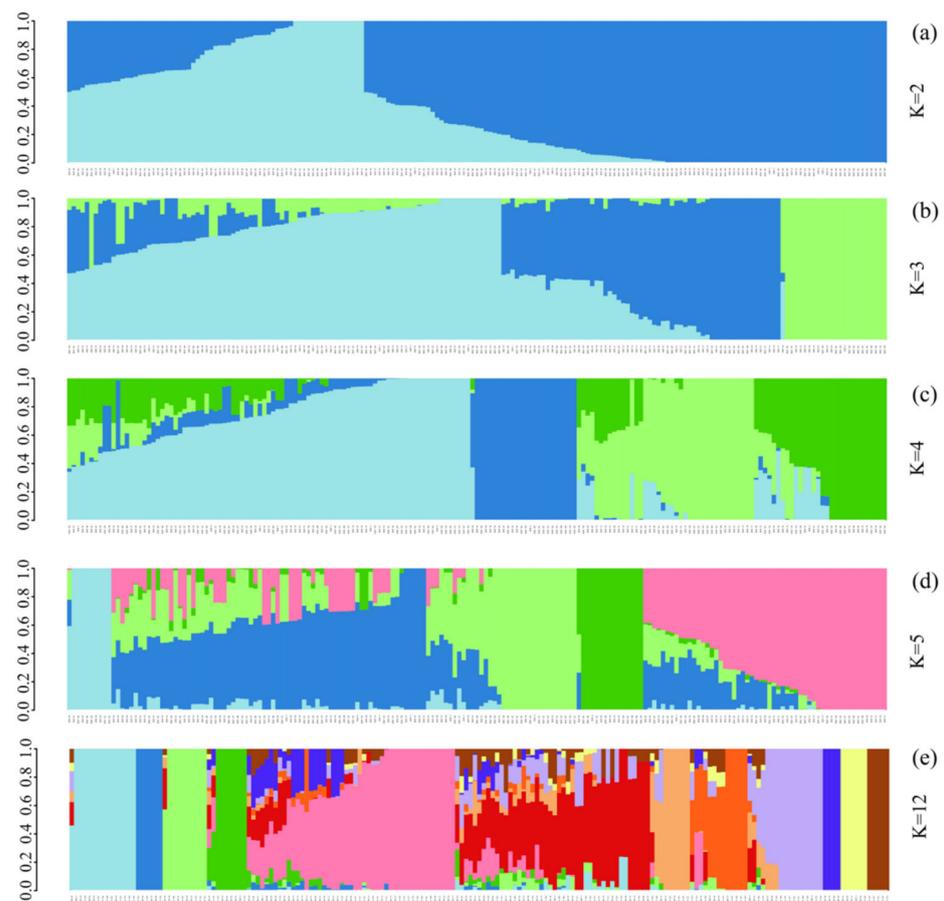
Source of Variation	df	Sum Squares	Variance of Components	Percentage of Variation (%)
Among populations	4	3879.027	8.75943	2.07
Within populations	365	151,380.206	414.74029	97.93
Total	369	155,259.232	423.49972	

### 3.3. ADMIXTURE Analysis of 185 Chinese Chestnut Landraces

ADMIXTURE analysis was performed on 299,015 SNPs to explore the population structure of all Chinese chestnuts. The number of populations (K) was set from 1 to 12, and the K optimum was selected based on the cross-validation error compared to other K values (Figure S1). As we were focused on the regional distribution of chestnuts, especially the north–south division, we focused on cases in which k ranged from 2 to 5.

A total of 185 chestnuts were divided into two groups when  $k = 2$  (Figure 1a, Table S1). The first group had 67 samples, and the second group had 118 samples. A total of 28 samples in the first group and 11 samples in the second group were from the North of the Yellow River region. A total of 27 samples and 63 samples in the first and second group, respectively, were from the Eastern Coastal region. There were 5 and 17 samples in the first and second group, respectively, from the Yangtze River Basin region. There were 4 and 20 samples in the first and second group, respectively, from the South Central region. There were 3 and 7 samples in the first and second group, respectively, from the Midwest region. Samples from the North of the Yellow River region were mainly in the first group, and samples from the other regions were mainly in the second group. These distributional patterns are consistent with the north–south geographic division around the Yellow River. However, we also observed a few irregular samples, which were mainly from Beijing (North of the Yellow River region) and Shandong Province (Eastern Coastal region). Some gene flow appears to have occurred between these two groups.

