Annotation:
Quantifying the genetic diversity present in plant populations is crucial for the success of selection of elite plants and breeding programs. The population structure and partition of genetic variance within and between plant populations allows inferences about the distribution and extent of variation available in plant breeding populations. The Chinese pine (Pinus tabuliformis Carriere) is a native species that is ecologically and economically important to Northern China. Studying the genetic variation and partition of P. tabuliformis will be invaluable for its efficient utilization. In this study, we selected 21 pairs of SSR primers to estimate the genetic diversity of 35 clones from four populations and 216 progenies from 22 full-sib families of P. tabuliformis. Our results showed that P. tabuliformis was a typical outcrossing plant ($F_{IS} = -0.053 / -0.165$) containing moderately high genetic diversity ($H_e = 0.542 / 0.605$) both for the clone populations and the progenies produced by the clone crosses via controlled pollination. The genetic differentiation between the clones ($F_{ST} = 0.076$) was lower than that between the full-sib progenies ($F_{ST} = 0.155$). Such diverse $F_{ST}$ may be promoted by different gene flows ($N_m = 3.881 / 1.546$) due to distinct pollination patterns. The clustering methods showed a lack of significant division of the clone populations, with a similar genetic admixture found for each of the clones. However, two groups were found between the full-sib families with genetic admixture between them. The analysis of molecular variance (AMOVA) showed that the most plentiful genetic variability exists between the cloned individuals (98%) as well as within the full-sib progenies (89%). Such distribution of genetic variation suggests rich internal variance and thus it is feasible to select the elite plants within each population or family. These results provide genetic materials and theoretical guidance for the selection and breeding of the Chinese pine.

Keywords: genetic variation; Pinus tabuliformis; progeny tests; SSR markers

1. Introduction

Biological diversity refers to the variation present within all the living organisms in the world and is presented at the diverse levels of genetic, species and ecosystem. Of the three levels at which biodiversity occurs, genetic diversity represents the amount of heritable variation within and among the populations of a species [1]. High genetic variation and diverse gene pools in a plant species are essential for its adaptation to novel pressures. For populations with a broad genetic base, recurrent selection will increase the frequency of favorable alleles and maintain the genetic variability of populations [2]. Thus, emphasizing the extent and distribution of heritable variation available in plant breeding populations
will promote the efficient utilization and safe conservation of germplasm resource. Several factors, including the evolutionary history (such as past bottlenecks and expansions of population), breeding system, environment and human factors, can directly determine the distribution and extent of genetic diversity in the populations [3]. Previous studies showed that much of a species’ available genetic variation may be partitioned among individuals within individual populations [4–7] or found among different populations [8,9].

For tree breeding, genetic tests are widely used because of their simple performance and integration of recurrent selection. In the case of progeny tests, full-sibs, half-sibs and inbred progeny tests are the most commonly involved. Recurrent selection among progeny tests is widely applied in different allogamous tree species, such as *Eucalyptus* [10,11], *Jatropha* [5] and *Pinus* [12]. Greater genetic gain per selection cycle is expected within full-sib progenies than within half-sib progenies [13,14]. The genetic structure of progenies in plant breeding is also widely used as a training set for genomic selection [15]. For such genomic selection among perennial trees, elite individual plants or clones are widely used for biparental crosses [16,17]. It is therefore crucial to obtain comprehensive information of the genetic diversity and population structure of the elite plants used for biparental crosses for the selection of base populations and the definition of breeding plans [18,19].

Molecular markers based on the polymerase chain reaction (PCR) are among the most important and efficient techniques for producing available data in terms of the genetic diversity of populations of interest. Simple sequence repeat (SSR) sequences, in which a basic unit of 1–6 nucleotides tandemly repeats many times across the genome, have been more extensively used in many studies of genetic diversity than other types of PCR-based molecular markers, for example, RAPD (random amplified polymorphism DNA), DAF (DNA amplified fingerprints), SCAR (sequence-characterized amplified regions), AFLP (amplified fragment length polymorphism) and ISSR (inter-simple sequence repeat) [10,20,21]. The utilization of SSR markers does not require any prior knowledge of a species genome and can be amplified using relatively simple techniques with a small amount of DNA. Compared to other molecular markers, SSR markers also have the characteristics of co-dominance, extensive polymorphisms and a large number of highly reproducible and informative alleles [22,23]. The use of SSR markers to estimate the genetic diversity of forest tree species has been frequently reported, for example within *Eucalyptus* [10,11], *Jatropha* [3], *Podocarpus elatus* [24], *Pinus kousaensis* [12], masson pine [7], *Pinus yunnanensis* [25], Chinese pine [26], *P. albicaulis* [27], *P. pinaster* [28], *Abies alba* [29], *Picea jezoensis* [30] and others.

