Population Genetic Analysis of Pinus Koraiensis in China Inferred Using EST-SSR Markers

Xiang Li  
State Key Laboratory of Tree Genetics and Breeding, School of Forestry, Northeast Forestry University

Minghui Zhao  
Stat Key Laboratory of Tree Genetics and breeding, School of Forestry, Northeast Forestry University

Yujin Xu  
State Key Laboratory of Tree Genetics and Breeding, School of Forestry, Northeast Forestry University

Yan Li  
State Key Laboratory of Tree Genetics and Breeding, School of Forestry, Northeast Forestry University

Mulalem Tigabu  
Southern Swedish Forest Research Centre, Swedish University of Agricultural Sciences

Xiyang Zhao (✉ zhaoxyphd@163.com)  
State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University  https://orcid.org/0000-0002-1744-5431

Research article

Keywords: Pinus koraiensis, EST-SSRs, genetic diversity, population structure, population differentiation, gene flow, China

DOI: https://doi.org/10.21203/rs.3.rs-117881/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background

*Pinus koraiensis* (commonly known as Korean pine), is a well-known conifer species in China with high economic, ornamental and ecological values. More than 50% of the *P. koraiensis* forests in the world are distributed in Northeast China, a region with abundant germplasm resources. However, these natural *P. koraiensis* populations are in danger of genetic erosion caused by continuous climate changes and frequent human activity. Little work has been conducted on the population genetic structure and genetic differentiation of *P. koraiensis* in China. Here, representative individuals from 16 natural *P. koraiensis* populations were sampled and genotyped, and polymorphic expressed sequence tag-simple sequence repeat (EST-SSR) markers were used to comprehensively evaluate genetic diversity, population structure and differentiation of *P. koraiensis* populations in China.

Results

A total of 480 samples from 16 populations were collected in the natural distribution area of *P. koraiensis*. Analysis of molecular variance (AMOVA) of the EST-SSR marker data showed that 33% of the total genetic variation was among populations and 67% was within populations. A high level of genetic diversity was found across *P. koraiensis* populations (average Na=10.33, Ne=2.514, He=0.521), and the highest levels of genetic diversity were found in Heihe (He=0.449), Zhanhe (He=0.413), Liangshui (He=0.370) and Tieli (He=0.414) populations. Moreover, pairwise Fst values revealed significant genetic differentiation among populations (mean Fst=0.177). Structure and Neighbor-joining (NJ) tree analyses and principal component analysis (PCA) revealed two genetic clusters: cluster 1 from Xiaoxinganling Mountains and cluster 2 from Changbaishan Mountains, which were consistent with the geographical distributions of the natural populations.

Conclusions

The findings provide new genetic information for future genome-wide association studies (GWAS), marker-assisted selection (MAS) and genomic selection (GS) in natural *P. koraiensis* breeding programs and can aid the development of conservation strategies for this valuable conifer species.

Background

*Pinus koraiensis* (Sieb. et Zucc.), commonly known as Korean pine, is a perennial evergreen tree in the Pinaceae family and has five needles per fascicle [1–3]. It is an ancient and valued forest tree in China, and natural forests of this species have undergone long-term succession and are described as tertiary forest [4]. Compared with other Pinus species, such *Pinus tabulaeformis* and *Pinus sylvestris*, *P. koraiensis* is long-lived, living to more than 500 years, and it is a dominant species in mixed conifer and broadleaved forest [5–6]. Currently, *P. koraiensis* is distributed mainly in cool-temperate regions in Northeast China, Russian Far East, the Korea peninsula and Honshu, Japan. It typically occurs in mild regions with humidity higher than 70% and at altitudes from 600 m to 1500 m [7–8]. However, in China, it only grows from Changbai Mountains to Xiaoxinganling Mountains in Northeast China, mainly on slopes and rolling hills and in river valleys [9]. Nearly half of the germplasm resources of *P. koraiensis* in the world are in Xiaoxinganling Mountains in Yichun city China. This region contains the largest and most undisturbed primeval forest remaining in Asia and a natural climax community of *P. koraiensis* [10].