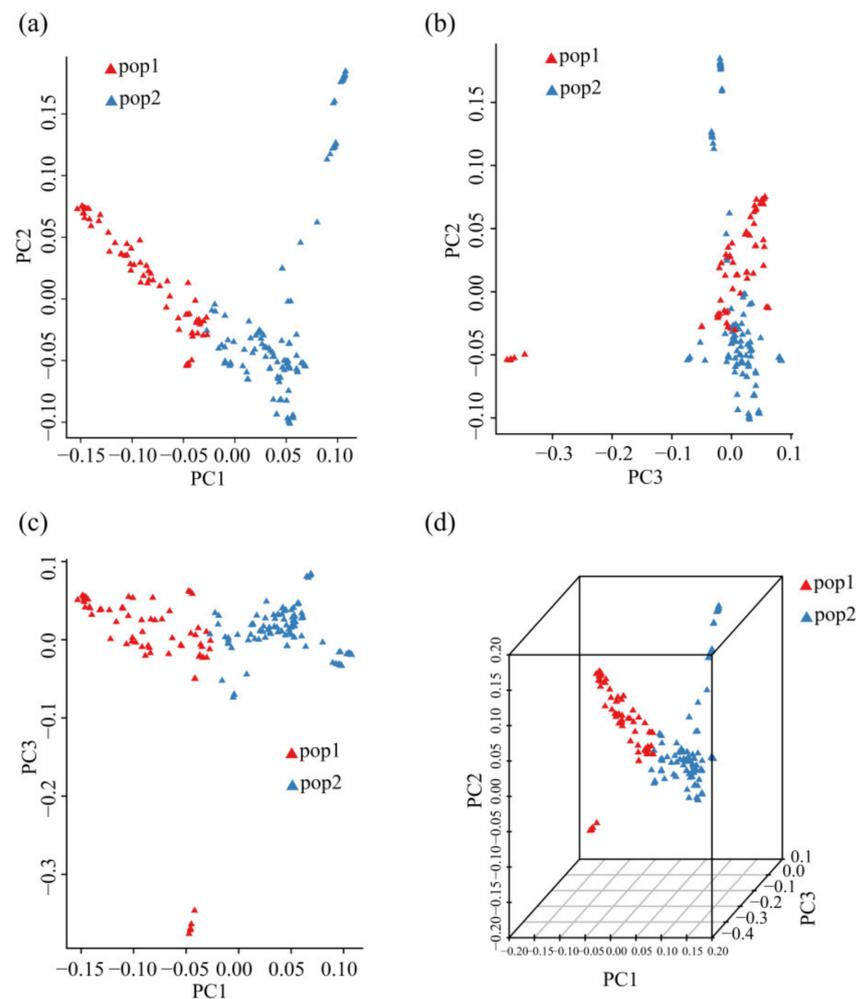
All chestnut samples were divided into three groups of 98, 63, and 24 samples when  $k = 3$  (Figure 1b). There were 10 samples in the first group, 28 samples in the second group, and 1 sample in the third group from the North of the Yellow River region. When  $k = 4$  (Figure 1c), we obtained four groups of 91, 24, 40, and 30 chestnut samples. There were 9, 1, 21, and 8 samples in the first, second, third, and fourth groups, respectively, from the North of the Yellow River region. When  $k = 5$  (Figure 1d), there were five groups of 10, 71, 34, 15, and 55 chestnut samples. There were 0, 12, 1, 1, and 25 samples in the first, second, third, fourth, and fifth groups, respectively, from the North of the Yellow River region.



**Figure 1.** Results of an ADMIXTURE analysis of 185 Chinese chestnut landraces. (a)  $k = 2$ , (b)  $k = 3$ , (c)  $k = 4$ , (d)  $k = 5$  and (e)  $k = 12$ . The  $K$  value represents the number of different subgroups, and different colors represent different subgroups.

