Ex situ conservation of *Pinus koraiensis* can preserve the species' genetic diversity but homogenizes differentiated population structure.

Yuewei Tong[a,b], Walter Durka[c,d], Wangming Zhou[a], Li Zhou[a], Dapao Yu[a], Limin Dai[a*]

Affiliations:

[a] CAS Key Laboratory of Forest Ecology and Management, Institute of Applied Ecology, Shenyang 110016, China

[b] University of Chinese Academy of Sciences, Beijing 100049, China

[c] Department of Community Ecology (BZF), Helmholtz Centre for Environmental Research-UFZ, Halle 06120, Germany

[d] German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher Platz 5e, D-04103 Leipzig, Germany.

* Corresponding author: Limin Dai.

E-mail: lmdai@iae.ac.cn.

Abstract: *Pinus koraiensis* is a conifer species of ecological and economic importance in northeast China that has been excessively exploited in recent years. To enhance the evaluation, management, and conservation of this species, we applied 9 pairs of simple sequence repeat (SSR) primers to study the genetic diversity and population structure of six maternal populations as well as their progeny populations in a clonal seed orchard (CSO). 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* (F_{ST} = 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 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, although differences in diversity were 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.

Key words: *Pinus koraiensis*, *ex situ* conservation, clonal seed orchard, genetic diversity, population structure

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 (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, is defined as an artificial plantation derived from vegetatively propagated wild trees with multiple provenances and multiple clones per provenance (Fernandes *et al.*, 2008). Although the main goal of CSO is to maximize the genetic gain in economic and/or adaptive traits, it is still important to consider its genetic diversity, because the CSO represents 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;

57 Chaloupkova *et al.*, 2019). Generally, the genetic diversity of seeds produced in a CSO is determined
58 by that of provenances and parent trees which, however, was unknown during the establishment.
59 Moreover, the diversity of seeds will depend on planting design (Yuan *et al.*, 2016) and the pollen
60 dispersal distance. Although most forest temperate forest trees are wind pollinated, pollen dispersal
61 distance is limited and may be non-random, e.g. due to preferential wind direction (Feng *et al.*, 2010).
62 If provenances are genetically differentiated and seeds are produced by random mating in a seed
63 orchard, higher genetic diversity is expected for seeds compared with the average source provenance.
64 However, while mixing genetically differentiated populations will increase genetic diversity,
65 potentially with hybrid vigor, this is connected with a potential risk of outbreeding depression (hybrid
66 breakdown) by disrupting coadapted gene complexes if provenances are adapted to different
67 environmental conditions (Johansen-Morris and Latta, 2006). Therefore, as CSOs are a tool for
68 breeding of numerous tree species around the world, studying the genetic diversity of maternal
69 provenances and offspring produced in a CSO can provide a reliable basis for evaluating the potential
70 risks in the management of seed orchards and plantations (El-Kassaby and Ritland, 1996; Nielsen and
71 Hansen, 2012; Sonstebo *et al.*, 2018).

72 *Pinus koraiensis* Siebold et Zucc. (Korean pine, Pinaceae), is one of the most ecologically and
73 economically important conifer species in the mountainous area of Northeastern China (Ma *et al.*,
74 1992). It is a dominant forest tree species in its habitat and produces high-quality timber and edible
75 pine nuts (Chen *et al.*, 2010; Aizawa *et al.*, 2012). In recent decades, the populations of natural broad-
76 leaved Korean pine mixed forests have been sharply reduced (Yu *et al.*, 2011), and the species has been
77 listed as nationally endangered in China (http://www.plant.csdb.cn/endangeredplants). Thus, the
78 protection of *P. koraiensis* has become urgent, especially for the genetic resources. There are several
79 hypotheses about the origin of Korean pine populations in continental Asia ranging from a single
80 glacial refugium, substantiated by the genetically depauperate plastid genomes that lack variation
81 (Aizawa *et al.*, 2012) to the hypothesis of several continental *in situ* relic populations (Potenko and
82 Velikov, 1998). For the nuclear genome of *P. koraiensis*, a gradient of genetic variation decreasing
83 from South (Korea) to North (Russia) (Kim *et al.*, 2005) indicates however that populations are not
84 homogenous and likely display genetic structure.

