



Article Population Genetic Diversity and Structure of an Endangered Salicaceae Species in Northeast China: *Chosenia arbutifolia* (Pall.) A. Skv.

Yu Wang ^{1,2,†}, Zhongyi Jiao ^{2,3,†}, Jiwei Zheng ^{2,3}, Jie Zhou ^{2,3}, Baosong Wang ^{2,3}, Qiang Zhuge ^{1,*} and Xudong He ^{2,3,*}

- ¹ College of Biology and the Environment, Nanjing Forestry University, Nanjing 210037, China; wangyu84yu@163.com
- ² Department of Tree Genetics and Breeding, Jiangsu Academy of Forestry, Nanjing 211153, China; tianyaguke2009@163.com (Z.J.); zjw932333@163.com (J.Z.); zjwin718@126.com (J.Z.); baosongwang1965@163.com (B.W.)
- ³ Willow Nursery of the Jiangsu Provincial Platform for Conservation and Utilization of Agricultural Germplasm, Jiangsu Academy of Forestry, Nanjing 211153, China
- Correspondence: qzhuge@njfu.edu.cn (Q.Z.); hxd_519@163.com (X.H.); Tel.: +86-25-8542-8701 (Q.Z.); +86-25-5274-4180 (X.H.)
- + These authors contributed equally to this work.

Abstract: Chosenia arbutifolia (Pall.) A. Skv. is a unique and endangered species belonging to the Salicaceae family. It has great potential for ornamental and industrial use. However, human interference has led to a decrease in and fragmentation of its natural populations in the past two decades. To effectively evaluate, utilize, and conserve available resources, the genetic diversity and population structure of C. arbutifolia were analyzed in this study. A total of 142 individuals from ten provenances were sampled and sequenced. Moderate diversity was detected among these, with a mean expected heterozygosity and Shannon's Wiener index of 0.3505 and 0.5258, respectively. The inbreeding coefficient was negative, indicating a significant excess of heterozygotes. The fixation index varied from 0.0068 to 0.3063, showing a varied genetic differentiation between populations. Analysis of molecular variance demonstrated that differentiation accounted for 82.23% of the total variation among individuals, while the remaining 17.77% variation was between populations. Furthermore, the results of population structure analysis indicated that the 142 individuals originated from three primitive groups. To provide genetic information and help design conservation and management strategies, landscape genomics analysis was performed by investigating loci associated with environmental variables. Eighteen SNP markers were associated with altitude and annual average temperature, of which five were ascribed with specific functions. In conclusion, the current study furthers the understanding of C. arbutifolia genetic architecture and provides insights for germplasm protection.

Keywords: willow; Chosenia arbutifolia; genetic diversity; population structure; SLAF-seq

1. Introduction

The Salicaceae family comprises over 300 species of trees and shrubs found across China and it is generally classified into three genera, namely *Populus, Salix,* and *Chosenia* [1]. Interestingly, the genus *Chosenia* is monotypic and contains only *Chosenia arbutifolia* (Pall.) A. Skv., which is distributed in the Greater and Lesser Khingan Mountains, the Changbai Mountains, and the montane regions of eastern Liaoning Province [1]. As an ancient tree species, *C. arbutifolia* was considered a transitional form between the divergence from *Populus* to *Salix* and was segregated into a separate genus by some botanists [2,3]. In Northeast China, *C. arbutifolia* serves primarily as landscape planting element, owing to its unique characteristics, including fast growth, frost tolerance, majestic architecture, beautiful shape, and scarlet branches. Moreover, as an important wood source, *C. arbutifolia*



Citation: Wang, Y.; Jiao, Z.; Zheng, J.; Zhou, J.; Wang, B.; Zhuge, Q.; He, X. Population Genetic Diversity and Structure of an Endangered Salicaceae Species in Northeast China: *Chosenia arbutifolia* (Pall.) A. Skv.. *Forests* 2021, *12*, 1282. https://doi.org/10.3390/f12091282

Academic Editor: Giovanbattista De Dato

Received: 1 September 2021 Accepted: 16 September 2021 Published: 18 September 2021

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



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). is also utilized for construction, furniture, and paper production [4]. The nature populations of *C. arbutifolia* have drastically decreased due to their weak regenerative capacity, unreasonable deforestation, and over-utilization, leading to the plant's recognition as a class II endangered species listed in the National Key Protected Wild Plants.

As the most important determinant of biodiversity, genetic diversity is driven by mutations, natural selection, and other factors that are fundamental for species evolution and adaption [5]. The level of genetic diversity indicates the adaptive ability to environmental changes and is a decisive factor for the long-term survival of species, especially for tree species with long life spans [6,7]. Therefore, the evaluation of genetic diversity and population structure is of major theoretical and practical significance for the management, development, utilization, and conservation of plant germplasms. Thus far, various efforts have been made to determine the genetic diversity and population structure of many common tree species, such as eucalypts [8], Chinese elm [9], oak [10], some fruit trees [11,12], and conifer species [13–15]. Furthermore, Salicaceae species, including *Populus nigra* L. [16], *Populus tomentosa* Carr. [17], *Populus deltoids* Bartr. [18], *Populus euphratica* Oliv. [19], *Populus trichocarpa* Torr. & Gray [20], *Salix viminalis* L. [21], *Salix psammophila* C. [22], *Salix purpurea* L. [23,24], and *Salix alba* L. [25], have been extensively characterized.