The Chinese pine (*Pinus tabuliformis* Carriere) is known as a conifer tree that serves as an important native species of boreal and temperate forest ecosystems [31,32]. It is continuously distributed across a natural range extending through Northeast, North and Northwest China (ranging between 31–44° N and 101–125° E) and grows at altitudes varying from 500 to 2000 m in areas with diverse environments [33]. Excellent traits, such as strong tolerance to harsh environments, fast growth and valuable timber and resin have made *P. tabuliformis* a popular afforestation tree [32]. The genomic assembly and evolutionary analyses have unveiled the genetic basis of such an adaptation [34]. The germplasm source of *P. tabuliformis* has been widely collected for the establishment of seed orchards. Several studies have been carried out to investigate the genetic diversity of its natural populations, elite clones and individual plants and progenies [26,35,36]. These works have reinforced the value of SSR molecular markers in studying the genetic diversity of *P. tabuliformis*. Among the applications of SSR markers, the identification of elite plants and the characterization of genetic diversity for the examined populations and/or crossed families are two predominant highlights. Therefore, taking into account the advantages of knowing the nature of the genetic variance to design plant breeding schemes, it is quite crucial for the success of selection and breeding via quantifying genetic diversity within and among populations and/or families. In this study, based on the SSR molecular markers, we specifically aimed to evaluate the extent and distribution of the genetic diversity of 35 clones from four geographic populations, as well as 216 full-sib progenies generated
from clone crosses, and consequently provided theoretical and practical guidance for future breeding of the Chinese pine.

2. Materials and Methods

2.1. Plant Sampling for DNA Extraction

All samples used for DNA extraction were collected in the National Seed Orchard of *P. tabuliformis* of Wanjigou, located on the southern foot of Daqing Mountain, Tumote Left Banner, Hohhot, Inner Mongolia Autonomous Region (110°09’ E, 40°41’ N). A total of 35 clones from four populations of *P. tabuliformis* were collected. Information about these clones is listed in Table 1. Among 35 clones, 22 combinations, with 18 clones as maternal plants and 20 clones as paternal plants, were used for biparental crosses to generate full-sib progenies. A total of 216 full-sib progenies were finally collected from these biparental crosses, with at least 5 samples contained for each family (Table S1). Fresh young shoots of individual samples were dried in silica gel for whole-DNA isolation.

Table 1. Information about the four populations of all the cloned individuals.

<table>
<thead>
<tr>
<th>Orchard Name</th>
<th>Location</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Number of Clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qigou Forest Farm National Pinus tabuliformis Seed Orchard</td>
<td>Pingquan city, Hebei Province</td>
<td>118°40’ E</td>
<td>41°22’ N</td>
<td>3</td>
</tr>
<tr>
<td>Heilhe Forest Farm National Pinus tabuliformis Seed Orchard</td>
<td>Ningcheng County, Neimenggu Province</td>
<td>112°02’ E</td>
<td>36°40’ N</td>
<td>16</td>
</tr>
<tr>
<td>Weichang Longtou Mountain Forest Farm</td>
<td>Weichang Manchu and Mongolian Autonomous County, Chengde City, Hebei Province</td>
<td>117°12’ E</td>
<td>42°18’ N</td>
<td>7</td>
</tr>
<tr>
<td>Lingkong Mountain National Nature Reserve</td>
<td>Qinyuan County, Shanxi Province</td>
<td>118°25’ E</td>
<td>41°35’ N</td>
<td>9</td>
</tr>
</tbody>
</table>