*P. koraiensis* not only has high economic value in terms of wood and pine seed production and ornamental value in northeastern China but also is considered to have high ecological value for soil and water conservation and diversity conservation in forest ecosystems. A significant and traditional industrial timber species, *P. koraiensis* is widely used for architecture, bridges, furniture and ships because of the light, soft, fine structure and straight texture of the wood and its strong corrosion resistance [11]. Furthermore, it has economic value because of nut production of its edible, nutritious and distinctly favored pine nuts, which contain abundant unsaturated fatty acids, vitamins and minerals [12]. It also has high medicinal value, able to lower cholesterol levels and alloy ultraviolet injury and tiredness. Natural *P. koraiensis* forest absorbs large amounts of carbon dioxide and releases large amounts of oxygen and anions [13]. Therefore, it is a prominent conifer tree species of great value for the maintenance and protection of the environment in China.

Genetic improvement of *P. koraiensis* began in the 1960s and developed slowly due to a lack of systematic breeding strategies and objectives [8]. In the early stages of selective breeding, large numbers of superior trees or natural populations were selected from natural forest to establish primary seed orchards, mainly through phenotype selection [14–16]. Studies have mainly focused on propagation technology [17], provenance division [18], progeny determination [19] and selection of improved varieties [20], while studies on molecular plant breeding, including studies of genetic diversity, genomic selection and construction of genetic maps, remain lacking [21]. Furthermore, the current management, utilization and protection of natural *P. koraiensis* forest resources are not ideal. Existing natural forest resources of this species are of great significance for the conservation of breeding resources, the development of gene resources and the study of population genetic diversity. However, in the past few decades, with the increasing demand for wood and cones of *P. koraiensis*, the area of natural *P. koraiensis* forest has decreased extensively, severely restricting the development, utilization and protection of germplasm resources [22]. Thus, to protect existing natural forest resources under the background of illegal logging and unpredictable biotic stress, the collection and evaluation of germplasm resources of *P. koraiensis* are urgently needed. Genetic diversity and population structure are key parameters of population genetics research. Analyses of genetic variation among and within populations can guide the formulation of conservation strategies. The use of molecular markers identified from whole-genome, chloroplast genome and transcriptome analysis is a primary method of revealing genetic diversity and population structure. However, because of the lack of genetic data, few studies of molecular marker-assisted breeding and population genetics in *P. koraiensis* have been conducted.
Unlike morphological and biochemical markers, many DNA molecular markers are codominant and highly polymorphic, and many have been identified in the genome and transcriptome [23–25]. Because of their low cost, easy detection by Polymerase Chain Reaction (PCR), high polymorphism, and codominance, simple sequence repeats (SSRs) are considered powerful and advantageous molecular tools for genetic diversity analysis, genome-wide association analysis, core collections and genetic linkage map construction in many plants and animals [26–28]. Furthermore, multiple EST-SSR markers can easily be developed from microsatellite loci of public transcriptome data. At present, there are few reports of analyses of genetic diversity in *P. koraiensis* based on DNA molecular markers. Studies to date have employed random amplified polymorphic DNA (RAPD) analysis [29], single primer amplification reaction (SPAR) [30], intersimple sequence repeat (ISSR) analysis [31–32] and expressed sequence tag-simple sequence repeat (EST-SSR) analysis [33]. All of these studies have identified high levels of genetic diversity in *P. koraiensis*, with the greatest levels of genetic differentiation occurring within populations. However, a systematic and comprehensive population genetic study is lacking because of the limitations of the species’ geographical distribution, the number of available molecular markers and population size. Therefore, widespread germplasm collection and abundant polymorphic markers are necessary to study the genetic relationships and diversity of *P. koraiensis*, which can be facilitated by abundant molecular markers developed from high-throughput sequencing, especially for some endangered and non-model plant species.