Despite its endangered status and environmental significance, genetic and genomic resources available for *C. arbutifolia* are scarce, with previous research having mainly focused on biological habitat, growth traits, breeding techniques, and transcriptome [26–28]. Furthermore, only one report exploring the phylogeography and population structure of *C. arbutifolia* in Japan through the analysis of chloroplast DNA markers has been published [29]. Therefore, the available genetic markers for *C. arbutifolia* are still insufficient compared to those for other Salicaceae species. In this study, a large number of single nucleotide polymorphism (SNP) markers were identified through specific locus amplified fragment sequencing (SLAF-seq). Moreover, the genetic diversity and population structure of ten *C. arbutifolia* species in Northeast China. Landscape genomics analysis was also performed to investigate the loci potentially under selection.

2. Materials and Methods

2.1. Plant Materials and DNA Extraction

A total of 142 *C. arbutifolia* plants from ten populations in Jilin Province and Heilongjiang Province in China were investigated (Figure 1). For each population, 10–15 individuals were collected, at least 50 m apart from each other (Table 1). The twigs of each plant were cut and brought back to the laboratory for water planting. Young leaves of all individuals were sampled and subjected to DNA isolation using a DNA extraction kit (DP305, Tiangen Biotech Co. Ltd., Beijing, China). The quality and concentration of isolated DNA were then assessed through electrophoresis in 0.8% agarose gels and Nanodrop (Thermo Fisher Scientific, Wilmington, DE, USA).

2.2. Librarary Construction and SLAF Sequencing

Since the whole genome of *C. arbutifolia* has not been published, the reference genome of *P. euphratica* was chosen to conduct marker discovery survey in silico by various restriction enzymes. Finally, the combination of restriction enzymes Rsal and Hae III was determined, and the genome information of *Oryza sativa* Indica (genome size 374.30 Mb, http://rapdb.dna.affra.go.jp/, accessed on 15 November 2019) was selected as a control to evaluate the quality of enzyme digestion. After digestion, a single nucleotide (A) was added to the 3'-end using dATP and ligated with dual-index adapter. The diluted restriction-ligation DNA was used as the template for PCR amplification. Then the length of restriction fragment was defined from 264 bp to 364 bp using SLAF tags. High-throughput sequencing was performed on an Illumina HiSeq 2500 platform (Illumina, Inc.; San Diago, CA, USA) at Beijing Biomarker Technologies Corporation (Beijing, China).

2.3. SLAF Tag Development and SNP Calling

After identification, filtering, inspection, and evaluation of the raw data generated via sequencing, high-quality reads were clustered based on sequence similarity, wherein different SLAF tags with over similarity above 90% among individuals were defined as polymorphic. The sequencing reads were aligned onto the reference sequence with the greatest sequencing depth of each SLAF tag using bwa software [30]. GATK [31] and Samtools [32] were used for SNP calling. Finally, the overlapping results of the two methods were considered as reliable SNPs and filtered based on integrity >0.8 and a minor allele frequency >0.5.

2.4. Genetic Diversity and Population Structure Analyses

The diversity indexes for each population including observed allele number (N_o), expected allele number (N_e), expected heterozygousity (H_e), Nei's diversity index (H), Shannon-Wiener index (I), and polymorphic information content (PIC) were determined using GenAlex 6.5 [33]. To evaluate the differentiation within or among populations, the index of inbreeding coefficient (F_{IS}), pairwise genetic differentiation index (F_{ST}), and analysis of molecular variance (AMOVA) were estimated using Arlequin 3 [34]. The software MEGA X was used for phylogenetic tree construction based on the neighbor-joining algorithm under the parameter model of Kimura 2 and 1000 bootstrap replicates. The software Admixture [35] based on maximum-likelihood method was used for population structure analysis. The number of populations (K) was set from 1 to 10 for clustering, and each individual was assigned to its respective populations according to the maximum membership probability.



Figure 1. Locations of ten C. arbutifolia (Pall.) A. Skv. populations in Northeast China.

2.5. Analysis of Environmental Adaptive Loci

The altitude data of each provenance were monitored using a portable GPS (HOLUX GR-260). The annual average temperature was accessed through the website of the National Meteorological Science Data Center (http://data.cma.cn, accessed on 20 December 2019). The Bayevn2 [36] program was employed for screening outlier loci potentially associated with the two environmental variables. A covariance matrix of allele frequencies was

calculated across populations using the full set of SNPs to avoid population-specific effects. Consequently, the Bayes Factor (BF) and nonparametric Spearman's rank correlation coefficient (ρ) were derived via the Markov Chain Monte Carlo (MCMC) method. The top 1% rank values of log10BF and ρ were adopted as thresholds to determine potential markers of environmental adaptability in this study. Furthermore, the putative function of these candidate markers sequences were identified using the Basic Local Alignment Search Tool (BLAST) in NCBI.

Number	Population	Abbreviation	Sample Size	Geographical Coordinates	Altitude (m)	Annual Average Temperature (°C)
1	Changbai, Jilin Province	СВ	15	41°28′ N, 128°13′ E	1006	3.3 (-2.8, 10.8) *
2	Manjiang, Jilin Province	MJL	15	41°55′ N, 127°37′ E	826	4.7 (-0.8, 11.5)
3	Wangqing, Jilin Province	WQ	15	43°11′ N, 129°56′ E	359	4.6 (-1.3, 11.7)
4	Hunchun, Jilin Province	HC	15	43°57′ N, 130°53′ E	233	6.5 (1.6, 12.3)
5	Tangwanghe, Heilongjiang Province	TWH	10	48°38′ N, 129°52′ E	246	1.6 (-4.7, 8.6)
6	Xiling, Heilongjiang Province	XLL	15	47°24′ N, 128°42′ E	401	2.1 (-3.2, 8.4)
7	Langxiang, Heilongjiang Province	LX	13	46°42′ N, 129°10′ E	280	2.3 (-3.4, 8.5)
8	Huzhong, Heilongjiang Province	HZ	14	52°03′ N, 123°34′ E	631	-3.6 (-11.1, 5.2)
9	Jiagedaqi, Heilongjiang Province	JGDQ	15	50°18′ N, 124°50′ E	360	-0.3 (-6.8, 7.2)
10	Xinlin, Heilongjiang Province	XL	15	51°34′ N, 124°21′ E	523	-2.3 (-9.2, 5.8)
Total			142			

Table 1. Details of provenances.