85 In the 1980s, a breeding program to protect the genetic resources of *P. koraiensis* was started in China
86 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) to 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℃ 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. 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 16.33 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. This work is guided on "Observation Methodology for Long-term Forest Ecosystem Research" of National Standards of the People's Republic of China (GB/T 33027-2016).

We selected six provenances, henceforward 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 (henceforward 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 1 Names, population ID, and original site coordinates of *Pinus koraiensis* populations analysed.

Population	ID	Latitude (°)	Longitude (°)	Annual Mean Temperature (°C)	Annual Mean Precipitation (mm)
Lushuihe	LSH	42.53	127.80	4.12	783.68
Huangnihe	HNH	43.55	128.01	3.76	635.42

Dahailin	DHL	44.52	128.86	1.14	619.72
Dongfanghong	DFH	46.58	133.58	3.17	648.99
Dailin	DL	47.18	128.85	0.30	574.59
Fenglin	FL	48.13	129.19	0.04	596.15

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)), 0.1mM 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

Genetic diversity indices

We calculated a set of population genetic diversity estimators using GeneAlEx 6.5.1 (Peakall and Smouse, 2012). The observed (N_a) and effective (N_e) number of alleles, the observed (H_o) and expected (H_e) 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_r) and the inbreeding coefficient F_{IS} (Weir and Cockerham, 1984) in FSTAT 2.9.3 (Goudet, 1995).

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) (1,000 permutations) based on the degree of genetic divergence among populations was performed using GeneAlEx 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 was 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).

Statistical analysis

Mean values of genetic diversity descriptors were compared among sample types or populations by T-test in R (Team, 2013). We tested for clinal variation of genetic diversity descriptors with latitude with 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 (N_a) with an average of 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 (Tab. 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 (Tab. 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 N_a, I, H_e, uH_e and A_r (Tab. 2). When testing for clinal variation with latitude, N_a, I and A_r showed increasing values towards North (Tab. 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_e, uH_e, I: $p<0.05$; N_a, A_r: $p<0.1$). For all diversity descriptors values where higher in seeds than in mothers, significantly so in N_a, I, and A_r. The spatial patters 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 N_a and A_r. Clinal variation with latitude was only marginally significant for the same parameters as in mothers (N_a, I, A_r). The inbreeding coefficient was significantly negative in four seed populations, but F_{IS} was not correlated between mothers and seeds.

Table 2 Genetic diversity parameters for the mother and seed populations of six *P. koraiensis* analyzed at nine microsatellite loci: mean number of alleles (N_a), mean number of effective alleles (N_e), Shannon's index (I), observed heterozygosity (H_o), expected heterozygosity (H_e), unbiased expected heterozygosity (uH_e), allelic richness (A_r) and inbreeding coefficient (F_{IS}). In addition, correlation coefficients and p-value of correlation with latitude and mean values for mother and seed populations, and p-value of an ANOVA testing for differences between northern and southern populations as well as mothers and seeds.

Population	**Type**	N_a	N_e	I	H_o	H_e	uH_e	A_r	F_{is}
LSH	mother	5.2	3.3	1.212	0.649	0.603	0.616	5.222	-0.055
HNH	mother	5.2	3.2	1.210	0.680	0.610	0.622	5.222	-0.095 *
DHL	mother	5.3	3.3	1.220	0.693	0.612	0.624	5.333	-0.113 *
DFH	mother	5.3	3.7	1.235	0.600	0.593	0.605	5.333	0.009
DL	mother	5.8	3.6	1.333	0.720	0.638	0.652	5.778	-0.108 *
FL	mother	5.8	3.4	1.319	0.716	0.643	0.656	5.778	-0.092 *
Correlation mother vs. latitude	**r**	0.854	0.677	0.852	0.304	0.628	0.626	0.866	0.014
	p	0.030	0.139	0.031	0.558	0.181	0.184	0.026	0.979
ANOVA N vs. S	**p**	<0.001	0.520	<0.001	0.114	0.006	0.005	<0.001	0.422
LSH	seed	6.3	3.6	1.305	0.707	0.622	0.635	6.333	-0.116 *
HNH	seed	6.4	3.3	1.279	0.640	0.623	0.636	6.444	-0.007
DHL	seed	6.1	3.8	1.320	0.644	0.633	0.646	6.111	0.002
DFH	seed	6.6	3.6	1.303	0.667	0.612	0.625	6.556	-0.069 *
DL	seed	6.8	3.6	1.366	0.724	0.646	0.660	6.778	-0.100 *
FL	seed	6.7	3.5	1.364	0.782	0.657	0.671	6.667	-0.170 *

