SNPs Detection and Genetic Analysis of *Chionanthus Retusus* via Genotyping-by-Sequencing

Song Huang[1], Wei Zheng[1], Xiangpeng Liang[1], Qingda Duan[1], Juan Wang[1], Yaoqing Sun[1], Tianxiao Ma[1]*

[1] College of Forestry, Xinyang Agricultural and Forestry University, Xinyang 464000, China; 13803862987@126.com(S.H.)

*Correspondence author: Tianxiao Ma, email: matianxiao1979@163.com

Abstract

Chionanthus retusus is one of the well-known ornamental trees in East Asia and America. Not only its value in the market but also had the potential as a source for producing antioxidant. However, due to uncontrolled exploitation, the number of wild *C. retusus* in China is decreasing rapidly. The genetic study of *C. retusus* is limited. In order to investigate the genetic diversity and the distribution of *C. retusus* in China, 47 samples from 8 different provinces have been sequenced via restriction site-associated DNA sequencing (RAD-seq). Totally, 31, 402 high-quality single nucleotide polymorphisms (SNP) were obtained. According to the phylogenetic tree and the principal component analysis, the samples were divided into four populations, including 3 major populations and 1 hybrid population. Population1 were the samples from Jiangsu and Yunnan province and the Population2 were mainly from northern and northeast of China including Liaoning and Hubei province, while the Population4 were from Shandong and Henan province, which were in central China. As the admixture showed, the population3 were the offspring of the other 3 populations by hybridization. The mean heterozygosity of Chinese Fringe tree from different province is 0.42 %, with the highest heterozygosity, which is as high as 0.63 %, from Jiangsu province and the lowest heterozygosity, which is only 0.19, from Henan province. This is the first report about the genetic diversity and relationship of *Chionanthus retusus*, which will provide value information for further genetic study, genomic study, conservation and breeding.

Keywords: *Chionanthus retusus, genotyping-by-sequencing, restriction site-associated DNA sequencing, single nucleotide polymorphism, genetic diversity*

Introduction

Chionanthus retusus (Chinese fringetree, *C. retusus*) is one of the famous ornamental trees in the world, which belongs to the family *Oleaceae*. Its flower is pure white when it is blooming in April, all the trees are whitewashed (Figure 1). Thus, it is also well-known as "April Snow" in China. *C. retusus* is a species native to China, Korean, and Japan [1]. Besides used as an ornamental tree, *C. retusus* and *Chionanthus virginicus* (White fringetree, which is native to America) are also used as a folk medicine for treating skin inflammation, diarrhea, and stomachache in the past [2-3]. It is mainly because of the presence of a large number of flavonoids [4] and secoiridoids [5] which have the cytotoxicity and antioxidant activity [6-7] in the leaves and wood bark. Moreover, its leaves can be used for tea-making [8] and its flowers, fruits, and seeds can be used for producing oil [9-10].

However, due to the human activities and lack of protection, the habitat of *C. retusus* is sharply decreasing and fragmented. Therefore, it is listed as "vulnerable" in Japan [11]. In China, even though some individual trees have been locally protected, its number is still dramatically declining due to stolen for its market value and insufficient conservation [12]. It is in urge to pay attention to the protection of the wild germplasm resources of *C. retusus*.

DOI:10.2478/sg-2023-0012
edited by the Thünen Institute of Forest Genetics



Figure 1
A flowering *C. retusus* in Shanxi province.

Using genotyping and molecular marker to investigate the genetic diversity and variation have become one of the critical methods for understanding plant origins and diversity. Recently, there are several techniques have been established and widely used for single nucleotide polymorphisms (SNPs) and genetic diversity studies, including restriction fragment length polymorphism (RFLP) [13], amplified fragment length polymorphism (AFLP) [14], simple sequence repeats (SSRs) [15], etc. Due to the recent advancing development in the sequencing technique, genotyping-by-sequencing (GBS) is also considered as a cost-effective and powerful method for studying genetic diversity. Restriction site-associated DNA sequencing (RAD-seq) is one of the common GBS methods nowadays and has already been applied in various studies for exploring numerous SNP markers in the genomic level [16-17]. However, the genetic and genomic study of *C. retusus* is limited.

In this report, we collected 47 *C. retusus* samples from 17 cities in 8 different provinces in China, which were used to establish the RAD-seq library. We reported the subdivision, the heterozygosity and the population diversity of *C. retusus* based on RAD-seq.