* Values in brackets indicate the average of the minimum and maximum temperatures over the years.

3. Results

3.1. SLAF Tags and SNP Calling

A total of 424.88 Mb reads were generated via high-throughput SLAF-seq of the 142 *C. arbutifolia* individuals, with a mean Q_{30} value of 96.07% and an average GC content of 42.47% (Table 2). The most SLAF tags were obtained from the TWH provenance (mean 157,987.50 for each individual) and the least were found for the WQ provenance (mean 101,985.60 for each individual). The sequencing depth for provenances ranged from 27.91× (CB provenance) to $16.28 \times$ (TWH provenance), with an average depth of $23.11 \times$. In total, 1,188,126 SLAF tags were obtained, of which 211,848 were polymorphic and utilized for SNP calling. After mapping to the reference genome, 1,164,285 SNPs were screened. The number of SNPs for each provenance varied from 313,515.77 (LX) to 223,295.27 (WQ), with an average of 264,633.74 (Table 2). Moreover, there were slight differences in the heter ratio (mean 8.82) and integrity (mean 22.72) of SNPs among each provenance. Finally, a total of 105,857 SNPs were selected for the evaluation of *C. arbutifolia* populations.

3.2. Genetic Diversity and Genetic Differentiation

As shown in Table 3, all the N_o were the same (N_o = 2), and N_e was also quite similar (mean 1.5964) between provenances. The H_e varied from 0.3559 (WQ) to 0.3428 (CB), with an average value of 0.3505. The values of H and I ranged from 0.3708 (HC) to 0.3570 (CB) and 0.5324 (WQ) to 0.5158 (CB), with averages of 0.3661 and 0.5258, respectively. Slight differences between PIC values (mean 0.2810) were detected across the ten populations. However, all F_{IS} values were negative, with the highest value determined for the MJL provenance and the lowest for the CB provenance.

Provenances	Q ₃₀ (%)	GC Content (%)	No. of SLAFs	Sequencing Depth	No. of SNPs	Heter Ratio (%)	Integrity of SNP (%)
СВ	96.14	43.26	113,129.67	27.91	245,420.53	7.50	21.08
MJL	96.03	44.28	114,436.33	25.22	236,671.40	7.01	20.32
WQ	96.10	41.37	101,985.60	23.68	223,295.27	8.73	19.17
HC	96.17	42.91	112,710.00	23.55	239,874.33	9.17	20.60
TWH	95.76	42.51	157,987.50	16.28	304,336.10	9.95	26.13
XLL	95.89	41.12	129,231.53	20.78	273,826.33	9.02	23.51
LX	95.87	42.72	154,696.54	21.02	313,515.77	9.53	26.92
HZ	96.25	42.55	134,427.43	23.36	275,946.50	9.22	23.70
JGDQ	96.27	41.72	115,460.87	23.18	252,437.87	8.63	21.68
XL	96.25	42.21	134,513.27	26.11	281,013.27	9.43	24.13
Mean	96.07	42.47	126,857.87	23.11	264,633.74	8.82	22.72

Table 2. Statistics of SLAF tags and SNP markers for different provenances of Chosenia arbutifolia (Pall.) A. Skv.

Table 3. Genetic diversity of the ten Chosenia arbutifolia populations.

Population	No	Ne	H _e	Н	Ι	PIC	F _{IS}
СВ	2	1.5838	0.3428	0.3570	0.5158	0.2750	-0.2329
MJL	2	1.5832	0.3439	0.3582	0.5177	0.2763	-0.0033
WQ	2	1.6079	0.3559	0.3706	0.5324	0.2849	-0.1145
HC	2	1.6072	0.3557	0.3708	0.5323	0.2848	-0.1382
TWH	2	1.5833	0.3432	0.3662	0.5162	0.2755	-0.0223
XLL	2	1.6035	0.3544	0.3686	0.5308	0.2840	-0.1134
LX	2	1.6020	0.3535	0.3701	0.5297	0.2833	-0.1637
HZ	2	1.5949	0.3506	0.3664	0.5264	0.2815	-0.1021
JGDQ	2	1.5947	0.3502	0.3645	0.5257	0.2810	-0.1098
XL	2	1.6040	0.3546	0.3690	0.5310	0.2841	-0.1391
Mean	2	1.5964	0.3505	0.3661	0.5258	0.2810	-0.1112

 N_o , observed allele number; N_e , expected allele number; H_e , expected heterozygosity; H, Nei's diversity index; I, Shannon-Wiener index; PIC, polymorphism information content; F_{IS} , inter-individual fixation index.

As a measurement of genetic differentiation across populations, the F_{ST} presented in Table 4 indicated significant variation between different *C. arbutifolia* provenances. The highest F_{ST} value was detected between the XL and CB provenances (0.3063), followed by that between the JGDQ and CB provenances (0.3005). The lowest F_{ST} value was observed between XL and HZ provenances (0.0068). As described in Table 5, AMOVA revealed that the differentiation among individuals accounted for 82.23% of the total variation, while the remaining 17.77% was variation among populations.

