Ex situ conservation of *Pinus koraiensis* can preserve genetic diversity but homogenizes population structure

Yuewei Tong\(^a, b\), Walter Durka\(^c, d\), Wangming Zhou\(^a\), Li Zhou\(^a\), Dapao Yu\(^a\), Limin Dai\(^b, *\)

\(^a\) CAS Key Laboratory of Forest Ecology and Management, Institute of Applied Ecology, Shenyang 110016, China
\(^b\) Department of Community Ecology (BZF), Helmholtz Centre for Environmental Research-UFZ, Halle 06120, Germany
\(^c\) University of Chinese Academy of Sciences, Beijing 100049, China
\(^d\) German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher Platz 5e, Leipzig D-04103, Germany

**A R T I C L E  I N F O**

**Keywords:**
*Pinus koraiensis*  
Ex situ conservation  
Clonal seed orchard  
Genetic diversity  
Population structure

**A B S T R A C T**

*Pinus koraiensis* is a conifer species of ecological and economic importance in northeast China that has been excessively exploited in recent years. A clonal seed orchard (CSO) was established including potentially genetically differentiated provenances from the whole distribution area, we applied nine simple sequence repeat (SSR) markers to study the genetic diversity and population structure of six maternal populations as well as their progeny populations. The results showed a high genetic diversity in both maternal and progeny populations, with the average expected heterozygosity of 0.617 and 0.632, respectively. The level of genetic diversity in the progeny populations was slightly higher than that in the maternal populations, and almost all diversity descriptors were correlated between maternal and progeny populations, indicating that the CSO could preserve the established species’ gene pool. An overall low level of genetic differentiation of *P. koraiensis* (Fst = 0.029 and 0.025 for maternal and progeny populations) was found. The six maternal populations clustered into two groups by Bayesian cluster analysis with the two northernmost populations comprising one group and the other four populations as another genetically differentiated group, potentially indicating regional adaptation. The northern group had significantly higher levels of within population diversity. Population structure was not evident anymore in the progeny populations, suggesting that mating among differentiated provenances within the CSO homogenized the different gene pools, although differences in diversity were still maintained. These results will inform efforts for the conservation and management of *P. koraiensis* and provide guidance for future studies of population genetics and breeding programs.

1. Introduction

Forests are considered one of the most complex terrestrial ecosystems due to the high level of biodiversity, and play an important role in the function of forest ecosystems (Geburek and Konrad, 2008; Litkowiec et al., 2018). Genetic diversity, widely recognized as a key component of biodiversity (Millennium Ecosystem Assessment, 2005), provides the essential basis for the adaptation and resilience of species to environmental stress and change (Potter et al., 2017). As long-lived, immobile, and often widespread life forms, forest tree species need high levels of genetic diversity to adapt to changing environmental influences (Loo et al., 2014; Potter et al., 2017). Unfortunately, the genetic resources of forest trees are threatened by an increased use of timber and reduction of the area of forestland (Rajora and Mosseler, 2001). For example, deforestation, which directly eliminates numbers and sizes of the locally adapted populations is an immediate threat to forest tree genetic diversity (Ledig, 1992). Therefore, maintaining or enhancing genetic diversity in forest tree species is an urgent global necessity for genetic resource conservation (Ratnam et al., 2014; Holliday et al., 2017). This may require the integration of both in situ (on site) and ex situ (off site) conservation strategies, especially in high-value species and those with small and vulnerable populations (Dawson et al., 2013; Pritchard et al., 2014; Schwartz et al., 2017).