Materials and Methods

Samples collection and genomic DNA extraction

All the samples were collected from forest or wild fields. The location and the representing number of the samples were listed in Table 1. Several leaves were collected and stored in the liquid nitrogen. Then the genomic DNA was extracted using CTAB-method [18]. Agarose electrophoresis and Nanodrop 2000C (Thermo Fisher, USA) were used to inspect the integrity and purity of the extracted DNA. Qubit 2.0 Fluorometer (Invitrogen, USA) was used to measure the concentration of the DNA. Only the qualified DNA was sent to Novogene Bioinformatics Technology CO. Ltd (Beijing, China) for library construction and sequencing.

The genomic DNAs were digested by HaeIII and the digested DNA samples were used for the GBS library construction [19]. Illumina HiSeq was used as the sequencing platform and 150 bp paired-end (PE) sequences were obtained. The quality of the sequencing also was estimated by the number of clean reads, sequencing depth, and Q30.

As there was no C. retusus genome assembled till now, the sequences assembled from RAD-seq of sample "jin_1_1" was selected as the reference.

SNP detection and annotation

The qualified sequencing samples was compared with the reference sequences via BWA [20], SNPs were detected via SAMTOOLS [21] and the results were filtered by the condition "dp4-miss0.5-maf0.1" via VCFTOOLS [22]. In other words, only the SNPs which have reading depth higher than 4 and existed in more than 50% with the minor allele percentage more than 10% in the sample pool would be selected for further analysis.

Genetic structure and diversity analysis

The phylogenetic tree was constructed by TreeBest using the neighbor-joining method and bootstrap 1000 times. The genetic structure was constructed by PLINK [23] and admixture (K=2~8). GCTA [24] was used for principal components analysis (PCA) which can be used for assisting the structure analysis. The principal components were plotted in 2D and 3D plots for visualization. The observed heterozygosity (Ho), expected heterozygosity (He) and analysis of molecular variance (AMOVA) were calculated using the filtered SNPs via Arlequin [25].

Table1
The location and representing number of each sample.

Collecting locations	Number of samples in the province	number of samples
Henan province	15	41.43.45.47.50.51.52.53.69.70.71.75.85.86.87
Shanxi province	11	60.61.62.63.64.65.66.67.68.78.79
Shandong province	7	42.44.46.48.49.76.77
Jiangsu province	3	54.55.56
Hubei province	3	72.73.74
Yunnan province	3	80.81.82
Liaoning province	3	57.58.59
Shaanxi province	2	83.84

The genomic DNAs were digested by HaeIII and the digested DNA samples were used for the GBS library construction [19]. Illumina HiSeq was used as the sequencing platform and 150 bp paired-end (PE) sequences were obtained. The quality of the sequencing also was estimated by the number of clean reads, sequencing depth, and Q30.

As there was no *C. retusus* genome assembled till now, the sequences assembled from RAD-seq of sample "jin_1_1" was selected as the reference.

SNP detection and annotation

The qualified sequencing samples was compared with the reference sequences via BWA [20], SNPs were detected via SAMTOOLS [21] and the results were filtered by the condition "dp4-miss0.5-maf0.1" via VCFTOOLS [22]. In other words, only the SNPs which have reading depth higher than 4 and existed in more than 50 % with the minor allele percentage more than 10 % in the sample pool would be selected for further analysis.

Genetic structure and diversity analysis

The phylogenetic tree was constructed by TreeBest using the neighbor-joining method and bootstrap 1000 times. The genetic structure was constructed by PLINK [23] and admixture (K=2~8). GCTA [24] was used for principal components analysis (PCA) which can be used for assisting the structure analysis. The principal components were plotted in 2D and 3D plots for visualization. The observed heterozygosity (Ho), expected heterozygosity (He) and analysis of molecular variance (AMOVA) were calculated using the filtered SNPs via Arlequin [25].

Results

Enzyme digestion and sequencing results

47 samples from 8 different provinces have been sequenced. After filtering the low-quality result, 74.39 Gb high quality clean data has been obtained (Q20 ≥ 95.35 %, Q30 ≥ 88.79 %). The capture efficiency of HaeIII cleavage sites was more than 99 % in all the samples and the cut efficiency of it was more than 97 %, which provided fair-well data for following analysis. The average GC content was 37.13 % with standard deviation of 0.23, which also suggested that the sequencing quality was reliable (Table 2). More than 78 % of reads were mapped to the reference with the depth ranged from 8.8 to 14.3, about half of which had a minimum depth of 4 (Table S1).