Correlation seed vs. latitude	**r**	0.792	0.06	0.772	0.609	0.616	0.623	0.732	0.509
	p	0.061	0.910	0.072	0.199	0.193	0.187	0.098	0.302
ANOVA N vs. S	**p**	0.065	0.882	0.008	0.038	0.016	0.014	0.069	0.135
Mean	**mother**	**5.4**	**3.4**	**1.255**	**0.676**	**0.617**	**0.629**	**5.444**	**-0.076**
Mean	**seed**	**6.5**	**3.6**	**1.323**	**0.694**	**0.632**	**0.645**	**6.481**	**-0.077**
Correlation mother-seed	**r**	0.773	0.273	0.937	0.440	0.978	0.977	0.744	-0.142
	p	0.071	0.600	0.006	0.382	<0.001	<0.001	0.090	0.789
ANOVA mothers vs. seeds	**p**	<0.001	0.178	0.031	0.557	0.170	0.165	<0.001	0.976

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 (Table 4, 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.

Table 3 Non-hierarchical AMOVA for all, mother and seed populations of *P. koraiensis*, and a hierarchical AMOVA with mothers and seeds as groups for all populations, as well as with the two northern and four southern populations as two geographic groups for mother populations and seed populations; significant values are indicated with * ($P < 0.05$).

Populations	Source of variation	df	Sum of square	Variance components	Percentage of variation
All	Among Pops	11	122.680	0.235	4.25% *
	Within Pops	288	1522.160	5.285	95.75% *
	Total	299	1644.840	5.520	100%
Mother	Among Pops	5	49.707	0.189	3.49% *
	Within Pops	144	752.000	5.222	96.51% *
	Total	149	801.707	5.411	100%
Seed	Among Pops	5	42.293	0.124	2.27% *
	Within Pops	144	770.160	5.348	98.73% *
	Total	149	812.453	5.473	100%
All	Between groups: mothers vs. seeds	1	30.680	0.143	2.56% *
	Among Pops	10	92.000	0.157	2.80% *
	Within Pops	288	1522.160	5.285	94.63% *
	Total	299	1644.840	5.585	100.00%
Mother	Between groups: N vs. S	1	16.857	0.130	2.37% *
	Among pops	4	32.850	0.120	2.19% *
	Within pops	144	752.000	5.222	95.44% *
	Total	149	801.707	5.471	100.00%
Seed	Between groups:	1	8.883	0.008	0.15%

N vs. S				
Among pops	4	33.410	0.120	2.19% *
Within pops	144	770.160	5.348	97.66% *
Total	149	812.453	5.476	100.00%

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 3).

Table 4 Pairwise F_{ST} for all populations of *P. koraiensis*, significant values are indicated with * ($P < 0.05$).

mothers						seeds						
LSH	HNH	DHL	DFH	DL	FL	LSH	HNH	DHL	DFH	DL	FL	
0.000												LSH
0.011	0.000											HNH
0.015 *	0.012	0.000										DHL
0.011	0.014	0.013	0.000									DFH
0.033 *	0.023 *	0.018 *	0.032 *	0.000								DL
0.017 *	0.014 *	0.012	0.016 *	0.019 *	0.000							FL
0.016	0.019	0.019	0.023	0.031	0.018	0.000						LSH
0.029	0.021	0.014	0.027	0.020	0.023	0.021 *	0.000					HNH
0.020	0.019	0.011	0.018	0.019	0.018	0.014	0.013	0.000				DHL
0.014	0.020	0.018	0.016	0.035	0.019	0.011	0.028 *	0.015	0.000			DFH
0.021	0.021	0.017	0.025	0.020	0.016	0.009	0.017 *	0.012	0.013	0.000		DL
0.029	0.029	0.023	0.028	0.027	0.019	0.014	0.014 *	0.016 *	0.023 *	0.010	0.000	FL

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).