2.2. DNA Extraction and SSR Genotyping

Whole genomic DNA of each sample was extracted from the dried leaves using a plant genomic DNA kit (ThermoFisher Scientifictm, Shanghai, China). To obtain SSR markers with high allele polymorphisms within *P. tabuliformis*, 59 pairs of SSR markers, collected from previous reports (Table S2), were preliminarily used for polymorphism investigation among 26 individuals, with one individual for each of the four clonal populations and each of the 22 full-sib families. Polymerase chain reactions (PCRs) were conducted in 20 µL, containing 10 µL 2× MasterMix (Tsingeke, Wuhan, China), 1.0 µL DNA (50 ng/µL), 1.0 µL of each primer (10 mMol), and 7 µL ddH2O. The PCRs were begun with initial denaturation at 98 °C for 2 min, followed by 35 cycles of 98 °C for 10 s, appropriate annealing temperature (Table S2) for 10 s, extension of 72 °C for 10 s, and a final-extension at 72 °C for 5 min. A negative control was carried out in each PCR run with no DNA contained in the reaction system. A total of 21 out of 59 pairs of SSR primers were obtained on the basis of these criteria: (1) SSR DNA fragments were easily PCR-amplified with strong signals to avoid non-specific amplification; (2) allele sizes of each SSR DNA fragments were discernible and differed among at least three individuals. Information about the screened primer pairs is listed in Table S2. PCR amplifications of the screened SSR markers were conducted among the remaining 225 individuals by running the above procedure with the corresponding annealing temperatures (Table S2).

SSR products were first detected in 3% agarose gels to ensure the successful amplification of SSR loci. To precisely genotype alleles in each SSR locus, SSR products were precipitated using 2.75 V of 100% ethanol and 0.1 V of 7.5 mol/L ammonium acetate, followed by centrifuging at 1200 rpm for 30 min. The pellets were collected by carefully removing all supernatant, followed by washing with 70% ethanol (v/v) before air-drying for 3 min and then resuspending in sterile distilled water. The purified products were pooled in triplex combinations according to fluorochrome-labelled forward primers, including three fluo-
rochrome moieties of HEX (hexachloro-6-carboxyfluorescein), FAM (6-carboxyfluorescein), and NED (7-benzo-5′-fluoro-2,4,7-trichloro-5-carboxyfluorescein). SSR genotyping was performed in ABI3730 capillary electrophoresis (3730xl, Applied Biosystems, Carlsbad, CA, USA) following the manual. The capillary electrophoresis volume of each triplex was 10 µL, containing 1.0 µL 10× diluted SSR products, 8.5 µL deionized formamide and 0.5 µL GeneScan™ 500 ROX™ dye size standards. Against the GeneScan™ 500 ROX™ size standards, allele sizes of SSR DNA fragments were scored and recorded automatically in base pairs as individual Gene Scan files using GENEMARKER Version2.2.0 (Soft genetics, State College, PA, USA).

2.3. Genetic Statistical Analyses for Clonal Populations and Full-Sib Families

The Hardy–Weinberg equilibrium (HWE) of each SSR locus was determined in GenALEX 6.503 [37] using Chi-Square (χ²) tests, with the degrees of freedom (df), probability and the significance P values statistically calculated. Genetic variations of the clonal populations and full-sib families were estimated using GenALEX 6.503 (https://biology-assets.anu.edu.au/GenALEX/Download.html, accessed on 14 July 2023) by calculating the parameters, including number of alleles (Na), number of effective alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), polymorphism information contents (PIC), Shannon’s information index (I), gene flow (Nm) and F-statistical parameters of Wright’s inbreeding coefficient of an individual (I) relative to the total (T) (FIT), Wright’s inbreeding coefficient of an individual (I) relative to the subpopulation (S) (FIS) and Wright’s fixation index indicating the effect of subpopulations (S) compared to the total population (T) (FST). The analysis molecular variance (AMOVA) was also carried out in GenALEX 6.503 to measure the partitioning of total genetic variability within and among populations/families. PowerMarkerV3.25 [38] was performed to calculate major allele frequency (MAF), gene diversity, Nei’s genetic distance and Nei’s genetic identity. According to the biparental genotypes at each SSR locus, the full-sib progenies having special genotypes inherited from both maternal and paternal plants were identified as true hybrids. The hybrid rate identified by each SSR locus was represented by P1, which was calculated by dividing the number of true hybrids identified at each SSR locus by the number of progenies. The hybrid rate of each family was indicated by P2, which was calculated by dividing the number of true hybrids of each family by the number of progenies. To identify whether the gene diversity of the full-sib progenies was affected by genetic differentiation between the clones, a mantel test between progeny gene diversity and biparental genetic distance was performed in the GenALEX v.6.503.