After SNP calling with SAMTOOLS, 535,491 SNPs were detected in 47 samples including sample "jin_1_1". After filtering the raw SNPs with the condition "dp4-miss0.5-maf0.1", 31, 402 high-quality SNPs were gotten and used for further analysis.

To detect the heterozygous rate of tassel trees, we used a native perl script to count sites with depth of at least 4. The heterozygous site percentage ranged from 0.71 % to 1.51 %, which was consistent with cross-pollination in tassel trees.

Genetic structure and phylogenetic analysis

As there is no reference genome of *C. retusus,* RAD-seq is an effective way to learn about its genetic information. The 47 *C. retusus* samples from 8 provinces were divided into 4 groups in the phylogenetic tree which was constructed with TreeBest based on the filtered SNPs. Group G1 contains 4 samples, mainly from the central and eastern provinces, including 2 from Shanxi Province, 1 from Shandong province and 1 from Henan province (Figure 2). Group G2 consisted of 16 samples, mainly from central, eastern and northeastern provinces, among which 5 were from Shanxi Province, 3 from Jiangsu Province, 3 from Liaoning Province, 3 from Shandong Province, and 2 from Henan Province (Figure 2). Group G3 contains 9 samples, mainly from central China, including 4 from Henan Province, 2 from Hubei Province, 2 from Shanxi Province, and 1 from Shandong Province, while group G4 consisted of 18 samples mainly from central and southern China, among which 8 were from Henan Province, 3 were from Yunnan Province, 2 were from Hubei Province, 2 were from Shanxi Province, 2 were from Shaanxi province, 2 were from Shandong Province, and 1 was from Hubei province (Figure 2). The samples from Shanxi Province were distributed in all groups and were major in group G1 and G2, while the samples from Henan Province were major in group G4. Furthermore, the samples were divided into four groups clearly based on PCA analysis, but the composition of the groups was slightly different from the result of phylogenetic tree above (Figure 3). Three components were selected for the PCA and their variance ratio were 15.57 %, 12.75 %, and 12.43 % respectively, which could explain more than 40 % of the variance. According to the PCA analysis, the samples were divided into the Liaoning province group in purple, the Yunnan province group in pink, the Jiangsu province group in green and the other four province, locating in central of China (Figure 3a-d). The samples from Jiangsu province group had some characteristics closed to the samples from Henan, Shanxi, Shaanxi, Shandong and Hubei province (Figure 3c-d).

Based on the genetic structure analysis, the CV error was smallest when K=3, which was selected to do the population analysis. The samples were divided into four populations, including 3 major populations and 1 hybrid population. Population1 were the samples from Jiangsu and Yunnan province and the Population2 were mainly from northern and northeast of China including Liaoning and Hubei province, while the Population4 were from Shandong and Henan province, which were in central China (Figure 4). Interestingly, most of the samples from Population3, which were located in central China, had the genetic background of Population 1, Population2 and Population3. It was suggested that the Population3 were hybrid which were mainly from the southern of China (Figure 2). This indicates that the central fringe tree hybridized with the southern and northeastern fringe trees to form the central fringe tree variety with diverse genetic information.

Table2
The result from the quality control of the sequencing.