### 3.4. PCA of 185 Chinese Chestnut Landraces

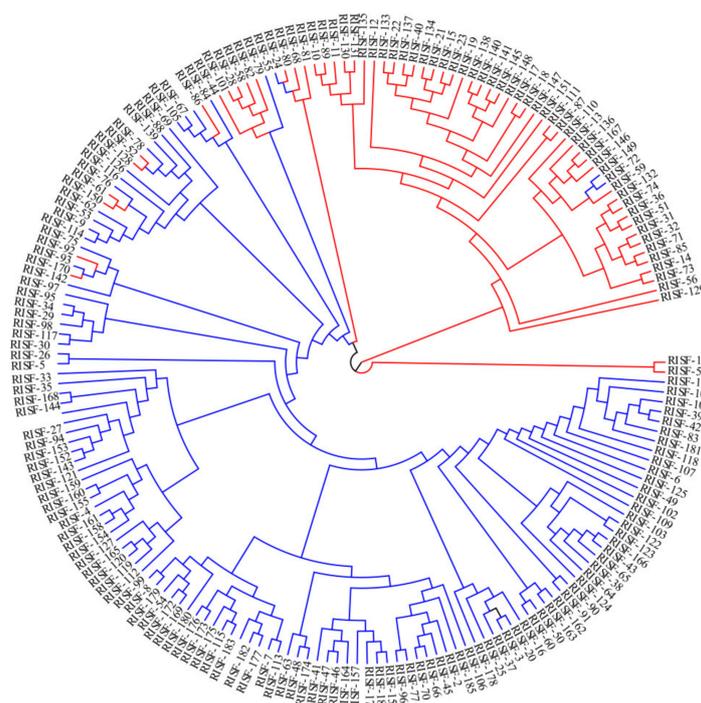
Tassel v5.0 software (<https://tassel.bitbucket.io/>, accessed on 23 February 2021) was used to analyze the results of PCA based on  $N \times \text{SNP}$  matrix [28]. PCA was performed on 299,015 SNPs from all chestnut samples; they were annotated based on the results of the ADMIXTURE when  $k = 2$ . PC1, PC2, and PC3 explained approximately 15.7%, 8.1%, and 6.0% of the total variance, which captured  $\sim 1/3$  of the genetic information of the samples. Pop1 was located on the left side of the PC1 axis, and pop2 was located on the right side of this axis, which indicates that PC1 clearly separated the two populations. However, these two populations were not separated along PC2 (Figure 2a). Most samples were clustered along PC3 in the PC2 vs. PC3 plot. Only a few pop1 samples were clearly separated from most samples along PC3, and a few pop2 samples were clearly separated from most samples along PC2 (Figure 2b). The two populations were clearly separated along PC1 in the PC1 vs. PC3 plot (Figure 2c), and a few pop1 samples were located far away from the other samples along PC3. The above conclusions can be readily observed in the plot with all three axes (Figure 2d). Samples were clearly separated by population, with the exception of a small portion of outliers of pop1 and pop2; the ADMIXTURE analysis for  $k = 2$  divided all Chinese chestnut samples into two different populations.



**Figure 2.** PCA of 185 Chinese chestnut landraces. (a) PC1 vs. PC2, (b) PC2 vs. PC3, (c) PC1 vs. PC3, (d) PC1 vs. PC2 vs. PC3.

### 3.5. Phylogenetic Analysis of 185 Chinese Chestnut Landraces

A total of 299,015 SNPs of all chestnut samples were used in the phylogenetic analysis, and samples in the tree were labeled according to the classification in the ADMIXTURE analysis when  $k = 2$  (Figure 1a, Table S1). The phylogenetic tree shows the evolutionary relationships among the different groups (Figure 3). The evolutionary branches of closely related varieties tended to be clustered within *C. mollissima* (Figure 3). The two populations were separated in the tree; only a few samples were not clustered with their respective populations. According to the clusters in Figure 3, all the individual samples were divided into the north and the south region of Yellow River Basin Region, and were represented in red and blue, respectively. However, some individual samples of the south region (shown in blue) mixed in the north region (shown in red), such as RISF 149, 72, and 24. These individual samples (belonging to the south region which was divided by the geographic position) were divided into the north region by the results of genetic distance, which indicated that different samples had gene interaction. These outlier samples were mainly from Shandong Province and Beijing City (Table S1). Combining the above results, these outliers were identified for further study and discussion (Tables S1 and S2).