To estimate the genetic relationships among the clones, as well as among the full-sib progenies, Nei’s genetic distances were used to reconstruct unrooted phylogenetic trees by neighbor-joining (NJ) in POWERMARKER V3.25 and the resulted NJ trees were visualized in MEGA V7.0 [39]. The principal coordinate analysis (PCoA) was carried out in GenALEX 6.503 to divide the samples into different clusters. Furthermore, genetic groups for clone populations and full-sib families in all the SSR loci were determined using a Bayesian-model-based clustering approach in STRUCTURE v2.3.4 [40], with the admixture model and the independent allele frequencies assumed among populations/families. Ten independent runs were conducted for each K (=1–9) with 100,000 MCMC replications and 10,000 discarded as burn-in. To determine the optimal K values of the clone populations and full-sib progenies, plots of ad hoc posterior probability models of Delta K and the corresponding K values [41] were used to identify the optimal K value at the maximum Delta K. Alternatively, plots between the K values and mean logeP(D) values of the corresponding L(K) were also used to find the optimal K values at the maximum L(K) [42].

3. Results

3.1. Polymorphisms of the Selected SSR Markers

A set of SSR markers were used to screen the genetic variations and infer the population structures of the clones and full-sib progenies of P. tabulaeformis. In this study,
the Chi-Square tests showed that all the SSR loci conformed to the HWE, except that the loci of 800, pt0119, pt0319 and pt0450 departed from the HWE within the full-sib progenies (Tables S3 and S4). The size of the amplified SSR DNA fragments varied from 127 to 279 bp in length among the selected 21 SSR loci (Table S2). A total of 84 alleles were scored over these SSR loci for the clones and 77 alleles for the full-sib progenies, respectively, with an average of 4.0 and 3.5 per locus for the clones and full-sib progenies, respectively (Tables S5 and S6). The allelic diversity was greatly variable over all the SSR loci. The $H_o$ ranged from 0.269 to 0.898, and $H_e$ from 0.389 to 0.806 (Tables S5 and S6). The mean $H_e$ was lower than $H_o$ both for the clones ($H_o = 0.636$, $H_e = 0.605$) and the full-sib progenies ($H_o = 0.634$, $H_e = 0.542$) across all the SSR loci (Tables S3 and S4).

### 3.2. Genetic Diversity

The genetic diversity of the clones and full-sib progenies of *P. tabulaeformis* has been investigated using the selected 21 SSR molecular markers. Among the four clone populations, we found $N_a$ varied from 3.048 to 5.238 with an average of 4.0 per population, and $N_e$ ranged from 2.504 to 3.240 with an average of 2.8 per population. These values of $N_a$ and $N_e$ were higher than those within the full-sib families, with $N_a$ varying from 2.905 to 4.048, and $N_e$ from 2.185 to 2.649 (Table 2). The $I$, $H_o$ and $H_e$ varied among the four clone populations from 0.978 to 1.289, from 0.619 to 0.650 and from 0.569 to 0.662, respectively (Table 2). Similarly, these parameters of the full-sib progenies were lower than those of the clone populations, with the $I$ from 0.845 to 1.068, the $H_o$ from 0.571 to 0.724, and the $H_e$ from 0.496 to 0.583 (Table 2). In summary, the clone populations showed a higher genetic diversity than the full-sib progenies did (Table 2). Of note, the clone population of Ningcheng showed the highest genetic diversity of the clone populations, and the full-sib progenies produced by N37 and W2 showed the highest genetic diversity of the progenies (Table 2). The full-sib family with the second-highest genetic diversity was the P84-N22 family, followed by the N17-Q85 family (Table 2).

Meanwhile, no significant correlation was found between the genetic distance between clone parents and the gene diversity of full-sib progenies (Figure S1). In addition, the average value of hybrid rate identified by each SSR locus $P1$ was 14.61% (Figure S2) and the average value of hybrid rate of each family $P2$ was 97.96% (Table S11).

<table>
<thead>
<tr>
<th>Population</th>
<th>$\varphi$♂ Clones</th>
<th>$N_a$</th>
<th>$N_e$</th>
<th>$I$</th>
<th>$H_o$</th>
<th>$H_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone populations</td>
<td>Ningcheng (N)</td>
<td>5.238</td>
<td>3.240</td>
<td>1.289</td>
<td>0.650</td>
<td>0.662</td>
</tr>
<tr>
<td></td>
<td>Pingquan (P)</td>
<td>3.048</td>
<td>2.614</td>
<td>0.978</td>
<td>0.619</td>
<td>0.577</td>
</tr>
<tr>
<td></td>
<td>Qinyuan (Q)</td>
<td>4.048</td>
<td>2.821</td>
<td>1.124</td>
<td>0.630</td>
<td>0.613</td>
</tr>
<tr>
<td></td>
<td>Weichang (W)</td>
<td>3.619</td>
<td>2.504</td>
<td>1.016</td>
<td>0.646</td>
<td>0.569</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>3.988</td>
<td>2.795</td>
<td>1.102</td>
<td>0.636</td>
<td>0.605</td>
</tr>
</tbody>
</table>