*Ex situ* conservation strategies preserve plants or plant germplasm away from their areas of natural occurrence (Given, 1994). A clonal seed orchard (CSO), one of these strategies, often consist of trees selected from a single provenance, however, also multiple provenances may be combined, thus broadening the genetic basis (Zobel et al., 1958; Giertych, 1975). The main goals of CSOs are to achieve genetic gain in economically relevant traits due to initial selection, and to maintain genetic variation by enhancing panmixia (Chaloupkova et al., 2019). Because CSOs represent the link between breeding programs and...
reforestation activities, inadequate genetic diversity in seed crops may result in accumulation of inbreeding, which directly reduces potential productivity (Kang et al., 2001; Kang et al., 2005; Sonstebo et al., 2018; Chaloupkova et al., 2019). Generally, the genetic diversity of seeds produced in a CSO is determined by that of provenances and parent trees which, however, was unknown during the establishment. Moreover, the diversity of seeds will depend on planting design (Yuan et al., 2016) and the pollen dispersal distance. Although most forest temperate forest trees are wind pollinated, pollen dispersal distance is limited and may be non-random, e.g. due to preferential wind direction (Feng et al., 2010). If provenances are genetically differentiated and seeds are produced by random mating in a seed orchard, higher genetic diversity is expected for seeds compared with the average source provenance. However, while mixing genetically differentiated populations will increase genetic diversity, potentially with hybrid vigor, this is connected with a potential risk of outbreeding depression (hybrid breakdown) by disrupting coadapted gene complexes if provenances are adapted to different environmental conditions (Johansen-Morris and Latta, 2006). Therefore, as CSOs are a tool for breeding of numerous tree species around the world, studying the genetic diversity of maternal provenances and offspring produced in a CSO can provide a reliable basis for evaluating the potential risks in the management of seed orchards and plantations (El-Kassaby and Ritland, 1996; Nielsen and Hansen, 2012; Sonstebo et al., 2018).

*Pinus koraiensis* Siebold et Zucc. (Korean pine, Pinaceae), has its centre of distribution in Northeastern China, and ranges from South Korea into Russia and Japan (Aizawa et al., 2012). It is one of the most ecologically and economically important conifer species in the mountainous area of Northeastern China (Ma et al., 1992). The species is a dominant forest tree species in its habitat and produces high-quality timber and edible pine nuts (Chen et al., 2010; Aizawa et al., 2012). In recent decades, the populations of natural broad-leaved Korean pine mixed forests have been sharply reduced (Yu et al., 2011), and the species has been listed as nationally endangered in China (http://www.plant.csdb.cn/endangeredplants). Thus, the protection of *P. koraiensis* has become urgent, especially for the genetic resources. There are several hypotheses about the origin of Korean pine populations in continental Asia ranging from a single glacial refugium, substantiated by the genetically depauperate plastid genomes that lack variation (Aizawa et al., 2012) to the hypothesis of several continental in situ relic populations (Potenkon and Velikov, 1998). For the nuclear genome of *P. koraiensis* a gradient of genetic variation decreasing from South (Korea) to North (Russia) (Kim et al., 2005) indicates however that populations are not homogenous and likely display genetic structure.

In the 1980s, a breeding program to protect the genetic resources of *P. koraiensis* was started in China and a clonal seed orchard (CSO) was established including provenances from the whole distribution area. However, the genetic diversity and population structure of the different provenances had not be assessed as well as their progeny. Therefore, our objectives of this study were, based on six populations of *P. koraiensis* in the CSO, to (i) estimate genetic diversity and population structure of *P. koraiensis* provenances, and (ii) compare the genetic variation of maternal populations and offspring populations that represent first generation hybrids produced in the CSO.

### 2. Materials and methods

#### 2.1. Study site and plant material

We studied the Hongwei Seed Orchard which is located in Lushuihe town, northeast China (42° 28′ 30“ N 127° 47′ 00“E, 760 ~ 800 m a.s.l., 2.7°C mean annual temperature, 871 mm mean annual precipitation) and was established in 1989. This clonal seed orchard (CSO) was established from cuttings of 660 *Pinus koraiensis* mother trees which were selected by phenotypic evaluation (volume, height, diameter at breast height, bole straightness, and branching habits) from 17 provenance areas in Northeast China, thus covering large parts of the species’ range. Each provenance has 20 to 60 mother trees (clones) and each clone is represented by 10 to 20 ramets. The CSO has a total area of 16.3 ha with rectangular planting scheme with 3 m distance between trees in 17 plots. All ramets were distributed using random block designs, and avoiding close vicinity of ramets from the same clone.