**Figure 3.** Phylogenetic tree of the 185 Chinese chestnut landraces.

#### 4. Discussion

In this study, we conducted a genetic analysis of 185 Chinese chestnut landraces to explore the genetic diversity, population structure, and domestication history of Chinese chestnut. Samples were obtained from five geographic regions belonging to 14 different provinces or cities. In a previous study, a high level of genetic diversity and genetic differentiation of 279 chestnut individuals from 10 populations in Shandong province by SSR markers analysis indicated an abundant genetic diversity of Chinese chestnut resources [31]. Genetic diversity and structure analysis of chestnut populations can be a useful strategy for conservation, decision-making, and management planning [31,32]. Moreover, 95 cultivars of Chinese chestnut from ten provinces were analyzed by SSR analysis and showed a high richness in genetic diversity [11]. By estimating the genetic variability of sweet chestnut in southwest Bulgaria, Lusini et al. (2014) [33] showed that a combination of natural events and human impacts affected the genetic diversity and spatial structure. Previous studies highlighted that there was no direct relationship between geographic distribution and genetic diversity, that is, no distinct relationship between the genetic distance and the geographic distance [31]. Our results were consistent with these studies that explain the genetic variation mainly from the intra-population level.

PCA, ADMIXTURE analysis, and phylogenetic analysis were used for genetic analysis of 185 Chinese chestnut landraces. ADMIXTURE analysis of all 299,015 SNPs revealed that  $k = 2$  was the optimal number of groupings. We found that the best  $k$  value was 12 based on the CV error. However, Chinese chestnuts were divided into two main types, the north and south regions, according to the geographical, ecological, and climatic conditions, and a variety characteristics when  $k$  value was 2. Furthermore, taking the Yellow River as the dividing line between the northern and southern chestnut groups is a relatively correct viewpoint, and when  $k$  value was 2 chestnut varieties could be distinguished clearly in this study. Therefore, we divided and analyzed our groups according to the  $k$  value 2. We found that a north–south geographic division was consistent with the classification of the samples. Gene flow was detected between the two groups, but some samples from Shandong Province and Beijing City were outliers that were not clustered with their respective groups. PCA and phylogenetic analysis were performed on the SNP data, and the classification of the ADMIXTURE analysis when  $k = 2$  was used to visualize the

grouping of populations; the results of these analyses were consistent with those of the ADMIXTURE analysis. Pop1 and pop2 were clearly separated; however, samples from Beijing City and Shandong Province did not fit the north–south geographic division.

Geographical factors have a substantial influence on the characteristics of Chinese chestnut varieties, and the phenotypic diversity of Chinese chestnut varies among geographical groups [11]. Generally, the genetic relationships among chestnut resources are related to their geographical origin, but this is not always the case. Shifts in the economic center of gravity and the migration of people from northern to southern regions have also affected the historical expansion of Chinese chestnut in ancient China. There have been several migrations of people from northern to southern China due to the frequent wars that took place in northern China during the Tang (from the year of 618 to 907) and Song (from the year of 960 to 1279) Dynasties. Chinese chestnut is one of the oldest domesticated fruit trees in China [33], and it is an extremely important source of nuts and grain rich in starch. We speculate that the expansion of the Chinese chestnut benefited from the labor and advanced technology provided by northern immigrants, and the suitable climate and fertile soil in southern China accelerated the spread of chestnut from northern to southern regions. The lack of fit of Chinese chestnut samples from Shandong Province and Beijing City with the north–south geographic division might be explained by other geographical and historical factors. A study of 29 pairs of SSR primers for analyzing the genetic diversity of 26 Chinese chestnuts from different regions in Shandong Province revealed geographic variation in the genetic features. However, a few samples of Chinese chestnut from different regions and even geographically distant samples are closely related genetically; this might stem from similarity in environmental conditions and artificial selection criteria [34]. Shandong Province features both banks of the Yellow River; it is thus an area where chestnuts from both the north and south occur. For example, Chinese chestnuts in Linyi City, Shandong Province are located far from the Yellow River in Shandong Province and belong to the second subpopulation. Chinese chestnuts of Beijing City and Hebei Province belong to the Yanshan Mountain ecotype, and this region has been one of the most important areas for the production of Chinese chestnut since ancient times. On the basis of historical records, the scale of Chinese chestnut planting was very large in the Han dynasty (from 202 BC to 220 AD), and this greatly enriched the genetic diversity of chestnut in the Yanshan region, especially the Yanshan mountain ecotype. Chinese chestnuts in North China gradually expanded westward and southward and penetrated into other ecological populations. Supplies and materials from various places circulate through Beijing, which is the capital of China and the historical political center, and this might explain why the germplasm resources in North China have historically been of great significance to the evolution of chestnut.