Table 2. Genetic diversity of the clone populations and full-sib families. Capital letters in parentheses represent each of the clone populations and correspond to those in the full-sib families.
3.3. Genetic Differentiation and Structure

Three clustering methods, including NJ tree, PCoA and STRUCTURE assignment, were employed to categorize the clones and the full-sib progenies, respectively. The Nei’s genetic distance showed that Ningcheng and Qinyuan have the shortest genetic distance (0.087), the largest genetic identity (0.917) and the lowest \( F_{ST} \) (0.022) (Table S9), supporting their closest genetic relationship. The similar genetic divergence from Qinyuan to Pingquan and Weichang and the highest genetic divergence between Pingquan and Weichang (Table S9) both supported that Weichang and Pingquan were successively sisters to Ningcheng and Qinyuan. However, the NJ tree and PCoA analysis both indicated that no clear divisions or cluster groups were found among clones from these four populations (Figure 1). Nonetheless, the clones from Weichang and Pingquan were clustered in the earlier diverged clades and most clones of Ningcheng and Qinyuan were grouped together in the recently diverged clades (Figure 1). The STRUCTURE showed that all clone individuals can be clustered into two groups (\( K = 2 \)) based on the maximum \( \ln P(D) \) and delta \( K \) values (Figure 2). However, the genetic groups of the clones were not correlated with the geographic populations but rather with the genetic components. That is, all the clones contained two genetic groups despite their different proportions (Figure 2). The result of AMOVA showed that 2% of the total genetic variation was partitioned among clone populations and 98% of the total variation was explained by genetic variation within the clone population (Table 3).

For the full-sib progenies, the NJ tree showed that 22 full-sib families can be clustered into two clades. The first clade included the families of N121-Q71, N121-W130, N15-W10, N17-Q85 and N17-W24; the second clade comprised the remaining 17 families. The result of PCoA also supported the genetic relationships revealed by the NJ tree (Figure 3). These two genetic clades (\( K = 2 \)) were also recovered in the population structure of the 22 full-sib families according to the peaks of \( \ln P(D) \) and delta \( K \) values (Figure 4). Of note, a clear genetic mixture between these two clades was observed (Figure 4). The result of AMOVA indicated that 89% of the total genetic variation was partitioned within each of the full-sib families, and only 11% of the total variation was explained by the genetic variation among the full-sib families (Table 3).
correlated with the geographic populations but rather with the genetic components. That is, all the clones contained two genetic groups despite their different proportions (Figure 2). The result of AMOVA showed that 2% of the total genetic variation was partitioned among clone populations and 98% of the total variation was explained by genetic variation within the clone population (Table 3).

Figure 1. Principal coordinates analysis (A) and neighbor-joining tree (B) of the clones of *P. tabulaeformis*.

![Principal coordinates analysis (A) and neighbor-joining tree (B) of the clones of *P. tabulaeformis*.](image)

Table 3. Analysis of Molecular Variance (AMOVA) for the clone populations and the full-sib families.

<table>
<thead>
<tr>
<th>Samples</th>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>Est. Var.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone populations</td>
<td>Among populations</td>
<td>3</td>
<td>27,640</td>
<td>9.213</td>
<td>0.143</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>Within populations</td>
<td>66</td>
<td>458.975</td>
<td>6.954</td>
<td>6.954</td>
<td>98%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>69</td>
<td>486.614</td>
<td>7.097</td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Among families</td>
<td>21</td>
<td>241.391</td>
<td>21.019</td>
<td>0.767</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>Within families</td>
<td>410</td>
<td>2474.262</td>
<td>6.035</td>
<td>6.035</td>
<td>89%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>431</td>
<td>2915.653</td>
<td>6.802</td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

Figure 2. Using the corresponding Delta K statistics (A) and the mean posterior probability (loge P(D) (±SD) values (B) of each K to identify the optimal STRUCTURE assignment (C) of the 35 clones from four populations of *P. tabulaeformis* (K = 2).

![Using the corresponding Delta K statistics (A) and the mean posterior probability (loge P(D) (±SD) values (B) of each K to identify the optimal STRUCTURE assignment (C) of the 35 clones from four populations of *P. tabulaeformis* (K = 2).](image)
Figure 3. Principal coordinates analysis (A) and neighbor-joining tree (B) of the full-sib progenies of \( P. \) tabulaeformis.