We selected six provenances, hereafter called “populations” (Table 1), representing a S-N cline with the two northernmost populations (DL, FL) separated from the southern populations by the river Heilong Jiang (Amur). The minimum and maximum geographic distances among populations ranged from 100 km to 650 km. For each population, we collected both needles (hereafter referred to as “mothers”) and seeds from one random ramet per clone for each of 25 clones (Table 1). Seeds were germinated and then the needles of one random seedling per clone were sampled. All samples were stored at −80 °C before DNA extraction.

<table>
<thead>
<tr>
<th>Population ID</th>
<th>Latitude (°)</th>
<th>Longitude (°)</th>
<th>Annual Mean Temperature (°C)</th>
<th>Annual Mean Precipitation (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lushuihe</td>
<td>42.53</td>
<td>128.80</td>
<td>4.12</td>
<td>783.68</td>
</tr>
<tr>
<td>Huangnihe</td>
<td>43.55</td>
<td>128.01</td>
<td>3.76</td>
<td>655.42</td>
</tr>
<tr>
<td>Dahailin</td>
<td>44.52</td>
<td>128.86</td>
<td>1.14</td>
<td>619.72</td>
</tr>
<tr>
<td>Dongfanghong</td>
<td>46.58</td>
<td>133.58</td>
<td>3.17</td>
<td>648.99</td>
</tr>
<tr>
<td>Dalin</td>
<td>47.18</td>
<td>128.85</td>
<td>0.30</td>
<td>574.59</td>
</tr>
<tr>
<td>Fenglin FL</td>
<td>48.13</td>
<td>129.19</td>
<td>0.04</td>
<td>596.15</td>
</tr>
<tr>
<td>Huangnihe</td>
<td>43.55</td>
<td>128.01</td>
<td>3.76</td>
<td>655.42</td>
</tr>
<tr>
<td>HNH</td>
<td>43.5</td>
<td>128.01</td>
<td>3.76</td>
<td>655.42</td>
</tr>
<tr>
<td>LSH</td>
<td>42.53</td>
<td>128.80</td>
<td>4.12</td>
<td>783.68</td>
</tr>
<tr>
<td>Dailin</td>
<td>47.18</td>
<td>128.85</td>
<td>0.30</td>
<td>574.59</td>
</tr>
</tbody>
</table>

#### 2.2. Genotyping

Total genomic DNA was extracted from leaf tissue using the DNAsecure Plant Kit DP320 (TIANGEN) according to the manufacturer’s instructions. The quality and concentration of genomic DNA were assessed by agarose gel electrophoresis. Samples were genotyped at eight dinucleotide and one trinucleotide SSR markers (P5, P6, P29, P45, P51, P52, P62, P63 and P79; see supplement Table S1) (Yu et al., 2012; Jia et al., 2017). PCRs for all loci were performed separately in a 20 μl reaction volume containing 40 ng of genomic DNA, 0.3 μM concentrations of each primer (the forward primers were labelled with a fluorescent dye (FAM or HEX)), 0.1 mM dNTP (TransGen Biotech, Beijing, China), 2 μl 10× buffer and 1.0 U Taq DNA Polymerase (TransGen Biotech, Beijing, China). Cycling parameters were initial denaturation at 94 °C for 5 min; followed by 35 cycles of denaturation at 94 °C for 30 s, annealing for 35 s (for annealing temperature, see supplement Table S1), extension at 72 °C for 40 s; and final extension at 72 °C for 3 min. The amplification products were separated on the ABI3730XL DNA analyzer (Applied Biosystems) using GS-500 LIZ (Applied Biosystems) as internal size standard. Allele binning and genotyping were performed with GeneMarker® software (SoftGenetics LLC, State College, PA, USA).