In the same way, previous studies have shown that *Castanea* species are greatly affected by geographical factors and human activities. The historical distribution of European chestnut in glacial refugia in the Mediterranean basin has had a major effect on current patterns of genetic diversity [34]. SSR markers have shown that the potential ancestral sources of sweet chestnut in Britain and Ireland are more closely related to lineages from Western Europe rather than Eastern European lineages [35]. The Western European regions of Portugal, Spain, France, Italy, and Romania served as refugia during the Last Glacial Maximum. The geographical distribution of Chinese chestnut is divided by the Yellow River. The physiological characteristics of the recalcitrant seeds (difficult germination and short lifetime) of Chinese chestnut hindered the spread of the seeds over long distances under natural conditions; consequently, seeds can only be effectively spread via the aid of birds or rodents [36,37]. These findings confirmed our speculation that the north–south grouping of chestnut populations was influenced by human activities. Similarly, the population dynamics of sour jujube (*Ziziphus acidujubba*) and common walnut (*Juglans regia*) were also greatly affected by human behavior [38,39].

Chinese chestnut, which is considered a woody grain, is well known for its short growth cycle, high yield, strong adaptability, and its ability to grow in mountainous areas.

With its deep roots and luxuriant leaves, Chinese chestnut can not only increase the income of farmers, but also regulate the climate, fertilize the topsoil, and provide ecosystems services [1]. This suggests that the migration of populations towards the south due to wars might have resulted in the introduction of northern Chinese chestnut populations to the south as a food crop during the Tang and Song Dynasties (from the year of 618 to 1279). This might explain the close genetic relationships between some southern subpopulations and northern subpopulations. The dual function of Beijing City and Hebei Province as political and economic centers has likely enhanced the genetic diversity of chestnut. Chinese chestnut in Shandong Province occurs on both banks of the Yellow River, which might be associated with similar artificial selection criteria [31]. Genetic relationships among chestnut germplasm resources are complicated after long periods of natural selection and artificial breeding [40]. Overall, analyses of greater numbers of chestnut samples and populations might generate additional insights.

## 5. Conclusions

A lack of knowledge of the genetic relationships among chestnut populations, patterns of genetic diversity, and the domestication history of chestnut impedes future chestnut breeding efforts. SNPs, a third-generation molecular marker, were used in this study to explore the population structure of chestnuts in China. Our study showed that the genetic diversity level of the five Chinese chestnut populations from different regions was relatively high in observed heterozygosity. The population in the mid-western regions showed the highest genetic diversity, while the population north of the Yellow River showed the lowest. Molecular variance analysis showed higher variation within the group, which indicated the genetic variation of chestnut mainly from the intra-populations. ADMIXTURE analysis with  $k = 2$  revealed a north–south division of the samples, with the Yellow River as the geographical boundary. The results of the PCA and phylogenetic analysis were consistent with the results of the ADMIXTURE analysis. However, chestnuts from Shandong Province and Beijing City were outliers that were not clustered with their respective groups. Therefore, we speculate that the historical distribution of Chinese chestnut has been shaped by human activities, including several migration events driven by wars. The chestnuts in Shandong Province and Beijing City occur in an area with samples from both northern and southern regions, which is likely related to their geographical and political importance. Ultimately, we aimed to obtain information that could aid the genetic diversity evaluation and conservation strategies of chestnut genetic resources in China.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants11243524/s1>, Table S1. Source information and ADMIXTURE results ( $k = 2$ ) of 185 Chinese Chestnut landraces; Table S2. Outliers of the phylogenetic analysis; Figure S1. K values in the ADMIXTURE analysis of 185 Chinese chestnut landraces.