4. Discussion

Comprehensive information about the genetic diversity and population structure are considered to be prerequisites for management and breeding programs, as their distribution and extent, measured by the variation in allelic frequencies and genetic composition within a population, is often integral to the capacity of adaptation, evolution and maintenance of populations \[43,44\]. In this study, the application of SSR markers proved to be efficient in...
evaluating the variability of the clone populations and full-sib families of *P. tabuliformis*. Although several reports have investigated the genetic diversity and population structure of clone populations and full-sib families of *P. tabuliformis* [26,36], the presented SSR data in terms of the genetic diversity of *P. tabuliformis* should also be considered as an important genetic source for future breeding plans. In addition, this is the first study to investigate the genetic diversity of both the clones and the full-sib progenies via biparental crosses. Our results showed that the selected SSR markers in this study were highly informative and polymorphic, with the number of effective alleles (Ne) varying from 2.404 to 2.795, lower than those of the natural populations of *P. tabuliformis* [26] but higher than those of the seed orchard of Pingquan [36]. Ne has been considered as an important parameter that evaluates the number of alleles displaying significant frequency in the population [45,46]. The Shannon diversity index (I) in this study ranged from 0.959 to 1.102, suggesting the existence of high genetic variability within and between the clone populations and the full-sib families. Although the I revealed in this study was much lower than that reported by Zhou et al. [26], high genetic diversity within the examined clone populations and full-sib families was also supported because of the smaller size of samples in this study than in those reported by Zhou et al. [26,47]. Furthermore, the polymorphism information contents (PIC) in this study varied from 0.618 to 0.592, supporting the capacity and efficiency of the selected SSR markers for genetic diversity identification in *P. tabuliformis* [48]. Of note, the PIC values in this study were higher than those of previously reported SSR markers [26,35,36], and thus the genetic diversity of *P. tabuliformis* will be overestimated in this study. It has been documented that the use of genomic data can accurately scan the genetic variation of target species populations (e.g., [49]). However, this is seemingly impossible for *P. tabuliformis*, due to its large genome size [34] and correspondingly huge sequencing costs. Alternatively, developing and using various molecular markers, for example RAPD and ISSR, to screen the genetic diversity for clonal identity and progeny test is also effective and powerful for forestry trees and crops [50–53].

The expected heterozygosity (He) is an important parameter in measuring gene diversity [54]. Moderately high genetic diversity was found for both the clones and the full-sib progenies of *P. tabuliformis* (mean Ho of 0.636/0.634; mean He of 0.605/0.542). Such genetic variability within accessions from *P. tabuliformis* was similar to that of the natural individuals from different geographic populations of *P. tabuliformis* [26] but higher than that of samples of *P. tabuliformis* from Pingquan [40]. Meanwhile, the genetic diversity of examined samples of *P. tabuliformis* was higher than that of some Pinus species, such as *P. strobus* (He of 0.531) [55], *P. tabuliformis* (He of 0.529) [56], *P. koraiensis* (He of 0.521) [12], or similar to that of some other Pinus species, such as *P. echinata* (mean He of 0.765) [57], *P. cembra* (mean He of 0.581) [58], and *P. densiflora* (mean He of 0.671) [59]. This moderately high genetic variability within and between the examined clone populations of *P. tabuliformis* may be a result of a wind outcrossing (*F_{IS} = -0.053*) and the unique reproductive biology characteristics of *P. tabuliformis* [60]. As a typical wind-pollinated plant, the pollen of *P. tabuliformis* can spread over 85 m and promote gene flow within populations [61]. Meanwhile, the winged seeds of *P. tabuliformis* can travel longer distances than its pollen and promote gene flow within and between populations. This will further increase the genetic diversity of *P. tabuliformis* via genomic combination. Of note, higher gene flow (Nm = 3.881) and lower genetic differentiation (*F_{ST} = 0.076*) were found for clonal populations of *P. tabuliformis*. The three clustering methods supported no clear genetic isolation between the clone populations. The optimal K (=2) value indicated that all clones comprised two genetic clusters. We thus concluded that all clones were generated from the same progenitors. According to the bio-ecological characters of *P. tabuliformis*, we speculated that the wind dispersal pattern and long generation time might have lessened genetic divergence within and between clonal populations by homogenizing genetic variation that has accumulated along generation time despite little such accumulation. However, lower gene flow (Nm = 1.546) and higher genetic differentiation (*F_{ST} = 0.155*) were observed in the full-sib progenies. The controlled pollination between the known clones from different
populations (with $F_{IS} = -0.165$) to generate the F1 progenies of full-sib families might have limited pollen/seed flow between full-sib families. The clustering analyses of the 22 full-sib families showed that they formed two groups. This supported the capacity of SSR markers to efficiently categorize the families based on the existence of genetic homogeneity and heterogeneity. In this study, the full-sib families that were clustered together presented a greater level of genetic similarity, regardless of whether they shared one progenitor; for example, the families of N121-Q71, Q121-W130, N15-W10, N17-Q85 and N17-W24 were clustered together and showed similar genetic structure as well as a pattern similar to those of the remaining families. We thus speculated that the higher genetic differentiation between full-sib progenies may be due to the hybrid effects already described in biparental crosses, for example, nucleo-cytoplasmic interaction and maternal effect. That there is no significant correlation between the genetic diversity of the full-sib progenies and between-clone genetic distances also supports that genetic differentiation between biparental clones likely cannot directly determine the genetic diversity of hybrid progenies. Previous phylogeographical studies showed that natural populations of *P. tabuliformis* can be divided into three geographical clades and derived from different ancient lineages [62,63]. Compared to other forest trees, for example *P. albicaulis* [27], *P. pinaster* [28], *Abies alba* [29] and *Picea jezoensis* [30], natural populations of *P. tabuliformis* show lower total and mean genetic diversity [62,63]. These comparisons suggest that different geographical clades of *P. tabuliformis* should be protected to avoid habitat fragmentation, genetic diversity loss and population extinction. This also suggests that elite plants selected from different natural germplasm sources can be used for afforestation in the same region, as gene flow between them would increase the genetic diversity and promote the adaptation capacity of novel varieties of *P. tabuliformis*. In addition, considering the diverse genetic components of each of the progenies, biparental crosses between different clones are suggested in future breeding to provide a large amount of genetic diversity in the progenies [64].