Table 4. Pairwise differentiation (FST) among provenances of Chosenia arbutifolia.

	СВ	HZ	HC	JGDQ	LX	MJL	TWH	WQ	XLL
HZ	0.2993								
HC	0.2410	0.1317							
JGDQ	0.3005	0.0206	0.1351						
LX	0.2752	0.1288	0.1202	0.1356					
MJL	0.2106	0.2476	0.1904	0.2530	0.2181				
TWH	0.2944	0.1135	0.1311	0.1176	0.1060	0.2450			
WQ	0.2491	0.1551	0.0927	0.1608	0.1461	0.2086	0.1577		
XLL	0.2839	0.1246	0.1260	0.1271	0.0916	0.2280	0.0953	0.1470	
XL	0.3063	0.0068	0.1439	0.0265	0.1428	0.2555	0.0121	0.0166	0.1344

3.3. Popalution Structure and Phylogenetic Relationships

As demonstrated in Figure 2a, the optimum K value was 3, which indicated that all 142 *C. arbutifolia* individuals originated from three primitive groups. Furthermore, as shown in Figure 2b, group 1 (green pool) included 69 individuals (48.59%), which were

from those in the LX (13 individuals), XLL (15 individuals), HC (15 individuals), WQ (15 individuals), TWH (10 individuals), and HZ (1 individual) provenances. Group 2 (yellow pool) consisted of 43 individuals (30.28%), which were mainly from the JGDQ (15 individuals), XL (15 individuals), HZ (13 individuals) provenances. The remaining 30 individuals (21.13%) originating from the CB (15 individuals) and MJL (15 individuals) provenances were also merged into group 3 (gray pool).

Table 5. Analysis of molecular variance of the ten Chosenia arbutifolia provenances.

Source of Variation	df	Sum of Squares	Variance Components	Percentage of Variaion (%)
Among Populations	9	48,691.24	167.0238	17.77
Between Individuals	274	212,784.1	772.9311	82.23
Total	283	261,475.3	939.9548	100



Figure 2. Rational group number and population structure of *Chosenia arbutifolia*. (**a**) The optimal K value estimated via Admixture; (**b**) Population structure for the 142 *Chosenia arbutifolia* individuals. The green, yellow, and gray colors represent group 1, group 2, and group 3 when K = 3, respectively.

An NJ tree was constructed for the 142 *C. arbutifolia* individuals (Figure 3a). Interestingly, except for plants from the CB population, individuals originating from the same population could not be clustered into a single clade. A dendrogram based on the ten populations was also established (Figure 3b). Similar to Figure 3b, the CB population had a relatively distant genetic relationship with the other nine populations. In general, populations located closely based on geographical distance were also divided into the same subclade. For instance, WQ and HC were both in Jilin province, TWH and LX were both in the lesser Khingan Mountains area, and JGDQ with XL were both in the greater Khingan Mountains area (Figure 1).

3.4. Association Analysis of Environmental Variables

In total, 18 SNP markers associated with different environmental variables were identified, of which seven were closely related to altitude, while the other 11 were significantly correlated with the annual average temperature (Table 6). Of these 18 marker sequences, eight had successful hits against the NCBI database, of which three were related with altitude, including Dicer-like (DCL) protein, mitogen-activated protein kinase (MAPK), and prefoldin subunit 5 (Table 7). The other three markers were related to annual average temperature, including the uncharacterized LOC118052887, sulfite oxidase, and hippocampus abundant transcript-like protein (Table 7). The last two sequences (marker54369 and marker67792) had no function, but were quiet similar to the sequences of *P. trichocarpa*. All the annotated putative genes originated from *Populus alba* L. and *P. euphratica*.



Figure 3. Phylogenetic tree of *Chosenia arbutifolia*. (a) The dendrogram was constructed at the level of individuals and (b) based on populations.

SNP ID	Position	Log ₁₀ (BF)	ρ	Altitude	Annual Average Temperature
Marker29376	240	4.2178	0.2145	*	
Marker33590	55	2.4780	0.1907	*	
Marker36276	81	0.5129	0.1925	*	
Marker36423	243	1.2850	0.2411	*	
Marker46246	255	0.5878	0.2333	*	
Marker54369	62	2.1703	0.2289	*	
Marker63252	254	0.4395	0.2018	*	
Marker18816	178	1.0925	0.1576		*
Marker25405	146	3.2490	0.1490		*
Marker28698	57	1.2597	0.1633		*
Marker30836	222	1.5497	0.1428		*
Marker33966	46	1.4143	0.1576		*
Marker34330	104	1.3260	0.1715		*
Marker36320	196	1.2136	0.1797		*
Marker44043	121	238.4900	0.1517		*
Marker47618	212	1.8375	0.1618		*
Marker57557	26	1.3072	0.1611		*
Marker67792	153	4.0701	0.1518		*

Table 6. Summary of candidate SNP markers associated with two different environmental variables.

* The SNP showed an association with the specific environmental variables.

Table 7. Sequence alignment of the candidate SNP markers for functional identification.