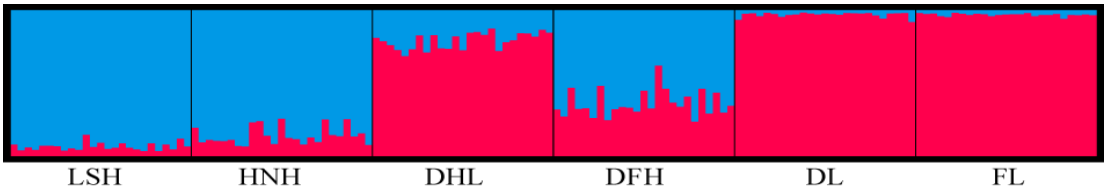

Fig. 1. Results of the STRUCTURE analysis of mother populations at K = 2. Each individual is represented by a single vertical bar, which is partitioned among gene pools. Colours represent genetic clusters and the coloured segments show the individual's estimated ancestry proportion.

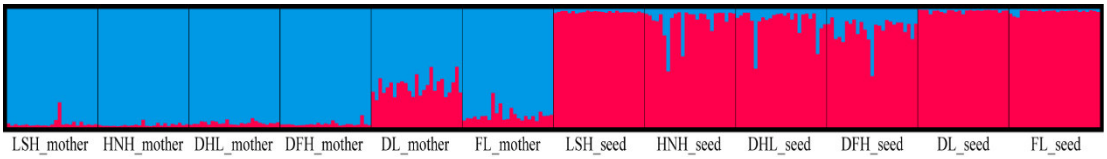

Fig. 2. Results of the STRUCTURE analysis of all populations at K = 2.

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_e 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 Daxinganling mountains maintained an increased level of diversity.

Secondly, the level of genetic diversity of the seed populations in our study was actually higher than that of the mother populations. The number of alleles increased on average across loci and populations by one allele from 5.4 to 6.5 and H_e increased from 0.617 to 0.632. Thus, the progeny populations not only maintain but actually surpass the maternal level of genetic diversity. This confirms the hypothesis that the *ex situ* conservation can preserve the species' gene pool, provided that a diverse wild gene pool is encompassed. In contrast, both in natural forests and orchards, usually a constant genetic diversity was 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 Liebst, 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) according to Wright's study (Wright, 1965), which is consistent with the findings of most species with outcrossing breeding system and wind pollination (Belletti *et al.*, 2012; Iwaizumi *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 (Miyaki, 1987). Such pollination and seed dispersal syndromes are beneficial to increase gene flow within and among populations (Nybom, 2004; Petit *et al.*, 2005), and consequently to reduce genetic differentiation among populations.

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 Daxinganling 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 Daxinganling 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 (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. More complex scenarios have to be considered and the high diversity detected in the Daxinganling 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 *P. 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 *P. 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 *P. 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 *P. koraiensis* resources, but also enriches its genetic diversity. Lastly, however, our finding that *P. 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 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 therefor may outperform foreign genotypes. To assess the relevance of the "Mix or Match" debate (Lesica and Allendorf, 1999; Bucharova *et al.*, 2019), more extensive experimental research should be carried out.

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 Lushuihe Town that helped us with the sampling.

References

Agrawal, A.F., 2009. Spatial Heterogeneity and the Evolution of Sex in Diploids. American Naturalist. 174, S54-S70. https://doi.org/10.1086/599082

Ai, C., Xu, L.A., Lai, H.L., Huang, M.R., Wang, Z.R., 2006. Genetic diversity and paternity analysis of a seed orchard in *Pinus massoniana*. Scientia Silvae Sinicae. 42, 146-150. (In Chinese).