The molecular variance analysis (AMOVA) measures the total variation of a species within and between populations. The AMOVA showed that the highest proportion of genetic variation was found within clone populations (98%), with a relatively lower variation between populations (2%). Such a genetic variation pattern has been frequently reported in tree species populations, such as in *Jatropha curcas* [5] and in other *Pinus* species [12,56], as well as in the natural populations and cultivated varieties of *P. tabuliformis* [26,32,52]. From these results, it was recommended that the internal variance of *P. tabuliformis* should be fully considered to avoid genetic narrowing and future gain compromise during the recurrent selection and recombination. Meanwhile, a similar genetic distribution pattern was detected within the full-sib families, with a significant increase of genetic differentiation between progenies (11% of the total variance portioned among families) compared to that of the clonal population (2%). The distribution and extent of genetic variation within and between the 22 full-sib families suggests plentiful genetic diversity within individual families, and such variation can be used in future breeding, although there is no consensus for how to consider the larger internal variance and the smaller variance among families during breeding [65,66]. A high-quality genome assembly of *P. tabuliformis* [30] has provided an opportunity to use genome-wide association study (GWAS) to combine phenotypes and genetic variation for the present accessions. Such an association will be useful for the selection of elite plants or the performance of new biparental crosses between more divergent samples to maximize heterotic effect and create more new varieties.

*Pinus tabuliformis* plays an important ecological and economic role in China and provides forest, wood and resin by-products. However, overlogging of natural forests over a sustained period has led to a dramatic reduction of natural forest cover and rapid decrease of genetic diversity among its remaining species populations. Plantation forests are a useful and effective avenue for restoring *P. tabuliformis* by selecting elite plants with some superior characteristics. However, afforestation often results in 100% monospecific forests, and consequently the planted forests are sensitive to disease and insect pest incidence. Thus, a better understanding of the extent and distribution of genetic diversity of the plants used for
Afforestation is essential for long-term strategies to maintain the health and sustainability of *P. tabuliformis* [66]. In this study, clone crosses created many diverse progenies, although little genetic differentiation was observed between clones. This provides a clue that biparental crosses may create new varieties to enrich the existing genetic pool of *P. tabuliformis*. In this way, more full-sib progenies can be created for afforestation through biparental crosses, especially those involving natural elite plants with higher genetic variability [62,63]. Given the long-distance dispersal ability and outcrossing of *P. tabuliformis*, ecological forests can be planted in areas close to natural forests to reinforce gene flow [3,49] and consequently the enriched genetic diversity will increase population evolutionary potential in novel environments, as well as adaptation to climate changes and habitat fragmentation.