Variables	Marker ID	Putative Genes	Origin
Altitude	Marker29376	Dicer-like (DCL) Protein	P. alba L.
	Marker46246	Mitogen-activated Protein Kinase	P. euphratica Oliv.
	Marker63252	Prefoldin Subunit 5	P. alba
Annual Average Temperature	Marker33966	Uncharacterized LOC118052887	P. alba
	Marker36320	Sulfite Oxidase	P. alba
	Marker47618	Hippocampus Abundant Transcript-like Protein	P. euphratica

4. Discussion

As the most important determinant of biodiversity, genetic diversity can be studied at different levels, including population, individual, tissue and organ, or the molecular level. Highly variable genetic markers, such as SSRs and SNPs, facilitate the evaluation of genetic diversity within and between populations. However, population size is a critical factor that can have a significant influence on the accuracy of genetic parameters [37]. Due to unreasonable deforestation, the distribution range of natural C. arbutifolia in China is being continuously reduced, thus decreasing the number of individual plants available for collection, which in turn accounts for the small sample size in the current study (Table 1). In general, the larger sample size, the more rare alleles will be detected. For some tree species, previous reports have suggested that a sample size of 25–29 [38] or less than 30 [39] per population should be sufficient. In contrast, numerous lines of evidence have demonstrated that the parameters of genetic diversity can be accurately estimated through the study of small populations with less than 10 individuals, as previously reported for *Quercus* susber L. [40] and Eucalyptus occidentalis Endl. [41]. Taken together, due to species-specific biological characteristics and different sampling strategies, there is no definitive conclusion regarding the optimal population size [42].

H_e is commonly considered an essential measure for genetic diversity, and the larger it is, the lower the population consistency, and the higher the genetic diversity. In the present study, the H_e value ranged from 0.3559 to 0.3428, with an average value of 0.3505, which is significantly smaller than that of *Salix arbutifolia* Pall. in Japan (mean $H_e = 0.6026$) [29]. Moreover, the current H_e value was also lower than that obtained for *Salix* species, such as S. viminalis (mean $H_e = 0.616$) [21], S. psammophila (mean $H_e = 0.689$) [22], S. purpurea $(\text{mean H}_e = 0.7365)$ [43], Salix eriocephala Michx. (mean H $_e = 0.6857$) [43], as well as Populus species, such as *P. tomentosa* (mean $H_e = 0.446$) [17], *P. deltoids* (mean $H_e = 0.487$) [18] and *Populus simonii* Carr. (mean $H_e = 0.677$) [44]. However, the H_e values in all of these reports were calculated using SSR markers, which are extremely variable and greatly dispersed across the genome. Using the same SNP markers, the He value estimated herein was much higher than that previously reported for S. purpurea (mean $H_e = 0.2301$) [23], while comparable to those for *Ulmus parvifolia* Jacq. (mean $H_e = 0.3315$) [9] and *Elaeis guineensis* Jacq. (mean H_e from 0.29 to 0.33) [45]. Furthermore, the I and PIC value measured based on SNP markers in this study (mean I = 0.5258, mean PIC = 0.2810) were considerably lower than those determined for *S. psammophila* (mean I = 1.345, mean PIC = 0.714) [22] and *P. tomentosa* (mean I = 0.80, mean PIC = 0.385) [17] with SSR markers, yet consistent with values reported for U. parvifolia (mean I = 0.5041, mean PIC = 0.2686) [9] and E. guineensis (PIC from 0.23 to 0.269) [45] based on the same SNP markers. Therefore, we believe that high values of genetic parameters, such as effective alleles, heterozygosity, and Shannon-Wiener index, are accompanied by an increase in the allelic variation of molecular markers, indicating that the type and efficiency of the marker will have tremendous influence on the estimation of genetic parameters. As a result, the authentic features of population genetic diversity might be most accurately revealed through the selection of more polymorphic markers.

Changes in population structure, which are mainly affected by genetic mutation, mating system, genetic drift, and selection, could be reflected by F statistics, which are ordinarily measured via F_{IS} , the total inbreeding coefficient of population (F_{IT}), and F_{ST} . In this study, all the F_{IS} values were negative, which is consistent with the report on *P. tomentosa* [17], indicating that a significant excess of heterozygotes exists among these populations across the entire natural distribution area in Northeast China. Similarly, negative F_{IS} values were observed in partial populations of S. *purpurea* [23] and S. *arbutifolia* [29], which is more likely due to interspecific hybridization or hybridization of inbred cultivars. In contrast, some Japanese populations of *S. arbutifolia* showed positive values of F_{IS} , suggesting a deficiency of heterozygosity which might lead to bi-parental inbreeding [29]. It is plausible that the population structure is affected by the geographical disjunction and habitat quality of these populations in Japan. However, for willows, the vegetative reproductive strategy is another critical factor contributing to an excess of homozygotes,

as observed for S. purpurea and S. eriocephala [43]. Moreover, the null alleles generated using SSR markers may also be attributable to the positive value of F_{IS} [44]. The overall F_{ST} value in our study (from 0.0068 to 0.3063) indicated significant differentiation among provenances, which is comparable with data for S. purpurea (from 0.064 to 0.420) [23] and S. arbutifolia (from 0.01 to 0.41) [29], but much higher than that reported for S. viminalis (from 0.040 to 0.119) [21], S. psammophila (from 0.008 to 0.016) [22], S. alba (F_{ST} = 0.07) [25], and several *Populus* species [44]. As described by Wright [46], $F_{ST} > 0.25$ represents an enormous genetic differentiation, $0.25 > F_{ST} > 0.15$ indicates a comparatively large genetic differentiation, $0.15 > F_{ST} > 0.05$ shows a moderate genetic differentiation, and $F_{ST} < 0.05$ is negligible. According to the criteria, large genetic differentiation was found for eight provenance combinations, among which were HZ and CB, as well as JGDQ and CB (Table 4). Of the eight combinations, all six provenances from the Heilongjiang province were related to CB, which is the farthest geographic provenance in Jilin province (Figure 1). Such highlevel differentiation was also observed in the phylogenetic tree (Figure 3a). The lowest F_{ST} value was detected between XL and HZ (0.0068), which belongs to the same administrative division as Jiagedaqi. Similar results were also observed on S. purpurea [23]. This suggests that gene flow was impeded by pollen competition among the local male individuals. Therefore, all of these evidences have proven that the genetic differentiation of C. arbutifolia in China is significantly associated with geographic distribution.