#### 2.3. Data analysis

##### 2.3.1. Genetic diversity indices

We calculated a set of population genetic diversity estimators using GeneAlEx 6.5.1 (Peakall and Smouse, 2012). The observed (*A*) and effective (*A*<sub>e</sub>) number of alleles, the observed (*H*<sub>e</sub>) and expected (*H*<sub>c</sub>) heterozygosity and the Shannon diversity index (*I*) were all calculated for each locus, over all loci, and for each population. In addition, we calculated allelic richness (*A*<sub>r</sub>) and the inbreeding coefficient *F*<sub>S</sub> (Weir and Cockerham, 1984) in FSTAT 2.9.3 (Goudet, 1995).
2.3.2. Population structure analysis

Genetic population structure was assessed and quantified in several steps. First, Principal coordinates analysis (PCoA) was used to illustrate genetic distances between individuals. Next, analysis of molecular variance (AMOVA) (1000 permutations) based on the degree of genetic divergence among populations was performed using GenAlEx 6.5.1 (Peakall and Smouse, 2012). In addition, genetic structure was assessed with a Bayesian clustering algorithm using STRUCTURE version 2.3.4 (Hubisz et al., 2009). The range of possible number of clusters (K) tested from 1 to 13 for all populations and 7 for mother and seed populations (number of populations plus 1), for which a series of ten independent runs was performed with a burn-in period of 10 000 steps followed by 100 000 MCMC replicates. We used provenance as location prior (LOCPRIOR) (Hubisz et al., 2009) under the admixture model. Note that using a location prior will enable detection of weak genetic differentiation among locations only when differentiation is actually present (Manual of Structure software). The ad hoc statistic ΔK (Evanno et al., 2010) together with inspecting the log-likelihood values was used to identify the most likely number of clusters. The ten runs from the most probable K were averaged by applying a FullSearch algorithm by CLUMPP 1.1.2 (Jakobsson and Rosenberg, 2007).

2.3.3. Statistical analysis

Mean values of genetic diversity descriptors were compared among sample types or populations by T-test in R (R Core Team, 2013). We tested for clinal variation of genetic diversity descriptors with latitude using linear models in R.

3. Results

3.1. Genetic diversity

The nine SSR loci assayed were all polymorphic in Pinus koraiensis. Overall, there were a total of 73 different alleles with an average of A = 6.0 alleles per locus. Among loci, locus P51 revealed the highest level of polymorphism across all diversity descriptors, while P5 and P29 showed lowest values (Tab. S2). Among loci, observed heterozygosity (H_o) varied from 0.457 to 0.890 with an average of 0.685, expected heterozygosity (H_e) was just slightly lower than H_o and ranged from 0.422 (P29) to 0.853 (P51), with an average of 0.624 (Tab. S2). A_r ranged from 2.5 (P5) to 11.3 (P51), and the average A_r was 6.7 (Table S2). The F_{is} showed negative and significant values in loci P6, P45, P51, P52 and P63, and the overall F_{is} calculated across 9 loci was ~ 0.109 (Table S2, P < 0.05), indicating the existence of heterozygosity excess.

Genetic diversity parameters and inbreeding coefficients at population level are given in Table 2. In mothers, for all the diversity descriptors, the two northernmost populations had the highest values. When grouped into northern and southern populations, the northern group had significantly higher values for A, I, H_o, uH_o and A_r (Table 2). When testing for clinal variation with latitude, A, I and A_r showed increasing values towards North (Table 2). The F_{is} showed significantly negative values in four populations indicating slight departure from Hardy-Weinberg equilibrium with heterozygosity excess. Almost all diversity descriptors were correlated between mothers and seeds (H_o, uH_o, I: p < 0.05; A, A_r: p < 0.1). For all diversity descriptors values where higher in seeds than in mothers, significantly so in A, I, and A_r. The spatial patterns found for mothers were similarly found, however with reduced strength, in seeds as the two northern populations had significantly higher values than the southern populations for I, H_o, H_e and uH_e, but only marginally significantly higher values for A and A_r. Clinal variation with latitude was only marginally significant for the same parameters as in mothers (A, I, A_r). The inbreeding coefficient was significantly negative in four seed populations, but F_{is} was not correlated between mothers and seeds.