Aizawa, M., Kim, Z.S., Yoshimaru, H., 2012. Phylogeography of the Korean pine (*Pinus koraiensis*) in northeast Asia: inferences from organelle gene sequences. Journal of Plant Research. 125, 713-723. https://doi.org/10.1007/s10265-012-0488-4

Assessment, M.E., 2005. Ecosystems and Human Well-being: Synthesis. Island Press, Washington, DC

Belletti, P., Ferrazzini, D., Piotti, A., Monteleone, I., Ducci, F., 2012. Genetic variation and divergence in Scots pine (*Pinus sylvestris* L.) within its natural range in Italy. European Journal of Forest Research. 131, 1127-1138. https://doi.org/10.1007/s10342-011-0584-3

Bernhardsson, C., Floran, V., Ganea, S.L., Garcia-Gil, M.R., 2016. Present genetic structure is congruent with the common origin of distant Scots pine populations in its Romanian distribution. Forest Ecology and Management. 361, 131-143. https://doi.org/10.1016/j.foreco.2015.10.047

Bucharova, A., Bossdorf, O., Hoelzel, N., Kollmann, J., Prasse, R., Durka, W., 2019. Mix and match: regional admixture provenancing strikes a balance among different seed-sourcing strategies for ecological restoration. Conservation Genetics. 20, 7-17. https://doi.org/10.1007/s10592-018-1067-6

Chaloupkova, K., Stejskal, J., El-Kassaby, Y.A., Frampton, J., Lstiburek, M., 2019. Current advances in seed orchard layouts: Two case studies in conifers. Forests. 10, 93. https://doi.org/10.3390/f10020093

Chen, M.M., Feng, F.J., Sui, X., Li, M.H., Zhao, D., Han, S.J., 2010. Construction of a framework map for *Pinus koraiensis* Sieb. et Zucc. using SRAP, SSR and ISSR markers. Trees-Structure and Function. 24, 685-693. https://doi.org/10.1007/s00468-010-0438-5

Dawson, I.K., Guariguata, M.R., Loo, J., Weber, J.C., Lengkeek, A., Bush, D., Cornelius, J., Guarino, L., Kindt, R., Orwa, C., Russell, J., Jamnadass, R., 2013. What is the relevance of smallholders' agroforestry systems for conserving tropical tree species and genetic diversity in circa situm, in situ and ex situ settings? A review. Biodiversity and Conservation. 22, 301-324. https://doi.org/10.1007/s10531-012-0429-5

Durka, W., Michalski, S.G., Berendzen, K.W., Bossdorf, O., Bucharova, A., Hermann, J.-M., Hoelzel, N., Kollmann, J., 2017. Genetic differentiation within multiple common grassland plants supports seed transfer zones for ecological restoration. Journal of Applied Ecology. 54, 116-126. https://doi.org/10.1111/1365-2664.12636

El-Kassaby, Y.A., Ritland, K., 1996. Genetic variation in low elevation Douglas-fir of British Columbia and its relevance to gene conservation. Biodiversity and Conservation. 5, 779-794. https://doi.org/10.1007/bf00051786

Evanno, G., Regnaut, S., Goudet, J., 2010. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology. 14, 2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Feng, F.J., Sui, X., Chen, M.M., Zhao, D., Han, S.J., Li, M.H., 2010. Mode of pollen spread in clonal seed orchard of *Pinus koraiensis*. Journal of Biophysical Chemistry. 1, 33-39. https://doi.org/10.4236/jbpc.2010.11004

Fernandes, L., Rocheta, M., Cordeiro, J., Pereira, S., Gerber, S., Oliveira, M.M., Ribeiro, M.M., 2008. Genetic variation, mating patterns and gene flow in a *Pinus pinaster* Aiton clonal seed orchard. Annals of Forest Science. 65. https://doi.org/10.1051/forest:2008049

Frankham, R., Ballou, J.D., Briscoe, D.A., 2002. Introduction to conservation genetics. Cambridge University Press, Cambridge, UK.