Natural selection drives species adaption to the environment through genetic changes between generations [47]. The variation in gene frequency among populations facilitates phenotypic differentiation [48]. Thus, the identification of genes or genomic loci associated with natural selection is of major relevance for protecting germplasm resources. In our study, out of the 105,857 efficient SNP markers, only 18 were related to environmental variables, of which five had specific functions, and one was uncharacterized. The Dicer protein was initially described in animals. Its ortholog in plants is the DCL protein. Research involved into the function of DCL focuses on the presence and function of siRNAs, which are major regulators of growth, development, and resistance to biotic or abiotic stress [49]. Similarly, as a prominent signaling molecule, mitogen-activated protein kinase (MAPK) participates in the transduction of developmental and environmental signals into programmed and adaptive responses in plants, thereby regulating gene expression [50]. While sulfite oxidase genes were previously cloned in Arabidopsis thaliana, Solanum tuberosum, and Populus, functional research has mostly concentrated on A. thaliana, where sulfite oxidase could protect plants against sulfur dioxide [51]. There are few reports on the function of prefoldin subunit 5 and hippocampus abundant transcript-like protein in plants, which should be addressed in further studies.

5. Conclusions

To our knowledge, this is the first population genomics study of *C. arbutifolia* to compare individuals from natural populations in Northeast China. The greater marker density determined using SLAF-seq proved to be a reliable marker system for evaluation of genetic relatedness and population structure, which can be applied in association analysis. All the populations showed moderate diversity. However, enormous genetic differentiations were also detected between provenances, mainly caused by geographical isolation. Based on the AMOVA analysis, the genetic variation mainly came from individuals, so that representative individuals could be selected to enhance the selection efficiency of hybrid parents for breeding. Due to the continuous reduction of natural population, a core germplasm resource bank should be established for the conservation of *C. arbutifolia*. Given the limited sample size in this study, it would be worthwhile to make further collections and expand the geographic coverage. In addition, more environmental variables should be considered in future research, including humidity, rainfall, and soil components.

Author Contributions: X.H. and Q.Z. conceived and designed the experiments; Y.W. and Z.J. performed the experiments and wrote the paper; J.Z. (Jiwei Zheng) and J.Z. (Jie Zhou) analyzed the data; Z.J., J.Z. (Jiwei Zheng) and B.W. collected the samples. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 31670662), and Independent Scientific Research Project of Jiangsu Academy of Forestry (Grant No. 2018022-1).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to Hongxue Wang in Yichun Academy of Forestry and Jun Ren in Jilin Academy of Forestry for their assistance with sample collection.