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

**Funding:** This research was funded by the National Key R&D Program of China (2019YFD1001600) and Zhejiang Science and Technology Major Program on Agriculture New Variety Breeding (2021C02070-4).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** RAW VCF and SRA Database Information can be obtained at request from the corresponding author, all other data are comprised in the manuscript.

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

## References

1. Zhang, S.; Wang, L.T.; Fu, Y.J.; Jiang, J.C. Bioactive constituents, nutritional benefits and woody food applications of *Castanea mollissima*: A comprehensive review. *Food Chem.* **2022**, *393*, 133380. [[CrossRef](#)] [[PubMed](#)]
2. Bao, Y.; Jiang, G.; Gu, C.; Yang, H.; Jiang, Y. Structural changes in polysaccharides isolated from chestnut (*Castanea mollissima* Bl.) fruit at different degrees of hardening. *Food Chem.* **2010**, *119*, 1211–1215.
3. Brewer, L.G. Ecology of Survival and Recovery from Blight in American Chestnut Trees (*Castanea dentata* (Marsh.) Borkh.) in Michigan. *Bull. Torrey Bot. Club* **1995**, *122*, 40–57. [[CrossRef](#)]
4. Anagnostakis, S.L. Chestnut blight: The classical problem of an introduced pathogen. *Mycologia* **1987**, *79*, 23–37. [[CrossRef](#)]
5. De Vasconcelos, M.D.C.B.M.; Bennett, R.N.; Rosa, E.A.S.; Cardoso, J.V.F. Primary and Secondary Metabolite Composition of Kernels from Three Cultivars of Portuguese Chestnut (*Castanea sativa* Mill.) at Different Stages of Industrial Transformation. *J. Agric. Food Chem.* **2007**, *55*, 3508–3516. [[CrossRef](#)]
6. Rutter, P.; Miller, G.; Payne, J. Chestnuts. In *Genetic Resources of Temperate Fruit and Nut Crops*; Moore, J.N., Ballington, J.R., Jr., Eds.; The International Society for Horticultural Science: Wageningen, The Netherlands, 1990; pp. 761–788.
7. Zohary, D.; Hopf, M.; Weiss, E. *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin*; Oxford University Press: Oxford, UK, 2012; pp. 220–227.
8. LaBonte, N.R.; Zhao, P.; Woeste, K. Signatures of Selection in the Genomes of Chinese Chestnut (*Castanea mollissima* Blume): The Roots of Nut Tree Domestication. *Front. Plant Sci.* **2018**, *9*, 810. [[CrossRef](#)]
9. Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends Genet.* **2008**, *24*, 133–141. [[CrossRef](#)]
10. Ong, Q.; Nguyen, P.; Thao, N.P.; Le, L. Bioinformatics Approach in Plant Genomic Research. *Curr. Genom.* **2016**, *17*, 368–378. [[CrossRef](#)]
11. Jiang, X.B.; Tang, D.; Gong, B.C. Genetic diversity and association analysis of Chinese chestnut (*Castanea mollissima* Blume) cultivars based on SSR markers. *Braz. J. Bot.* **2017**, *40*, 235–246. [[CrossRef](#)]
12. Tang, R.; Liu, E.X.; Zhang, Y.Z.; Schinnerl, J.; Sun, W.B.; Chen, G. Genetic diversity and population structure of *Amorphophallus albus*, a plant species with extremely small populations (PSESP) endemic to dry-hot valley of Jinsha River. *BMC Genet.* **2020**, *21*, 102. [[CrossRef](#)]