Sample	Raw Base(bp)	Clean Base(bp)	Effective Rate(%)	Error Rate(%)	Q20(%)	Q30(%)	GC Content(%)
57	1778090400	1778014080	100	0.03	96.81	91.09	37.2
58	2060330112	2060295840	100	0.03	96.4	90.06	37.14
59	1908112896	1908080064	100	0.03	96.31	89.91	37.28
41	1882378944	1882304640	100	0.03	96.27	89.77	36.86
42	1707708960	1707582816	99.99	0.03	96.31	89.81	37.2
43	1730256768	1730224800	100	0.03	95.91	88.96	37.31
44	2118818016	2118716064	100	0.03	96.59	90.45	36.84
45	1786419360	1786325760	99.99	0.03	96.38	89.97	37.55
46	1803811104	1803720384	99.99	0.03	96.59	90.45	37.24
47	1834366464	1834285248	100	0.03	95.59	88.16	37.4
48	1800475776	1800393120	100	0.03	97.17	91.78	37.09
49	2136291264	2136252096	100	0.03	96.42	90.08	36.53
64	2202710688	2202668928	100	0.03	96.16	89.53	36.91
62	1849620096	1849584960	100	0.03	95.43	87.85	37.23
63	1968549120	1968444000	99.99	0.03	95.66	88.43	37.33
54	1893888288	1893856320	100	0.03	95.8	88.58	37.47
55	1776820896	1776789504	100	0.03	95.78	88.54	37.58
56	1731249504	1731213504	100	0.03	96.3	89.84	37.2
60	1790937504	1790904096	100	0.03	97.05	91.63	37.08
61	1985645376	1985602176	100	0.03	96.58	90.51	37.16
51	1947701088	1947613536	100	0.03	96.66	90.58	37.15
52	2172217536	2172100608	99.99	0.03	96.24	89.53	36.63
53	1829335104	1829246400	100	0.03	95.75	88.59	37.14
50	1992242592	1992157056	100	0.03	96.77	90.91	37.18
85	1582121088	1582011360	99.99	0.03	96.3	89.66	37.28
87	1595263968	1595191104	100	0.03	97.1	91.56	37.01
86	1490811552	1490748480	100	0.04	94.75	86.08	37.44
65	1477948608	1477855872	99.99	0.03	96.54	90.17	37
66	1413196416	1413138816	100	0.03	96.54	90.12	37.43
67	1472742720	1472678208	100	0.03	96.49	89.98	37.25
68	1460337408	1460270880	100	0.03	96.95	91.22	37.39
75	1454738976	1454628096	99.99	0.03	96.45	90	37.27
76	1693549440	1693433088	99.99	0.03	96.8	90.88	36.83
84	1737601056	1737474624	99.99	0.03	96.97	91.2	36.86
79	1422963648	1422841248	99.99	0.03	97.66	93.12	36.79
83	1290729312	1290672864	100	0.03	96	88.91	37.17
78	1493767584	1493704224	100	0.03	97.48	92.64	37
72	1369787328	1369682208	99.99	0.03	97.55	92.82	36.95
73	1446712704	1446593472	99.99	0.03	97.19	91.89	36.83
74	1435069440	1435010688	100	0.03	96.18	89.34	37.25
69	1291753152	1291694688	100	0.03	97.34	92.34	37
70	1470092832	1470022848	100	0.03	97.03	91.45	37.25
71	1483340832	1483231968	99.99	0.03	96.91	91.25	37.12
80	1583409024	1583341632	100	0.03	96.9	91.04	37.19
81	1567956672	1567884096	100	0.03	96.95	91.33	37.08
82	1469225088	1469167488	100	0.03	96.44	89.95	37.18
77	1495128096	1495022688	99.99	0.03	95.9	88.69	37.26