Geburek, T., Konrad, H., 2008. Why the conservation of forest genetic resources has not worked. Conservation Biology. 22, 267-274. https://doi.org/10.1111/j.1523-1739.2008.00900.x

Given, D.R., 1994. Principles and practice of plant conservation. Timber Press, Portland

Goudet, J., 1995. FSTAT (version 1.2): A computer program to calculate f-statistics. Journal of Heredity. 86, 485-486. https://doi.org/10.1093/oxfordjournals.jhered.a111627

Hofmann, M., Durka, W., Liesebach, M., Bruelheide, H., 2015. Intraspecific variability in frost hardiness of Fagus sylvatica L. European Journal of Forest Research. 134, 433-441. https://doi.org/10.1007/s10342-015-0862-6

Holliday, J.A., Aitken, S.N., Cooke, J.E.K., Fady, B., Gonzalez-Martinez, S.C., Heuertz, M., Jaramillo-Correa, J.-P., Lexer, C., Staton, M., Whetten, R.W., Plomion, C., 2017. Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. Molecular

Ecology. 26, 706-717. https://doi.org/10.1111/mec.13963
Hubisz, M.J., Falush, D., Stephens, M., Pritchard, J.K., 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources. 9, 1322-1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x
Iwaizumi, M.G., Miyata, S., Hirao, T., Tamura, M., Watanabe, A., 2018. Historical seed use and transfer affects geographic specificity in genetic diversity and structure of old planted *Pinus thunbergii* populations. Forest Ecology and Management. 408, 211-219. https://doi.org/10.1016/j.foreco.2017.10.026
Iwaizumi, M.G., Tsuda, Y., Ohtani, M., Tsumura, Y., Takahashi, M., 2013. Recent distribution changes affect geographic clines in genetic diversity and structure of *Pinus densiflora* natural populations in Japan. Forest Ecology and Management. 304, 407-416. https://doi.org/10.1016/j.foreco.2013.05.026
Jakobsson, M., Rosenberg, N.A., 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 23, 1801. https://doi.org/10.1093/bioinformatics/btm233
Jia, D., Zhen, Z., Zhang, H.G., Tang, J.H., 2017. EST-SSR marker development and transcriptome sequencing analysis of different tissues of Korean pine (*Pinus koraiensis* Sieb. et Zucc.). Biotechnology & Biotechnological Equipment. 31, 679-689. https://doi.org/10.1080/13102818.2017.1331755
Johansen-Morris, A.D., Latta, R.G., 2006. Fitness consequences of hybridization between ecotypes of Avena barbata: Hybrid breakdown, hybrid vigor, and transgressive segregation. Evolution. 60, 1585-1595. https://doi.org/10.1554/05-680.1
Kang, K.S., Lindgren, D., Mullin, T.J., 2001. Prediction of genetic gain and gene diversity in seed orchard crops under alternative management strategies. Theoretical and Applied Genetics. 103, 1099-1107. https://doi.org/10.1007/s001220100700