13. Unamba, C.I.N.; Nag, A.; Sharma, R.K. Next Generation Sequencing Technologies: The Doorway to the Unexplored Genomics of Non-Model Plants. *Front. Plant Sci.* **2015**, *6*, 1074. [[CrossRef](#)] [[PubMed](#)]
14. Rayda, B.A.; Fabienne, M.; Hajer, B.H.; Ahmed, R.; Sezi, E.; Narendra, K.; Mohsen, H.; Amine, A.; Riaz, U.; Essam, A.A. SNP discovery and structural insights into OeFAD2 unravelling high oleic/linoleic ratio in olive oil. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 1229–1243.
15. Liu, G.; Xie, Y.J.; Zhang, D.Q.; Chen, H.P. Analysis of SSR loci and development of SSR primers in *Eucalyptus*. *J. For. Res.* **2018**, *29*, 273–282. [[CrossRef](#)]
16. Li, Y.F.; Zhang, J.J.; Chang, S.X.; Jiang, P.K.; Zhou, G.M.; Shen, Z.M.; Wu, J.S.; Lin, L.; Wang, Z.S.; Shen, M.C. Converting native shrub forests to Chinese chestnut plantations and subsequent intensive management affected soil C and N pools. *For. Ecol. Manag.* **2014**, *312*, 161–169. [[CrossRef](#)]
17. Liu, Y.L.; Li, S.Q.; Wang, Y.Y.; Liu, P.Y.; Han, W.J. De novo assembly of the seed transcriptome and search for potential EST-SSR markers for an endangered, economically important tree species: *Elaeagnus mollis* Diels. *J. For. Res.* **2019**, *31*, 759–767. [[CrossRef](#)]
18. Taranto, F.; D'Agostino, N.; Greco, B.; Cardi, T.; Tripodi, P. Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. *BMC Genom.* **2016**, *17*, 943. [[CrossRef](#)]
19. Nunziata, A.; Ruggieri, V.; Petriccione, M.; De Masi, L. Single Nucleotide Polymorphisms as Practical Molecular Tools to Support European Chestnut Agrobiodiversity Management. *Int. J. Mol. Sci.* **2020**, *21*, 4805. [[CrossRef](#)]
20. Nunziata, A.; Ferlito, F.; Magri, A.; Ferrara, E.; Petriccione, M. The Hundred Horses Chestnut: A model system for studying mutation rate during clonal propagation in superior plants. *Forestry* **2022**, *95*, 678–685. [[CrossRef](#)]
21. Chang, L.; Wang, S.; Chang, X.; Wang, S.J.F.H. Structural and functional properties of starches from Chinese chestnuts. *Food Hydrocoll.* **2015**, *43*, 568–576.
22. He, J.F.; Zhao, Q.X.; Laroche, A.; Lu, Z.X.; Liu, H.K.; Li, Z.Q. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* **2014**, *5*, 484. [[CrossRef](#)]
23. Nunzio, D.A.; Francesca, T.; Salvatore, C.; Giacomo, M.; Valentina, F.; Susanna, G.; Monica, M.M.; Stefano, P.; di Valentina, R.; Wilma, S.; et al. GBS-derived SNP catalogue unveiled wide genetic variability and geographical relationships of Italian olive cultivars. *Sci. Rep.* **2018**, *8*, 15877.
24. Jiří, K.; Jaroslav, Č.; Jan, S.; Zuzana, F.; Jakub, D.; Milan, L.; Yousry, A. Genetic diversity of Norway spruce ecotypes assessed by GBS-derived SNPs. *Sci. Rep.* **2021**, *11*, 23119.
25. Ji, F.; Wei, W.; Liu, Y.; Wang, G.; Zhang, Q.; Xing, Y.; Zhang, S.; Liu, Z.; Cao, Q.; Qin, L. Construction of a SNP-Based High-Density Genetic Map Using Genotyping by Sequencing (GBS) and QTL Analysis of Nut Traits in Chinese Chestnut (*Castanea mollissima* Blume). *Front. Plant Sci.* **2018**, *9*, 816. [[CrossRef](#)] [[PubMed](#)]
26. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **2011**, *6*, e19379. [[CrossRef](#)]
27. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]