In this study, we collected germplasm resources from 480 individuals of 16 natural populations of *P. koraiensis* within the species main distribution area in northeastern China, and we analyzed genetic diversity using 15 EST-SSRs. We aimed to (1) investigate genetic variation using polymorphic EST-SSRs, (2) evaluate the genetic diversity and structure of natural populations, (3) conduct a comprehensive, range-wide genetic diversity study of *P. koraiensis* in China, and (4) propose a protection strategy. The results provide insight into the conservation of this species and lay a foundation for further studies of marker-assisted selection (MAS) and genomic selection (GS) in *P. koraiensis* for genetic improvement.

**Results**

**Genetic diversity at different loci among populations**

The genetic diversity analysis was performed on 480 individuals from 16 natural *P. koraiensis* populations using 15 EST-SSRs markers (Table 2). The allele size ranged from 151 bp at locus NEPK-65 to 301 bp at loci NEPK-168 and NEPK-184. In total, 155 alleles across all 15 loci were detected in the sampled individuals; the number of alleles per locus ranged from 4 (NEPK-67) to 21 (NEPK-145), with a mean value of 10.33. We identified 58 alleles as private alleles, accounting for 37.42% of the alleles. The effective number of alleles (Ne) ranged from 1.170 at locus NEPK-40 to 6.605 at locus NEPK-145, with an average of 2.514 per locus. Observed and expected heterozygosity ranged from 0.008 to 0.984 and from 0.145 to 0.849, respectively, with mean values of 0.374 and 0.521, respectively. Polymorphism information content (PIC) varied from 0.142 (NEPK-40) to 0.833 (NEPK-145), with a mean value of 0.461.

### Table 1

**Description of *Pinus koraiensis* sampling locations in China.**

| Group | Location | Sample size | Codes | Population codes | Latitude (N) | Longitude (E) | Average altitude (m) | Province |
|-------|----------|-------------|-------|------------------|--------------|---------------|----------------------|----------|
| G1    | Zhanhe   | 30          | ZH    | P6               | 48°1’23"     | 127°0’56.16"  | 358                  | H        |
| G1    | Heihe    | 30          | HH    | P4               | 49°24’47"    | 126°36’47.16"  | 131                  | H        |
| G1    | Liangshui| 30          | LS    | P5               | 47°43’48"    | 128°55’12"    | 231                  | H        |
| G1    | Tieli    | 29          | TL    | P7               | 47°58’48"    | 128°4’48"     | 213                  | H        |
| G1    | Hegang   | 30          | HG    | P8               | 47°21’29"    | 129°33’50.03"  | 63                   | H        |
| G2    | Liangzihe| 30          | LZH   | P1               | 47°1’54"     | 129°41’49.92"  | 500                  | H        |
| G2    | Lushuihe | 30          | LSH   | P3               | 42°24’00"    | 127°28’59.88"  | 732                  | J        |
| G2    | Linjiang | 30          | LJ    | P9               | 42°0’36"     | 127°13’12"    | 342                  | J        |
| G2    | Helong   | 30          | HL    | P2               | 42°32’47"    | 129°1’3.72"    | 452                  | J        |
| G2    | Wangqing | 31          | WQ    | P16              | 43°19’26"    | 129°45’14.04"  | 232                  | J        |
| G2    | Muleng   | 30          | ML    | P13              | 43°49’54"    | 129°45’19.08"  | 410                  | J        |
| G2    | Jiahe    | 30          | JH    | P10              | 44°0’36"     | 127°25’12"    | 196                  | J        |
| G2    | Maoershan| 30          | MES   | P14              | 45°16’22"    | 127°30’14.40"  | 536                  | H        |
| G2    | Hulin    | 30          | HLQY  | P11              | 45°46’12"    | 132°58’48"    | 84                   | H        |
| G2    | Boli     | 30          | BL    | P12              | 45°42’00"    | 130°25’12"    | 525                  | H        |
| G2    | Fangzheng| 30          | FZ    | P15              | 45°49’48"    | 128°48’00"    | 111                  | H        |
Table 2
Characteristics of the 15 polymorphic EST-SSR markers used in this study.