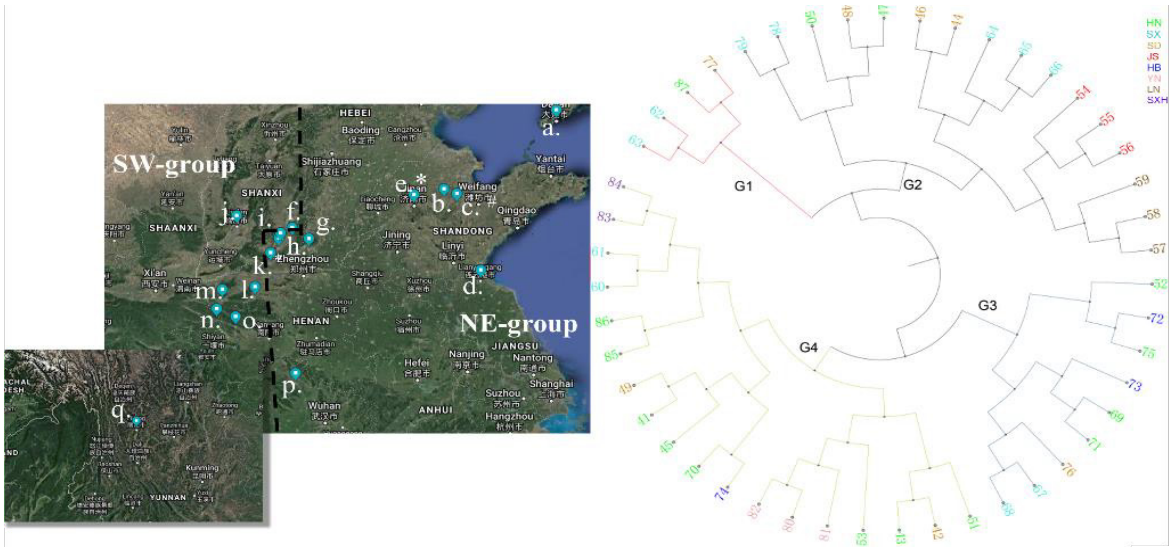


Figure 2

The constructed phylogenetic tree based on the high-quality SNP result. The similar colors are representing that the samples come from different cities in the same province. The map is obtained from Google Maps. The location of each point: a. Dalian, b. Zibo, c. Qingzhou, d. Lianyungang, e. Jinan, f. Lingchuan, g. Huixian, h. Zezhou, i. Jincheng, j. Linfen, k. Jiyuan, l. Songxian, m. Lushi, n. Shangnan, o. Xixia, p. Suizhou, q. Lijiang. * (Jinan and Jiyuan) are the cities which also contain samples could be divided into NE and SE group. # (Qingzhou) is the one that is different from the geological division.

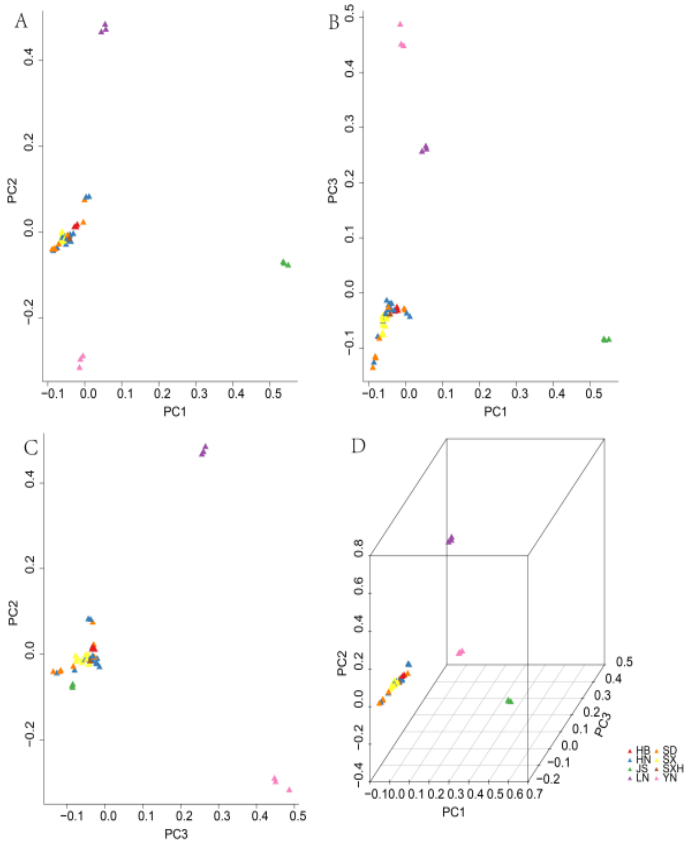


Figure 3

a), b), c) are the PCA plotted with PC1/PC2, PC2/ PC3, and PC1/PC3 respectively. d) is the PCA plotted with PC1/PC2/PC3 in 3D. The plane is the calculated plane regression based on the point plotted in the Fig which is showing that the points can be divided into 2 planes when plotting 3 principle components in 3D.

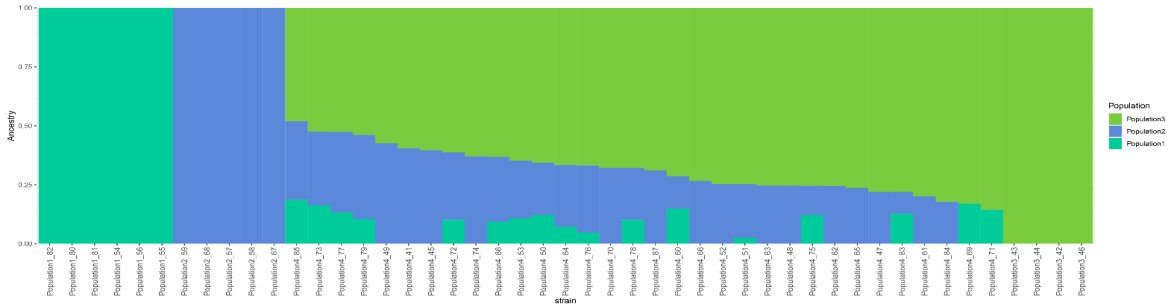


Figure 4
The genetic structure analysis of the *C. retusus*.