5. Conclusions

The selected 21 polymorphic SSR markers in this study were useful and efficient, enabling the evaluation of moderately high genetic diversity, the identification of two distinct genetic groupings for both the clones and full-sib progenies, and the estimation of the distribution of genetic variation within and between the populations or families of *P. tabuliformis*. The same origin pattern of all the clones, long generation time and wind-dispersal traits of pollen, might have led to lessened genetic differentiation between the clones, while the genetic diversity and controlled pollination of biparental crosses have laid the foundation of genetic variation and differentiation between the full-sib progenies. Genetically diverse progenies have been generated via different biparental crosses in spite of the slight genetic differentiation between parents. This suggests that biparental crosses between the selected elite clones would generate novel varieties. A high proportion of the genetic variation of the examined clones and progenies was found to be partitioned within individual populations or families, suggesting plentiful internal variation. For breeding programs, it is feasible to consider both the larger internal variance and the smaller variation among populations or families, because even a little genetic variation between populations or families will lead to the formation of diverse genetic variation, such as biparental crosses between the clones with lower genetic differentiation and similar genetic components. This study provides genetic materials for future breeding by using a genome-wide association study to investigate the association between novel phenotypes and genetic variation among these accessions of *P. tabuliformis*.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/horticulturae9111205/s1. Table S1 Information about the clones and the full-sib progenies; Table S2 Information about SSR Primers; Table S3 Hardy–Weinberg equilibrium test at each SSR locus for clones; Table S4 Hardy–Weinberg equilibrium test at each SSR locus for full-sib progenies; Table S5 Genetic parameters of clones at each SSR locus; Table S6 Genetic parameters of full-sib progenies at each SSR locus; Table S7 Genetic divergence and variation of clones at each SSR locus; Table S8 Genetic divergence and variation of full-sib progenies at each SSR locus; Table S9 Genetic distance, genetic identity and genetic differentiation between pairwise populations of clones; Table S10 Genetic distance between clones used for crossing combinations; Table S11 Identification rate of hybrid progenies for each of the clone-crossing combinations; Figure S1 Colinear relationship test between gene diversity within full-sib progenies and between-clone genetic distances; Figure S2 Identification rate of hybrid progenies at each SSR locus.

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

Funding: This research was funded by the Young Teachers’ Research Ability Improvement Project of Inner Mongolia Agricultural University (grant number BR230161), the Key R&D and Achievement Transformation Plan of Inner Mongolia Autonomous Region (grant number 2023YFDZ0017) and the National Natural Science Foundation of China (grant number 31160167).
Data Availability Statement: All data generated or analyzed during this study are included in this published article [and its supplementary information files]. In addition, the datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

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

References


23. Ferreira, V.; Castro, I.; Rocha, J.; Crespi, A.L.; Pinto-Carnide, O.; Amich, F.; Almeida, R.; Carnide, V. Chloroplast and nuclear DNA studies in Iberian Peninsula endemic *Silene scabriflora* subspecies using cpSSR and ISSR markers: Genetic diversity and phylogenetic relationships. *Biochem. Syst. Ecol.* 2015, 61, 312–318. [CrossRef]


27. Ferreira, V.; Castro, I.; Rocha, J.; Crespi, A.L.; Pinto-Carnide, O.; Amich, F.; Almeida, R.; Carnide, V. Chloroplast and nuclear DNA studies in Iberian Peninsula endemic *Silene scabriflora* subspecies using cpSSR and ISSR markers: Genetic diversity and phylogenetic relationships. *Biochem. Syst. Ecol.* 2015, 61, 312–318. [CrossRef]


37. Smouse, P.E.; Banks, S.C.; Peakall, R. Converting quadrat-based diversity to biodiversity: Both animals and alleles are diverse, but some are more diverse than others. *PloS ONE* 2017, 12, e0185499. [CrossRef]


42. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* 2000, 155, 945–959. [CrossRef]


63. Xia, H.H. Landscape Genomics of *Pinus tabuliformis* Carr. and Its Evolutionary History; Beijing Forestry University: Beijing, China, 2019.


66. Laviola, B.G.; Alves, A.A.; Rosado, T.B.; Bhering, L.L.; Formighieri, E.F.; Peixoto, L.d.A. Establishment of new strategies to quantify and increase the variability in the Brazilian Jatropha genotypes. *Ind. Crops Prod.* 2018, 117, 216–223. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.