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; tianyagyke2009@163.com (Z.J.); zjw932338@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.: +86-25-85428701 (Q.Z.); +86-25-52744180 (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* 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* populations were evaluated in order to determine the evolutionary history of *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).
Figure 1. Locations of ten C. arbutifolia (Pall.) A. Skv. populations in Northeast China.

| Number | Population                      | Abbreviation | Sample Size | Geographical Coordinates | Altitude (m) | Annual Average Temperature (°C) |
|--------|---------------------------------|--------------|-------------|--------------------------|--------------|-------------------------------|
| 1      | Changbai, Jilin Province        | CB           | 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         |                          |              |                               |

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

2.2. Library 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 Diego, 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 (No), expected allele number (Ne), expected heterozygosity (He), 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 (Fis), pairwise genetic differentiation index (Fst), 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.

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.

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 Q30 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× (TWH provenance), with an average depth of 23.11×. 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.

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

| Provenances | Q30 (%) | GC Content (%) | No. of SLAFs | Sequencing Depth | No. of SNPs | Heter Ratio (%) | Integrity of SNP (%) |
|-------------|---------|----------------|-------------|-----------------|-------------|-----------------|---------------------|
| CB          | 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           | 223,925.27  | 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               |

#### 3.2. Genetic Diversity and Genetic Differentiation

As shown in Table 3, all the No were the same (No = 2), and Ne was also quite similar (mean 1.5964) between provenances. The He 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 FIS values were negative, with the highest value determined for the MJL provenance and the lowest for the CB provenance.

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

| Population | No | Ne  | He  | H   | I   | PIC  | Fis   |
|------------|----|-----|-----|-----|-----|------|-------|
| CB         | 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 |

No, observed allele number; Ne, expected allele number; He, expected heterozygosity; H, Nei's diversity index; I, Shannon-Wiener index; PIC, polymorphism information content; Fis, inter-individual fixation index.

As a measurement of genetic differentiation across populations, the Fst presented in Table 4 indicated significant variation between different C. arbutifolia provenances. The highest Fst value was detected between the XL and CB provenances (0.3063), followed by that between the JGDQ and CB provenances (0.3005). The lowest Fst 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*.

|     | CB  | 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 |

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

| Source of Variation | df | Sum of Squares | Variance Components | Percentage of Variation (%) |
|--------------------|----|----------------|---------------------|----------------------------|
| 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                        |

3.3. Population 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).

![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*](image-url)
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).

![Figure 3](image_url)

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

### 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 (marker 54369 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. eurphratica*.

| SNP ID   | Position | Log10(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    | *        |                            |

Table 6. Summary of candidate SNP markers associated with two different 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*<sub>e</sub> 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*<sub>e</sub> 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*<sub>e</sub> = 0.6026) [29]. Moreover, the current *H*<sub>e</sub> value was also lower than that obtained for *Salix* species, such as *S. viminalis* (mean *H*<sub>e</sub> = 0.616) [21], *S. psammophila* (mean *H*<sub>e</sub> = 0.689) [22], *S. purpurea* (mean *H*<sub>e</sub> = 0.7365) [43], *Salix eriocephala* Michx. (mean *H*<sub>e</sub> = 0.6857) [43], as well as *Populus* species, such as *P. tomentosa* (mean *H*<sub>e</sub> = 0.446) [17], *P. deltoids* (mean *H*<sub>e</sub> = 0.487) [18] and *Populus simonii* Carr. (mean *H*<sub>e</sub> = 0.677) [44]. However, the *H*<sub>e</sub> 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 *H*<sub>e</sub> 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 high-level 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.
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.
5. 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, doi:10.1186/1297-9686-34-5-613.
6. 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, doi:10.1111/j.1461-0248.2008.01179.x.
7. 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, doi:10.1007/BF00120641.
22. [36x373]22.
21. [36x246]27.
12. [36x684]10.
9. [36x84]34.
33.
8.
7.
6.
5.
4.
3.
2.
1.