Kang, K.S., Lindgren, D., Mullin, T.J., Choi, W.Y., Han, S.U., 2005. Genetic gain and diversity of orchard crops under alternative management options in a clonal seed orchard of *Pinus thunbergii*. Silvae Genetica. 54, 93-96. https://doi.org/10.1515/sg-2005-0014
Kim, Z.S., Hwang, J.W., Lee, S.W., Yang, C., Gorovoy, P.G., 2005. Genetic variation of Korean pine (*Pinus koraiensis* Sieb. et Zucc.) at allozyme and RAPD markers in Korea, China and Russia. Silvae Genetica. 54, 235-246. https://doi.org/10.1515/sg-2005-0034
Ledig, F.T., 1992. Human impacts on genetic diversity in forest ecosystems. Oikos. 63, 87-108. https://doi.org/10.2307/3545518
Lesica, P., Allendorf, F.W., 1999. Ecological genetics and the restoration of plant communities: Mix or match? Restoration Ecology. 7, 42-50. https://doi.org/10.1046/j.1526-100X.1999.07105.x
Litkowiec, M., Lewandowski, A., Wachowiak, W., 2018. Genetic variation in Taxus baccata L.: A case study supporting Poland's protection and restoration program. Forest Ecology and Management. 409, 148-160. https://doi.org/10.1016/j.foreco.2017.11.026
Loo, J., Souvannavong, O., Dawson, L.K., 2014. Seeing the trees as well as the forest: The importance of managing forest genetic resources. Forest Ecology and Management. 333, 1-8. https://doi.org/10.1016/j.foreco.2014.08.014
Loveless, M.D., Hamrick, J.L., 1984. Ecological determinants of genetic-structure in plant-populations. Annual Review of Ecology and Systematics. 15, 65-95. https://doi.org/10.1146/annurev.es.15.110184.000433
Ma, J.L., Zhang, L.W., Cheng, D., Li, J.W., 1992. Geographic distribution of *Pinus koraiensis* in the world. Journal of Northeast Forestry University. 20, 40-47. (In Chinese).
Mandak, B., Hadincova, V., Mahelka, V., Wildova, R., 2013. European Invasion of North American *Pinus strobus* at Large and Fine Scales: High Genetic Diversity and Fine-Scale Genetic Clustering over Time in the Adventive Range. Plos One. 8. https://doi.org/10.1371/journal.pone.0068514
Miyaki, M., 1987. Seed dispersal of the Korean pine, *Pinus koraiensis*, by the red squirrel, *Sciurus vulgaris*. Ecological Research. 2, 147-157. https://doi.org/10.1007/BF02346923
Moritz, C., 1999. Conservation units and translocations: strategies for conserving evolutionary processes. Hereditas. 130, 217–228. https://doi.org/10.1111/j.1601-5223.1999.00217.x
Nielsen, U.B., Hansen, O.K., 2012. Genetic worth and diversity across 18 years in a Nordmann fir clonal seed orchard. Annals of Forest Science. 69, 69-80. https://doi.org/10.1007/s13595-011-0159-y
Nybom, H., 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology. 13, 1143-1155. https://doi.org/10.1111/j.1365-294X.2004.02141.x
Oakley, C.G., Agren, J., Schemske, D.W., 2015. Heterosis and outbreeding depression in crosses between natural populations of Arabidopsis thaliana. Heredity. 115, 73-82. https://doi.org/10.1038/hdy.2015.18