Diversity analysis

The degree of heterozygosity reflects the degree of genetic variation in the population, and the higher the degree of heterozygosity, the higher the genetic diversity within the population. The observed heterozygosity (Ho) and expected heterozygosity (He) Were calculated for all the provinces. It was interesting to find that the province from Population1, Population2 and Population4 had higher Ho and He, such as the samples from Jiangsu province have the highest Ho and He, 0.769 and 0.624 respectively. It was also observed that the samples from Henan Province had the lowest Ho and He, which suggested that the tassel trees in Henan were mainly imported from other provinces, especially from Shandong province.

Discussion

Diversity analysis

The degree of heterozygosity reflects the degree of genetic variation in the population, and the higher the degree of heterozygosity, the higher the genetic diversity within the population. The observed heterozygosity (Ho) and expected heterozygosity (He) Were calculated for all the provinces. It was interesting to find that the province from Population1, Population2 and Population4 had higher Ho and He, such as the samples from Jiangsu province have the highest Ho and He, 0.769 and 0.624 respectively. It was also observed that the samples from Henan Province had the lowest Ho and He, which suggested that the tassel trees in Henan were mainly imported from other provinces, especially from Shandong province.

Since the development of the sequencing technique and reduction of the sequencing cost, GBS has been one of the widely used techniques for investigating the genetic diversity and relationship in various plant species [26-27]. It is also considered as a rapid and effective study strategy for non-model organisms, even without the presence of the reference genomic data [19,28-29]. *C. retusus* is a well-known ornamental tree with high potential for various applications in society. However, there are only a few reports related to its genomic data [12]. All the 47 DNA samples of *C. retusus* are digested by restriction enzyme HaeIII and the fragments are sequenced. The GC distribution, reading depth, and Q30 suggest that the sequencing quality is acceptable. However, the amount of collected samples in this study is much lower than expected, thus the MAF is set as 0.1 and indeed this condition did not affect the filtered result.

The phylogenetic tree and PCA indicated that the collected *C. retusus* samples can be geologically divided into 4 groups. The north and northeast group in G2 were different from the southern group in G4.

The genetic structure also indicated that besides the samples can be geologically divided into 4 populations. Most of the samples were grouped into Population3, which were the hybrids from Population1 mainly in southern, Population2 in northeast and Population4 in Shandong province. It was suggested that the *C. retusus* in central China were the offspring by hybridization of the other 3 populations.

Baali-Cherif and Besnard report the same family plant *Olea europaea* has a high genetic diversity with the Ho and He, 0.64 and 0.61 [30-31]. In our study, He and Ho are varied and not related to the location at all. However, the samples from Jiangsu province have the highest Ho and He, 0.769 and 0.624, which might due to that the *C. retusus* collected in Jiangsu province are being protected in the field park thus they are less influenced by human behavior and habitat damage [32-33]. The Population1 mainly included the samples from Jiangsu and Yunnan province, which was one of the ancestors of *C. retusus* in central China.

SNP is the most accurate molecular marker technology at present [34]. Its resolution can reach single base, and it has many advantages, such as abundant sites, high genetic stability, objectified results, and suitable for automated and large-scale analysis. It can carry out accurate genetic diversity research on closely related species, even varieties with different genotypes. It can also be used for gene mapping and integration, gene function and marker-assisted breeding. It can provide scientific basis for rational conservation and utilization of gene resources and genetic improvement [35].