3.2. Genetic differentiation and population structure

The non-hierarchical AMOVA (Table 3) showed that 4.25%, 3.49% and 2.27% of genetic variation was among populations (P < 0.001) for all, mother, and seed populations, respectively, with the rest residing within populations. Pairwise F_{ST} values were low (0.009 to 0.035), but often significant (Tables 4 and S3). For the mother populations, nine out of 15 pairwise F_{ST} values were significant, while for the seed populations six were significant (Table 4). Pairwise F_{ST} values were significant mostly for comparisons including the two northern populations.

Furthermore, a hierarchical AMOVA analysis revealed significant differentiation (2.56%) between groups of mother and seeds (P < 0.001), and a similar amount of variation (2.80%) among populations within groups (P < 0.001) (Table 3). Similarly, a hierarchical AMOVA was carried out according to the results of the genetic diversity analysis which indicated that the two northern and four southern populations could be considered as two geographic groups. For the mother populations, genetic variation between groups was 2.37% (P < 0.001) and variation among populations within groups was 2.19% (P < 0.001, Table 3). However, in the seed populations, the variation between groups was not significant and 2.19% resided among populations within groups (Table 5).

The subsequent STRUCTURE analysis provided additional information on number of gene pools and the level of genomic mixture and admixture among populations. For mothers, the most probable division with the strongest support in terms of log-likelihood values was at K = 2 (Fig. S1), with southern populations LSH and HNH representing pure cluster I (blue), northern populations DL and FL pure cluster II (red) and intermediate populations DHL and DFH being admixed (Fig. 1). For seeds, only a single cluster was found, with L(K) being highest at K = 1 and no geographic separation of populations whatsoever (Fig. S2, S3). Analyzing all populations, the ΔK value reached a maximum for K = 2 with mothers (blue) and seed (red) populations representing the two clusters (Fig. 2, S4).

4. Discussion

4.1. Genetic diversity

The main task of seed orchards is to produce a large number of high genetic quality seeds for reforestation without reducing its genetic diversity (Chaloupkova et al., 2019). In this paper, we analysed genetic diversity and population structure of Pinus koraiensis in a CSO, to assess the degree of genetic differentiation among provenances and the changes in genetic variation between mothers and seeds.

Firstly, the average expected heterozygosity of the mother populations of Pinus koraiensis in this study was 0.617, indicating that genetic diversity was high. This result is similar to most current research revealing high levels of genetic diversity in Pinus species when assessed with SSR markers. For example, H_o was 0.782 for P. thunbergii (Iwaizumi et al., 2018), 0.601 in P. koraiensis (Feng et al., 2010), 0.55 in P. sylvestris (Bernhardsson et al., 2016), 0.586 also in P. sylvestris (Toth et al., 2017), 0.531 in P. strobus (Mandak et al., 2013), and 0.428 in P. yunnanensis (Xu et al., 2016). Meanwhile, we found a pattern of genetic diversity for P. koraiensis in this study, that is, the genetic diversity of the northern populations was higher than that of the southern populations. This is in contrast to results of Kim (Kim et al., 2005) who found a large-scale decline of variation from south (North Korea) to north (Russia). Thus, large scale patterns are not necessarily true on a more regional scale. This pattern went also undetected in Kim et al. (2005), potentially due to biased sample sizes or choice of genetic markers or of populations. Anyway, the results indicate that against a general cline of northward reduction of genetic diversity, populations in Xiaoxinganling mountains maintained an increased level of diversity.