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.
assembly, development of EST-SSR markers and population genetic analyses for the desert biomass willow, *Populus*
Chen, C.; Chu, Y.G.; Ding, C.J.; Su, X.H.; Huang, Q.J. Genetic diversity and population structure of black cottonwood (*Populus tomentosa*).

Blumstein, M.; Richardson, A.; Weston, D.; Zhang, J.; Muchero, W.; Hopkins, R. A new perspective on ecological prediction systems: Consequences for conservation and restoration. *For. Ecol. Manag.* 2008, 255, 1388–1399, doi:10.1016/j.foreco.2007.10.063.

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, doi:10.1039/jhered/ess061.

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. *RMC Genet.* 2020, 21, 2, doi:10.1186/s12863-019-0805-1.

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. *J. Hered.* 2010, 101, 1388–1399, doi:10.1016/j.jhered.2010.02.001.

Kadis, I. (*Chosenia Arbutifolia* (Salicaceae)): Life strategies and introduction perspectives. *Sib. J. For. Sci.* 2016, 3, 34–45, doi:10.15372/SJFS20160304.

Kadis, I. *Chosenia: An amazing tree of Northeast Asia.* *Australindia* 2005, 63, 8–18.

Rao, G.D.; Zeng, Y.F.; Sui, J.K.; Zhang, J.G. *In situ* conservation of wild relatives in the Chinese National Repository of Apple Germplasm Resources. *Tree Genet. Genomes* 2015, 11, 106, doi:10.1007/s11295-015-0913-7.

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, doi:10.1111/j.1471-2229.2005.01155.x.

Excoffier, L.; Laval, G.; Schneider, S. *Arlequin* (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform.* 2005, 1, 11769-3430500100, doi:10.1177/11769343050010003.
35. Alexander, D.H.; Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* 2011, 12, 246, doi:10.1186/1471-2105-12-246.

36. Günther, T.; Coop, G. Robust identification of local adaptation from allele frequencies. *Genetics* 2013, 195, 205–220, doi:10.1534/genetics.113.152462.

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, doi:10.1007/s11434-006-1219-9.

38. 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, doi:10.13759/j.cnki.dlxb.2017.09.007.

39. 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, 403–403, doi:10.1051/forest:2008014.

40. Laakili, A.; Belkadi, B.; Medraoui, L.; Alami, M.; Yatrib, C.; Pakhrrou, O.; Makhloufi, M.; El Antry, S.; Laamarti, 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, doi:10.1007/s12298-018-0538-z.

41. Elliott, C.P.; Byrne, M. Genetic diversity within and between natural populations of *Eucalyptus occidentalis* (Myrtaceae). *Silvae Genet.* 2003, 52, 169–173.

42. 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, doi:10.1007/s10722-004-0597-6.

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, doi:10.3732/ajb.0800321.

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, doi:10.1007/s11105-012-0527-2.

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, doi:10.3389/fpls.2019.00130.

46. Wright, S. The interpretation of population structure by F-Statistics with special regard to systems of mating. *Evolution* 1965, 19, 395–420, doi:10.1111/j.1558-5646.1965.tb01731.x.

47. Jump, A.S.; Hunt, J.M.; Martínez-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, doi:10.1111/j.1365-294X.2006.03027.x.

48. Savolainen, O.; Pyhäjärvi, T.; Knürr, T. Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Evol. Syst.* 2007, 38, 595–619, doi:10.1146/annurev.ecolsys.38.091206.095646.

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, doi:10.1016/j.febslet.2006.03.072.

50. Cristina, M.; Petersen, M.; Mundy, J. Mitozen-Activated protein kinase signaling in plants. *Annu. Rev. Plant Biol.* 2010, 61, 621–649, doi:10.1146/annurev-arplant-042809-112252.

51. Brychkova, G.; Xia, Z.; Yang, G.; Yesbergenova, Z.; Zhang, Z.; Davydov, O.; Fluur, R.; Sagi, M. Sulfite oxidase protects plants against sulfur dioxide toxicity. *Plant J.* 2007, 50, 696–709, doi:10.1111/j.1365-313X.2007.03080.x.