Park, J.M., Kwon, S.H., Lee, H.J., Na, S.J., El-Kassaby, Y.A., Kang, K.S., 2017. Integrating fecundity variation and genetic relatedness in estimating the gene diversity of seed crops: *Pinus koraiensis* seed orchard as an example. Canadian Journal of Forest Research. 47, 366-370. https://doi.org/10.1139/cjfr-2016-0223
Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. Bioinformatics. 28, 2537-2539. https://doi.org/10.1093/bioinformatics/bts460
Petit, R.J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D., Vendramin, G.G., 2005. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Molecular Ecology. 14, 689-701. https://doi.org/10.1111/j.1365-294X.2004.02410.x
Potenko, V.V., Velikov, A.V., 1998. Genetic diversity and differentiation of natural populations of *Pinus koraiensis* (Sieb. et Zucc.) in Russia. Silvae Genetica. 47, 202-208. https://doi.org/0000-0002-8442-4825
Potter, K.M., Jetton, R.M., Bower, A., Jacobs, D.F., Man, G., Hipkins, V.D., Westwood, M., 2017. Banking on the future: progress, challenges and opportunities for the genetic conservation of forest trees. New Forests. 48, 153-180. https://doi.org/10.1007/s11056-017-9582-8
Prat, D., Arnal, S., 1994. Allozyme variation and mating system in 3 artificial stands of Douglas-fir (*Pseudotsuga-menziesii* (mirb) franco) planted in Europe. Silvae Genetica. 43, 199-206.
Pritchard, H.W., Moat, J.F., Ferraz, J.B.S., Marks, T.R., Camargo, J.L.C., Nadarajan, J., Ferraz, I.D.K., 2014. Innovative approaches to the preservation of forest trees. Forest Ecology and Management. 333, 88-98. https://doi.org/10.1016/j.foreco.2014.08.012
Rajora, O.P., Mosseler, A., 2001. Challenges and opportunities for conservation of forest genetic resources. Euphytica. 118, 197-212. https://doi.org/10.1023/a:1004150525384
Ratnam, W., Rajora, O.P., Finkeldey, R., Aravanopoulos, F., Bouvet, J.-M., Vaillancourt, R.E., Kanashiro, M., Fady, B., Tomita, M., Vinson, C., 2014. Genetic effects of forest management practices: Global synthesis and perspectives. Forest Ecology and Management. 333, 52-65. https://doi.org/10.1016/j.foreco.2014.06.008
Roberds, J.H., Conkle, M.T., 1984. Genetic-structure in loblolly - pine stands - Allozyme variation in parents and progeny. Forest Science. 30, 319-329.
Schneller, J., Liebst, B., 2007. Patterns of variation of a common fern (*Athyrium filix*-femina; Woodsiaceae): Population structure along and between altitudinal gradients. American Journal of Botany. 94, 965-971. https://doi.org/10.3732/ajb.94.6.965
Schwartz, K.R., Parsons, E.C.M., Rockwood, L., Wood, T.C., 2017. Integrating *in-situ* and *ex-situ* data management processes for biodiversity conservation. Frontiers in Ecology and Evolution. 5, 120. https://doi.org/10.3389/fevo.2017.00120
Sonstebo, J.H., Tollefsrud, M.M., Myking, T., Steffenrem, A., Nilsen, A.E., Edvardsen, O.M., Johnskas, O.R., El-Kassaby, Y.A., 2018. Genetic diversity of Norway spruce (*Picea abies* (L.) Karst.) seed orchard crops: Effects of number of parents, seed year, and pollen contamination. Forest Ecology and Management. 411, 132-141. https://doi.org/10.1016/j.foreco.2018.01.009
Team, R.C., 2013. R: A language and environment for statistical computing.
Toth, E.G., Vendramin, G.G., Bagnoli, F., Cseke, K., Hoehn, M., 2017. High genetic diversity and distinct origin of recently fragmented Scots pine (Pinus sylvestris L.) populations along the Carpathians and the Pannonian Basin. Tree Genetics & Genomes. 13. https://doi.org/10.1007/s11295-017-1137-9
Vieira, M.L.C., Santini, L., Diniz, A.L., Munhoz, C.d.F., 2016. Microsatellite markers: what they mean and why they are so useful. Genetics and Molecular Biology. 39, 312-328. https://doi.org/10.1590/1678-4685-gmb-2016-0027
Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution. 38, 1358-1370. https://doi.org/10.2307/2408641
Wright, S., 1965. The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. Evolution. 19, 395-420.
Xu, Y.L., Cai, N.H., Woeste, K., Kang, X.Y., He, C.Z., Li, G.Q., Chen, S., Duan, A.A., 2016. Genetic diversity and population structure of *Pinus yunnanensis* by simple sequence repeat markers. Forest Science. 62, 38-47. https://doi.org/10.5849/forsci.15-064
Yu, D.P., Zhou, L., Zhou, W.M., Ding, H., Wang, Q.W., Wang, Y., Wu, X.Q., Dai, L.M., 2011. Forest Management in Northeast China: History, Problems, and Challenges. Environmental Management. 48, 1122-1135. https://doi.org/10.1007/s00267-011-9633-4
Yu, J.H., Chen, C.M., Tang, Z.H., Yuan, S.S., Wang, C.J., Zu, Y.G., 2012. Isolation and characterization of 13 novel polymorphic microsatellite markers for *Pinus koraiensis* (Pinaceae). American Journal of Botany. 99, 421-424. https://doi.org/10.3732/ajb.1200145
Yuan, H., Niu, S., El-Kassaby, Y.A., Li, Y., Li, W., 2016. Simple Genetic Distance-Optimized Field

511 Deployments for Clonal Seed Orchards Based on Microsatellite Markers: As a Case of Chinese Pine
512 Seed Orchard. Plos One. 11. https://doi.org/10.1371/journal.pone.0157646

513