Secondly, the level of genetic diversity of the seed populations in
and inbreeding coefficients of maternal and seed populations; significance, as well as with the two northern and four southern populations as two groups: N vs. S, found across generations. Roberds & Conkle found that there was no significant change allele frequencies between the maternal and progeny populations of *Pinus taeda* (Roberds and Conkle, 1984). Furthermore, for *Pinus massoniana* (Ai et al., 2006) and *Pseudotsuga menziesii* (Prat and Arnal, 1994) in CSOs is was shown that genetic diversity is maintained across generations. The underlying causes for the observed increase of diversity in our study likely is first the rich gene pool of 17 source populations representing the whole species range in China, second the efficient pollen dispersal of *P. koraiensis* up to 60 m (Feng et al., 2010) and third the planting strategy in this CSO that maximizes the probability of crossing between accessions. Therefore, the genetic diversity of the seed populations of *P. koraiensis* can not only be maintained, but increased in this CSO. Still, the geographic pattern of increased variation in northern populations is maintained, although with reduced effect in the seed populations.

### 4.2. Genetic structure and genetic differentiation

The population structure is important for establishing the appropriate scale and subunits for conservation management (Moritz, 1999). It is affected by mutation, gene flow, natural selection and genetic drift, and thus is related to the evolutionary history and biological characteristics of the species (Loveless and Hamrick, 1984; Schneller and Lieb, 2007). Population structure is manifested mainly by genetic differentiation among populations (Frankham et al., 2002). As expected for conifer tree species, we found an overall low level of population differentiation (*F_{ST} = 0.029* and 0.025 for mother and seed populations), which is consistent with the findings of most species with outcrossing breeding system and wind pollination (Belletti et al., 2012; Iwaiuzumi et al., 2013; Mandak et al., 2013; Durka et al., 2017). *Pinus koraiensis* is wind pollinated and outcrossing (Feng et al., 2010), and its seeds are dispersed mainly by birds and rodents (Miya, 1987). Such pollination and seed dispersal syndromes are beneficial to increase gene flow within and among populations (Nyholm, 2004; Petit et al., 2005), and consequently to reduce genetic differentiation among populations.
likely is too simple, as has been found in many Chinese forest species resulting in constant decrease of genetic variation from south to north that a scenario of a single refugium with later northward expansion (et al., 2005; Aizawa et al., 2012) have been stated. Our data suggest refugium hypothesis with a latitudinal cline of genetic variation (Kim et al., 2005) and a single-

study of the phylogeography of Pinus koraiensis (Potenko and Velikov, 1998) and a single-refugium hypothesis with a latitudinal cline of genetic variation (Kim et al., 2005; Aizawa et al., 2012) have been stated. Our data suggest that a scenario of a single refugium with later northward expansion resulting in constant decrease of genetic variation from south to north likely is too simple, as has been found in many Chinese forest species (Shi et al., 2014). More complex scenarios have to be considered and the high diversity detected in the Xiaoxinganling Mountains may indicate a role as a local refuge.

However, for the seed populations in the CSO, the genetic population structure of the six populations became similar and clustered into just one group. This occurred due to the efficient mixing of gene pools among mother populations in the CSO. Thus, while seeds produced in the CSO have efficiently mixed gene pools resulting in increased diversity, but still maintaining regional differences in diversity, the primary geographically organized population differentiation, which may be in part due to regional adaptation, was wiped out.