Conclusions

In this report, we have sequenced and obtained 31, 892 high-quality SNPs for genetic relationship and genetic diversity analysis. The result showed that most of the samples can be geologically divided into 4 groups. Furthermore, the samples from Jiangsu province have the highest genetic diversity, which suggested it is one of the ancestors of *C. retusus* in China. Among the 4 populations constructed from admixture based on the lowest CV error, the population3 in central China is offspring by hybridization of the other 3 populations. However, as there is no real chromosome-level reference, some information may be incomplete. It is in urge to sequence and assembly a high quality of *C. retusus*. This is the first report regarding the genetic relationship and diversity of C. retusus in China, which can provide theoretical basis for molecular marker-assisted breeding of *C. retusus*.

Supplementary Materials:

Table S1: The data after comparing sample sequences with the reference genomic data including the mapping rate and reading depth.

Author Contributions:

H.S. and T.M.; methodology, W.Z.; software, Y.S.; validation, X.L.; formal analysis, Q.D.; investigation, J.W.; resources, W.Z.; data curation, H.S.; writing—original draft preparation, H.S.; writing—review and editing, H.S.; visualization, T.M.; supervision, T.M.; project administration, T.M.; funding acquisition, all authors have read and agreed to the published version of the manuscript.

Funding:

This research was funded by the Forestry Science and Technology Development Project, the Science and Technology Development Center of State Forestry and Grassland Administration (KJZXSA2019041), and the Science and Technology Revitalization Forestry Project of Henan Provincial Department of Forestry [2018(68)], and the Dabie Mountain Forestry Resources Innovation Theory and Technology innovation team of Xinyang Agriculture and Forestry University (XNKJTD-004), and Comprehensive reform of forestry major (ZYZHGG201805).

Data Availability Statement:

Data will be made available on request.

Acknowledgments:

The authors wish to express their sincere thanks to the Key Laboratory of the State Forestry Administration for Cultivating Forest Resources in Central China and Henan Agricultural University for their kind assistance in the experiments.

Conflicts of Interest:

The authors declare no conflict of interest.

References

Chang MC, Qiu L, Green P (1996) Oleaceae. Flora of China 15, 272-319.

Duke J, Wain K (1981) Medicinal plants of the world. Computer index with more than 85000 entries. Handbook of Medicinal Herbs CRC press Florida, Boca Raton. 96.

Huxley A (1992) New RHS dictionary of gardening 1: 203–205. acmillan ISBN 0-333-47494-5.

Lee YG, Lee H, Jung JW, Seo KH, Lee DY, Kim HG, Ko JH, Lee DS, Baek NI (2019) Flavonoids from Chionanthus retusus (Oleaceae) Flowers and Their Protective Effects against Glutamate-Induced Cell Toxicity in HT22 Cells. International journal of molecular sciences. 20(14):3517. https://doi.org/10.3390/ijms20143517

Gülçin İ, Elias R, Gepdiremen A, Taoubi K, Köksal E (2009) Antioxidant secoiridoids from fringe tree (Chionanthus virginicus L.). Wood science and technology. 43(3-4):195. https://doi.org/10.1007/s00226-008-0234-1

Kwak JH, Kang MW, Roh JH, Choi SU, Zee OP (2009) Cytotoxic phenolic compounds from Chionanthus retusus. Archives of pharmacal research. 32(12):1681-1687. https://doi.org/10.1007/s12272-009-2203-0

Kim M-S, Lee E-H, Cho Y-J (2015) Anti-oxidative Activities of Extracts from Chionanthus retusus leaves, Fruits and FLower. Current Research on Agriculture and Life Sciences. 33(2):49-56.

Hu X-l, Jiang Q, Yin F-j, Liu T (2010) Extraction and Determination of Mineral Elements and Bioactive Components in Chinese Fringe Tree Young Leaf Tea and Flower Tea [J]. Food Science. 18.

Kim KS, Lee YH, Jang YS, Choi IH (2013) Analysis of fatty acid compositions and biodiesel properties of seeds of woody oil plants in Korea. Korean Journal of Plant Resources. 26(5):628-635. https://doi.org/10.7732/kjpr.2013.26.5.628

Gao J, Yin W (2016) Flower Essential Oil Composition of Chionanthus retusus. Chemistry of Natural Compounds. 52(5):934-935. https://doi.org/10.1007/s10600-016-1823-0

Soejima A, Maki M, Ueda K (1998) Genetic variation in relic and isolated populations of Chionanthus retusus (Oleaceae) of Tsushima Island and the Tôno region, Japan. Genes & genetic systems. 73(1):29-37.