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

References

- 1. Wang, Z.; Fang, C.F. Salicaceae. In Flora Reipublicae Popularis Sinicae; Science Press: Beijing, China, 1984; pp. 79–81.
- 2. Nakai, T. Chosenia, a new genus of Salicaceae. Bot. Mag. 1920, 34, 66–69. [CrossRef]
- 3. Skvortsov, A.K. Willows of Russia and adjacent countries. Taxonomical and geographical revision (English translation with additions). *Univ. Joensuu Fac. Math. Nat. Sci. Rep. Ser.* **1999**, *39*, 1–307.
- 4. Tu, Z.Y. Breeding and Cultivation of Salix; Jiangsu Science and Technology Press: Nanjing, China, 1982; pp. 154–196.
- Eding, H.; Crooijmans, R.P.; Groenen, M.A.; Meuwissen, T.H. Assessing the contribution of breeds to genetic diversity in conservation schemes. *Genet. Sel. Evol.* 2002, 34, 613. [CrossRef] [PubMed]
- Hughes, A.R.; Inouye, B.D.; Johnson, M.T.J.; Underwood, N.; Vellend, M. Ecological consequences of genetic diversity. *Ecol. Lett.* 2008, 11, 609–623. [CrossRef]
- Hamrick, J.L.; Godt, M.J.W.; Sherman-Broyles, S.L. Factors influencing levels of genetic diversity in woody plant species. *New For.* 1992, 6, 95–124. [CrossRef]
- Song, Z.J.; Zhang, M.M.; Li, F.G.; Weng, Q.J.; Zhou, C.P.; Li, M.; Li, J.; Huang, H.H.; Mo, X.Y.; Gan, S.M. Genome scans for divergent selection in natural populations of the widespread hardwood species *Eucalyptus grandis* (Myrtaceae) using microsatellites. *Sci. Rep.* 2016, *6*, 34941. [CrossRef] [PubMed]
- 9. Lyu, Y.Z.; Dong, X.Y.; Huang, L.B.; Zheng, J.W.; He, X.D.; Sun, H.N.; Jiang, Z.P. SLAF-seq Uncovers the genetic diversity and adaptation of Chinese Elm (*Ulmus parvifolia*) in Eastern China. *J. For.* **2020**, *11*, 80. [CrossRef]
- 10. Shi, X.M.; Wen, Q.; Cao, M.; Guo, X.; Xu, L.A. Genetic diversity and structure of natural *Quercus variabilis* population in China as revealed by microsatellites markers. *J. For.* **2017**, *8*, 495. [CrossRef]
- Gao, Y.; Liu, F.Z.; Wang, K.; Wang, D.J.; Gong, X.; Liu, L.J.; Richards, C.M.; Henk, A.D.; Volk, G.M. Genetic diversity of *Malus cultivars* and wild relatives in the Chinese National Repository of Apple Germplasm Resources. *Tree Genet. Genomes* 2015, *11*, 106. [CrossRef]
- 12. Cao, K.; Wang, L.R.; Zhu, G.R.; Fang, W.C.; Chen, C.W.; Luo, J. Genetic diversity, linkage disequilibrium, and association mapping analyses of peach (*Prunus persica*) landraces in China. *Tree Genet. Genomes* **2012**, *8*, 975–990. [CrossRef]
- Peláez, P.; Ortiz-Martínez, A.; Figueroa-Corona, L.; Montes, J.R.; Gernandt, D.S. Population structure, diversifying selection, and local adaptation in *Pinus patula*. Am. J. Bot. 2020, 107, 1555–1566. [CrossRef] [PubMed]
- 14. Zhang, D.Q.; Zhou, N. Genetic diversity and population structure of the endangered conifer *Taxus wallichiana* var. *mairei* (Taxaceae) revealed by Simple Sequence Repeat (SSR) markers. *Biochem. Syst. Ecol.* **2013**, *49*, 107–114. [CrossRef]
- 15. Neophytou, C.; van Loo, M.; Hasenauer, H. Genetic diversity in introduced Douglas-fir and its natural regeneration in central Europe. *For. Int. J. For. Res.* **2020**, *93*, 535–544. [CrossRef]
- Smulders, M.J.M.; Cottrell, J.E.; Lefèvre, F.; van der Schoot, J.; Arens, P.; Vosman, B.; Tabbener, H.E.; Grassi, F.; Fossati, T.; Castiglione, S.; et al. Structure of the genetic diversity in black poplar (*Populus nigra* L.) populations across European river systems: Consequences for conservation and restoration. *For. Ecol. Manag.* 2008, 255, 1388–1399. [CrossRef]
- 17. Du, Q.Z.; Wang, B.W.; Wei, Z.Z.; Zhang, D.Q.; Li, B.L. Genetic diversity and population structure of Chinese white poplar (*Populus tomentosa*) revealed by SSR markers. *J. Hered.* **2012**, *103*, 853–862. [CrossRef]
- 18. Chen, C.; Chu, Y.G.; Ding, C.J.; Su, X.H.; Huang, Q.J. Genetic diversity and population structure of black cottonwood (*Populus deltoides*) revealed using simple sequence repeat markers. *BMC Genet.* **2020**, *21*, 2. [CrossRef]
- 19. Wang, J.; Li, Z.J.; Guo, Q.H.; Ren, G.P.; Wu, Y.X. Genetic variation within and between populations of a desert poplar (*Populus euphratica*) revealed by SSR markers. *Ann. For. Sci.* **2011**, *68*, 1143–1149. [CrossRef]
- 20. Blumstein, M.; Richardson, A.; Weston, D.; Zhang, J.; Muchero, W.; Hopkins, R. A new perspective on ecological prediction reveals limits to climate adaptation in a temperate tree species. *Curr. Biol.* **2020**, *30*, 1447–1453.e4. [CrossRef] [PubMed]
- Berlin, S.; Trybush, S.O.; Fogelqvist, J.; Gyllenstrand, N.; Hallingbäck, H.R.; Åhman, I.; Nordh, N.-E.; Shield, I.; Powers, S.J.; Weih, M.; et al. Genetic diversity, population structure and phenotypic variation in European *Salix viminalis* L. (Salicaceae). *Tree Genet. Genomes* 2014, 10, 1595–1610. [CrossRef]