4.3. Implications for forest management and conservation of Pinus koraiensis

Pinus koraiensis is a second-class protected precious timber species in China, it has formed rich genetic variation by long-term geographical isolation and natural selection (Feng et al., 2010; Park et al., 2017). At present, natural populations of Pinus koraiensis are declining due to habitat fragmentation and reduced population size entailing also long-term risks for its genetic resources. On the basis of the genetic diversity and population structure of Pinus koraiensis in this study, several strategies for more effectively protecting this species may be proposed: (1) The population genetics of the Korean pine should be studied and integrated with analysis techniques of molecular biology (e.g. transcriptome), physiology (e.g. rates of photosynthesis, water use efficiency, frost resistance) (Hofmann et al., 2015), and ecology to evaluate the ability to survive future changes and planning conservation strategies. (2) Strengthening the protection of existing natural forests by conserving their habitats, prohibiting timber harvest and installing sustainable management regimes are essential actions for maintaining the genetic diversity of this species. (3) In this paper, we have shown that the genetic diversity of the progeny population in the CSO is slightly increased compared with the maternal population, providing evidence that the gene pool of Pinus koraiensis as present in the CSO can be preserved by the ex situ conservation. Therefore, strengthening the ex situ conservation efforts and combining them with in situ conservation, can not only protects the Pinus koraiensis resources, but also enriches its genetic diversity. Lastly, however, our finding that Pinus koraiensis in China is divided into a Southern and a Northern gene pool suggests that the two clusters represent regionally adapted gene pools. Thus the question

Table 4

<table>
<thead>
<tr>
<th>mothers</th>
<th>LSH</th>
<th>HNH</th>
<th>DHL</th>
<th>DFH</th>
<th>DL</th>
<th>FL</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSH</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.011</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.015*</td>
<td>0.012</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.031</td>
<td>0.014</td>
<td>0.013</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.033*</td>
<td>0.023*</td>
<td>0.018*</td>
<td>0.032*</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.017*</td>
<td>0.014*</td>
<td>0.012</td>
<td>0.016*</td>
<td>0.019*</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.016</td>
<td>0.019</td>
<td>0.019</td>
<td>0.023</td>
<td>0.031</td>
<td>0.018</td>
<td>0.000</td>
</tr>
<tr>
<td>0.029</td>
<td>0.021</td>
<td>0.014</td>
<td>0.027</td>
<td>0.020</td>
<td>0.023</td>
<td>0.021*</td>
</tr>
<tr>
<td>0.020</td>
<td>0.019</td>
<td>0.011</td>
<td>0.018</td>
<td>0.019</td>
<td>0.018</td>
<td>0.014</td>
</tr>
<tr>
<td>0.014</td>
<td>0.020</td>
<td>0.018</td>
<td>0.016</td>
<td>0.035</td>
<td>0.019</td>
<td>0.011</td>
</tr>
<tr>
<td>0.021</td>
<td>0.021</td>
<td>0.017</td>
<td>0.025</td>
<td>0.020</td>
<td>0.016</td>
<td>0.009</td>
</tr>
<tr>
<td>0.029</td>
<td>0.029</td>
<td>0.023</td>
<td>0.028</td>
<td>0.027</td>
<td>0.019</td>
<td>0.014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>seeds</th>
<th>LSH</th>
<th>HNH</th>
<th>DHL</th>
<th>DFH</th>
<th>DL</th>
<th>FL</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSH</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.011</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.015*</td>
<td>0.012</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.031</td>
<td>0.014</td>
<td>0.013</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.033*</td>
<td>0.023*</td>
<td>0.018*</td>
<td>0.032*</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.017*</td>
<td>0.014*</td>
<td>0.012</td>
<td>0.016*</td>
<td>0.019*</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.016</td>
<td>0.019</td>
<td>0.019</td>
<td>0.023</td>
<td>0.031</td>
<td>0.018</td>
<td>0.000</td>
</tr>
<tr>
<td>0.029</td>
<td>0.021</td>
<td>0.014</td>
<td>0.027</td>
<td>0.020</td>
<td>0.023</td>
<td>0.021*</td>
</tr>
<tr>
<td>0.020</td>
<td>0.019</td>
<td>0.011</td>
<td>0.018</td>
<td>0.019</td>
<td>0.018</td>
<td>0.014</td>
</tr>
<tr>
<td>0.014</td>
<td>0.020</td>
<td>0.018</td>
<td>0.016</td>
<td>0.035</td>
<td>0.019</td>
<td>0.011</td>
</tr>
<tr>
<td>0.021</td>
<td>0.021</td>
<td>0.017</td>
<td>0.025</td>
<td>0.020</td>
<td>0.016</td>
<td>0.009</td>
</tr>
<tr>
<td>0.029</td>
<td>0.029</td>
<td>0.023</td>
<td>0.028</td>
<td>0.027</td>
<td>0.019</td>
<td>0.014</td>
</tr>
</tbody>
</table>