He Y, Liu L, Yang S, Dong M, Yuan W, Shang F (2017) Characterization of the complete chloroplast genome of Chinese fringetree (Chionanthus retusus). Conservation genetics resources. 9(3):431-434. https://doi.org/10.1007/s12686-017-0704-6

Wang Z, Tanksley S (1989) Restriction fragment length polymorphism in Oryza sativa L. Genome. 32(6):1113-1118. https://doi.org/10.1139/g89-563

Lanteri S, Saba E, Cadinu M, Mallica G, Baghino L, Portis E (2004) Amplified fragment length polymorphism for genetic diversity assessment in globe artichoke. Theoretical and Applied Genetics. 108(8):1534-1544. https://doi.org/10.1007/s00122-003-1576-6

Smith J, Kresovich S, Hopkins M, Mitchell S, Dean R, Woodman W, Lee M, Porter K (2000) Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. Crop Science. 40(1):226-232. https://doi.org/10.2135/cropsci2000.401226x

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2012) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Molecular methods for evolutionary genetics. Springer; p. 157-178. https://doi.org/10.1007/978-1-61779-228-1_9

Pegadaraju V, Nipper R, Hulke B, Qi L, Schultz Q (2013) De novo sequencing of sunflower genome for SNP discovery using RAD (Restriction site Associated DNA) approach. BMC genomics. 14(1):556. https://doi.org/10.1186/1471-2164-14-556

Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant molecular biology reporter. 15(1):8-15. https://doi.org/10.1007/bf02772108

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using se-

quenced RAD markers. PloS one. 3(10):e3376. https://doi.org/10.1371/journal.pone.0003376

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics. 25(14):1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics. 25(16):2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R (2011) The variant call format and VCFtools. Bioinformatics. 27(15):2156-2158. https://doi.org/10.1093/bioinformatics/btr330

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics. 81(3):559-575. https://doi.org/10.1086/519795

Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. The American Journal of Human Genetics. 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources. 10(3):564-567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, Normandeau É, Laroche J, Larose S, Jean M (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PloS one. 8(1):e54603. https://doi.org/10.1371/journal.pone.0054603

Wang Y, Ning Z, Hu Y, Chen J, Zhao R, Chen H, Ai N, Guo W, Zhang T (2015) Molecular mapping of restriction-site associated DNA markers in allotetraploid upland cotton. PLoS One. 10(4):e0124781. https://doi.org/10.1371/journal.pone.0124781

Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD, Blaxter ML (2011) Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. PloS one. 6(4):e19315. https://doi.org/10.1371/journal.pone.0019315

Berthouly-Salazar C, Mariac C, Couderc M, Pouzadoux J, Floc'h J-B, Vigouroux Y (2016) Genotyping-by-sequencing SNP identification for crops without a reference genome: using transcriptome based mapping as an alternative strategy. Frontiers in plant science. 7:777. https://doi.org/10.3389/fpls.2016.00777

Baali-Cherif D, Besnard G (2005) High genetic diversity and clonal growth in relict populations of Olea europaea subsp laperrinei (Oleaceae) from Hoggar, Algeria. ANNALS OF BOTANY,2005, 96(5): 823-830. https://doi.org/10.1093/aob/mci232

Nelson MR, Marnellos G, Kammerer S, Hoyal CR, Shi MM, Cantor CR, Braun A (2004) Large-scale validation of single nucleotide polymorphisms in gene regions. Genome research. 14(8):1664-1668. https://doi.org/10.1101/gr.2421604

Nielsen R (2005) Molecular signatures of natural selection. Annu Rev Genet. 39, 197-218. https://doi.org/10.1146/annurev.genet.39.073003.112420

Ma B, Liao L, Peng Q, Fang T, Zhou H, Korban SS, Han Y (2017) Reduced representation genome sequencing reveals patterns of genetic diversity and selection in apple. Journal of integrative plant biology. 59(3):190-204. https://doi.org/10.1111/jipb.12522

Xue X.J, Du XY, Gai Y, Tang Y, Sun YX, Song LQ, Jiang ZW (2020) Application progress of SNPs in plants based on GBS sequencing. 48, 62-68. DOI:10.15889/j.issn.1002-1302.2020.13.012

Niu SZ, Song QF, Koiwa H, et al. (2019) Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (Camellia sinensis) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. BMC Plant Biology19, 328. https://doi.org/10.1186/s12870-019-1917-5