28. García-Arias, F.L.; Osorio-Guarín, J.A.; Zarantes, V.M.N. Association Study Reveals Novel Genes Related to Yield and Quality of Fruit in Cape Gooseberry (*Physalis peruviana* L.). *Front. Plant Sci.* **2018**, *9*, 362. [[CrossRef](#)]
29. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [[CrossRef](#)]
30. Lipka, A.E.; Tian, F.; Wang, Q.; Peiffer, J.; Li, M.; Bradbury, P.J.; Gore, M.A.; Buckler, E.S.; Zhang, Z. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **2012**, *28*, 2397–2399. [[CrossRef](#)]
31. Ai, C.X.; Li, G.T.; Zhang, L.S.; Liu, Q.Z. Study on the genetic diversity of natural chestnut populations in Shandong China by SSR markers. *Acta Hortic.* **2009**, *844*, 257–266. [[CrossRef](#)]
32. Allendorf, F.W.; Hohenlohe, P.A.; Luikart, G. Genomics and the future of conservation genetics. *Nat. Rev. Genet.* **2010**, *11*, 697–709. [[CrossRef](#)]
33. Lusini, I.; Velichkov, I.; Pollegioni, P.; Chiocchini, F.; Hinkov, G.; Zlatanov, T.; Cherubini, M.; Mattioni, C. Estimating the genetic diversity and spatial structure of Bulgarian *Castanea sativa* populations by SSRs: Implications for conservation. *Conserv. Genet.* **2014**, *15*, 283–293. [[CrossRef](#)]
34. Hao, H.N.; Li, Q.; Bao, W.J.; Wu, Y.W.; Ouyang, J. Relationship between physicochemical characteristics and in vitro digestibility of chestnut (*Castanea mollissima*) starch. *Food Hydrocoll.* **2018**, *84*, 193–199. [[CrossRef](#)]
35. Aravanopoulos, F.A.; Jarman, R.; Mattioni, C.; Russell, K.; Chambers, F.M.; Bartlett, D.; Martin, M.A.; Cherubini, M.; Villani, F.; Webb, J. DNA analysis of *Castanea sativa* (sweet chestnut) in Britain and Ireland: Elucidating European origins and genepool diversity. *PLoS ONE* **2019**, *14*, e0222936. [[CrossRef](#)]
36. Chen, X.N.; Zhang, B.; Chen, Y.J.; Hou, X.; Wang, J.; Chang, G. Effect of forest rodents on predation and dispersal of *Castanea mollissima* and *Quercus aliena* seeds on south- and north-facing slopes of Qinling Mountains. *Acta Ecol. Sin.* **2016**, *36*, 1303–1311. [[CrossRef](#)]
37. Cao, L.; Xiao, Z.S.; Guo, C.; Chen, J.; Zhang, Z.B. Scatter-hoarding rodents as secondary seed dispersers of a frugivore-dispersed tree *Scleropyrum wallichianum* in a defaunated Xishuangbanna tropical forest. *China. Integr. Zool.* **2011**, *6*, 227–234. [[CrossRef](#)]
38. Zhang, C.M.; Huang, J.; Yin, X.; Lian, C.L.; Li, X.G. Genetic diversity and population structure of sour jujube, *Ziziphus acidujubata*. *Tree Genet. Genomes* **2014**, *11*, 809. [[CrossRef](#)]
39. Feng, X.J.; Zhou, H.J.; Zulfiqar, S.; Luo, X.; Hu, Y.; Feng, L.; Malvolti, M.E.; Woeste, K.; Zhao, P. The Phylogeographic History of Common Walnut in China. *Front. Plant Sci.* **2018**, *9*, 1399. [[CrossRef](#)]
40. Bouffartigue, C.; Debille, S.; Fabreguettes, O.; Cabrer, A.R.; Lorenzo, S.P.; Flutre, T.; Harvengt, L. Two main genetic clusters with high admixture between forest and cultivated chestnut (*Castanea sativa* Mill.) in France. *Ann. For. Sci.* **2020**, *77*, 74. [[CrossRef](#)]