Fig. 1. Results of the STRUCTURE analysis of mother populations at K = 2. Each individual is represented by a single vertical bar, which is partitioned among gene pools. Colours represent genetic clusters and the coloured segments show the individual’s estimated ancestry proportion.

According to the genetic diversity parameters and the STRUCTURE analysis, the six mother populations can be divided into two groups, with the two northern populations in the Xiaoxinganling Mountains as one group displaying higher genetic diversity and the southern four populations in the Changbai Mountains and its adjacent highlands as another group with lower genetic diversity. These groups are separated by the large Heilong Jiang river valley which may have restricted gene flow between these geographic regions. While populations in the Xiaoxinganling Mountains have always been part of natural forests with low human disturbance, the southern region has long been subjected to strong human disturbance. In addition, the regions also differ climatically which may have resulted in adaptive divergence among these regions. It is important to mention that one of the nine genetic markers used are EST-SSR (P79), i.e. it is part of the expressed genome. Thus unlike random genomic SSRs, which are typically considered to be neutral (Shi et al., 2011; Vieira et al., 2016), the EST-SSR markers are more likely to be under selection. Thus our data suggest that adaptive divergence may be at least partially responsible for genetic divergence between the two gene pools.

Both adaptive divergence and differences in human intervention are unlikely to be responsible for differences in genetic variation between the two population groups. However, the maternal population genetic structure and level of diversity also reflects the glacial refugia and postglacial expansion processes to a certain extent. A comprehensive study of the phylogeography of Pinus koraiensis is lacking and both a multiple refugia hypothesis (Potenko and Velikov, 1998) and a single-refugium hypothesis with a latitudinal cline of genetic variation (Kim et al., 2005; Aizawa et al., 2012) have been stated. Our data suggest that a scenario of a single refugium with later northward expansion resulting in constant decrease of genetic variation from south to north likely is too simple, as has been found in many Chinese forest species

Fig. 2. Results of the STRUCTURE analysis of all populations at K = 2.
arises whether hybrid offspring with higher diversity and potential hybrid fitness effects, e.g. heterosis (Agrawal, 2009; Oakley et al., 2015), actually show higher fitness in the field compared to local seed that may have lower diversity, but is regionally/locally adapted and therefore may outperform foreign genotypes. To assess the relevance of the “Mix or Match” debate (Lesica and Allendorf, 1999; Bucharová et al., 2019), more extensive experimental research should be carried out.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by CFERN & BEIJING TECHNO SOLUTIONS Award Funds on excellent academic achievements and the Project of National Natural Science Foundation of China (NSFC 41877549 and 41701052). We would like to acknowledge Chengrui Mao, Yan Wang, Yang Tang, Hong Chen and all the people from the Forestry Bureau in Lushui Town that helped us with the sampling.

Author statement

All the co-authors have read and approved the submitted version of the manuscript, and all those entitled to authorship have been so named. The content of this manuscript is original and no other articles using this database have been submitted (or are going to be submitted in the next three months) elsewhere. All the materials, raw data, and protocols used in the article are available upon request and without any restriction, and will be published in a public repository if the article is accepted for publication.

References

Oakey, C.G., Agren, J., Schemske, D.W., 2015. Heterosis and outbreeding depression in