- Jia, H.X.; Yang, H.F.; Sun, P.; Li, J.B.; Zhang, J.; Guo, Y.H.; Han, X.J.; Zhang, G.S.; Lu, M.Z.; Hu, J.J. *De novo* transcriptome assembly, development of EST-SSR markers and population genetic analyses for the desert biomass willow, *Salix psammophila. Sci. Rep.* 2016, *6*, 39591. [CrossRef]
- 23. Gouker, F.E.; DiFazio, S.P.; Bubner, B.; Zander, M.; Smart, L.B. Genetic diversity and population structure of native, naturalized, and cultivated *Salix purpurea*. *Tree Genet. Genomes* **2019**, *15*, 47. [CrossRef]
- 24. Sulima, P.; Prinz, K.; Przyborowski, J. Genetic diversity and genetic relationships of purple willow (*Salix purpurea* L.) from natural locations. *Int. J. Mol. Sci.* 2017, *19*, 105. [CrossRef] [PubMed]
- 25. Degirmenci, F.O.; Acar, P.; Kaya, Z. Consequences of habitat fragmentation on genetic diversity and structure of *Salix alba* L. populations in two major river systems of Turkey. *Tree Genet. Genomes* **2019**, *15*, 59. [CrossRef]
- 26. Moskalyuk, T.A. *Chosenia Arbutifolia* (Salicaceae): Life strategies and introduction perspectives. *Sib. J. For. Sci.* **2016**, *3*, 34–45. [CrossRef]
- 27. Kadis, I. Chosenia: An amazing tree of Northeast Asia. Arnoldia 2005, 63, 8-18.
- Rao, G.D.; Zeng, Y.F.; Sui, J.K.; Zhang, J.G. *De novo* transcriptome analysis reveals tissue-specific differences in gene expression in Salix arbutifolia. Trees 2016, 30, 1647–1655. [CrossRef]
- Nagamitsu, T.; Hoshikawa, T.; Kawahara, T.; Barkalov, V.Y.; Sabirov, R.N. Phylogeography and genetic structure of disjunct *Salix* arbutifolia populations in Japan. *Popul. Ecol.* 2014, 56, 539–549. [CrossRef]
- 30. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *J. Bioinform.* **2009**, 25, 1754–1760. [CrossRef]
- McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010, 20, 1297–1303. [CrossRef]
- Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/Map format and SAMtools. J. Bioinform. 2009, 25, 2078–2079. [CrossRef]
- 33. Peakall, R.; Smouse, P.E. genalex 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295. [CrossRef]
- 34. Excoffier, L.; Laval, G.; Schneider, S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform.* **2005**, *1*, 117693430500100. [CrossRef]
- 35. Alexander, D.H.; Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform*. **2011**, *12*, 246. [CrossRef]
- 36. Günther, T.; Coop, G. Robust identification of local adaptation from allele frequencies. Genetics 2013, 195, 205–220. [CrossRef]
- 37. Zhao, R.; Cheng, Z.; Lu, W.F.; Lu, B.R. Estimating genetic diversity and sampling strategy for a wild soybean (*Glycine soja*) population based on different molecular markers. *Chin. Sci. Bull.* **2006**, *51*, 1219–1227. [CrossRef]
- Zhang, S.S.; Yang, W.Z.; Zhang, Y.P.; Gao, X.S.; Kang, H.M. Effects of *Poncirus polyandra* sample size on genetic diversity index by SSR markers. J. Northeast For. Uni. 2017, 45, 35–39, 44. [CrossRef]
- Miyamoto, N.; Fernández-Manjarrés, J.F.; Morand-prieur, M.-E.; Bertolino, P.; Frascaria-Lacoste, N. What sampling is needed for reliable estimations of genetic diversity in *Fraxinus excelsior* L. (Oleaceae)? *Ann. For. Sci.* 2008, 65, 403. [CrossRef]
- Laakili, A.; Belkadi, B.; Medraoui, L.; Alami, M.; Yatrib, C.; Pakhrou, O.; Makhloufi, M.; El Antry, S.; Laamarti, A.; Filali-Maltouf, A. Diversity and spatial genetic structure of natural Moroccan *Quercus suber* L. assessed by ISSR markers for conservation. *Physiol. Mol. Biol. Plants* 2018, 24, 643–654. [CrossRef]
- 41. Elliott, C.P.; Byrne, M. Genetic diversity within and between natural populations of *Eucalyptus occidentalis* (Myrtaceae). *Silvae Genet.* **2003**, *52*, 169–173.
- Singh, M.; Chabane, K.; Valkoun, J.; Blake, T. Optimum sample size for estimating gene diversity in wild wheat using AFLP markers. *Genet. Resour. Crop Evol.* 2006, 53, 23–33. [CrossRef]
- 43. Lin, J.; Gibbs, J.P.; Smart, L.B. Population genetic structure of native versus naturalized sympatric shrub willows (*Salix*; Salicaceae). *Am. J. Bot.* **2009**, *96*, 771–785. [CrossRef]
- 44. Wei, Z.Z.; Du, Q.Z.; Zhang, J.F.; Li, B.L.; Zhang, D.Q. Genetic diversity and population structure in Chinese indigenous poplar (*Populus simonii*) populations using microsatellite markers. *Plant Mol. Biol. Rep.* **2013**, *31*, 620–632. [CrossRef]
- 45. Xia, W.; Luo, T.t.; Zhang, W.; Mason, A.S.; Huang, D.Y.; Huang, X.L.; Tang, W.Q.; Dou, Y.J.; Zhang, C.Y.; Xiao, Y. Development of high-density SNP markers and their application in evaluating genetic diversity and population structure in *Elaeis guineensis*. *Front. Plant Sci.* **2019**, *10*, 130. [CrossRef] [PubMed]
- 46. Wright, S. The interpretation of population structure by F-Statistics with special regard to systems of mating. *Evolution* **1965**, *19*, 395–420. [CrossRef]
- 47. Jump, A.S.; Hunt, J.M.; Martinez-Izquierdo, J.A.; Penuelas, J. Natural selection and climate change: Temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Mol. Ecol.* **2006**, *15*, 3469–3480. [CrossRef]
- 48. Savolainen, O.; Pyhäjärvi, T.; Knürr, T. Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Evol. Syst.* **2007**, *38*, 595–619. [CrossRef]
- 49. Margis, R.; Fusaro, A.F.; Smith, N.A.; Curtin, S.J.; Watson, J.M.; Finnegan, E.J.; Waterhouse, P.M. The evolution and diversification of Dicers in plants. *FEBS Lett.* **2006**, *580*, 2442–2450. [CrossRef]

- 50. Cristina, M.; Petersen, M.; Mundy, J. Mitogen-Activated protein kinase signaling in plants. *Annu. Rev. Plant Biol.* **2010**, *61*, 621–649. [CrossRef] [PubMed]
- 51. Brychkova, G.; Xia, Z.; Yang, G.; Yesbergenova, Z.; Zhang, Z.; Davydov, O.; Fluhr, R.; Sagi, M. Sulfite oxidase protects plants against sulfur dioxide toxicity. *Plant J.* 2007, *50*, 696–709. [CrossRef] [PubMed]