| Locus     | Allele size range (bp) | Na | Ne  | I   | Ho  | He   | PIC | HWE  | NRA | Fis | Fit  | Fst  | Nm   |
|-----------|------------------------|----|-----|-----|-----|------|-----|------|-----|-----|------|------|------|
| NEPK-218  | 196–230                | 12 | 2.080 | 0.949 | 0.056 | 0.519 | 0.443 | *** | 6 | 0.647 | 0.893 | 0.695 | 0.109 |
| NEPK-40   | 196–238                | 13 | 1.170 | 0.399 | 0.063 | 0.145 | 0.142 | *** | 3 | 0.327 | 0.593 | 0.396 | 0.382 |
| NEPK-32   | 206–224                | 5  | 2.431 | 0.979 | 0.618 | 0.589 | 0.503 | *** | 1 | -0.236 | -0.053 | 0.148 | 1.438 |
| NEPK-53   | 197–218                | 5  | 1.902 | 0.697 | 0.008 | 0.474 | 0.367 | *** | 3 | 0.923 | 0.981 | 0.757 | 0.080 |
| NEPK-65   | 151–259                | 12 | 2.257 | 1.068 | 0.195 | 0.557 | 0.488 | *** | 5 | 0.259 | 0.649 | 0.526 | 0.225 |
| NEPK-71   | 215–239                | 7  | 2.013 | 0.772 | 0.635 | 0.503 | 0.392 | *** | 3 | -0.275 | -0.254 | 0.016 | 15.141 |
| NEPK-117  | 205–217                | 9  | 3.995 | 1.601 | 0.984 | 0.750 | 0.714 | *** | 2 | -0.594 | -0.290 | 0.191 | 1.058 |
| NEPK-72   | 222–234                | 7  | 1.649 | 0.637 | 0.043 | 0.394 | 0.324 | *** | 4 | 0.346 | 0.894 | 0.838 | 0.048 |
| NEPK-67   | 226–232                | 4  | 2.802 | 1.067 | 0.479 | 0.643 | 0.567 | *** | 1 | 0.034 | 0.259 | 0.233 | 0.822 |
| NEPK-38   | 229–255                | 15 | 1.211 | 0.498 | 0.096 | 0.174 | 0.172 | *** | 6 | 0.396 | 0.477 | 0.134 | 1.622 |
| NEPK-145  | 251–289                | 21 | 6.605 | 2.243 | 0.637 | 0.849 | 0.833 | *** | 6 | 0.229 | 0.309 | 0.104 | 2.154 |
| NEPK-168  | 241–301                | 18 | 3.550 | 1.602 | 0.279 | 0.718 | 0.674 | *** | 9 | 0.265 | 0.602 | 0.459 | 0.295 |
| NEPK-181  | 275–289                | 10 | 2.039 | 0.884 | 0.165 | 0.509 | 0.426 | *** | 5 | 0.375 | 0.667 | 0.467 | 0.285 |
| NEPK-213  | 276–298                | 12 | 1.846 | 1.039 | 0.421 | 0.458 | 0.435 | *** | 4 | -0.151 | 0.109 | 0.226 | 0.856 |
| NEPK-184  | 291–301                | 5  | 2.152 | 0.856 | 0.937 | 0.535 | 0.430 | *** | 0 | -0.743 | -0.719 | 0.014 | 17.691 |
| Average   |                        | 10.33 | 2.514 | 1.019 | 0.374 | 0.521 | 0.461 | 3.867 | 0.12 | 0.341 | 0.347 | 2.667 |

Genetic diversity within Pinus koraiensis populations

The levels of genetic diversity in the 16 populations are shown in Table 3. Across the sampled populations, Na varied from 2.667 (Helong) to 4.467 (Tieli), with a mean value of 3.271, and Ne ranged from 1.586 (Jiaohe) to 2.257 (Linjiang), with a mean value of 1.870. The populations with the highest levels of genetic diversity were Heihe (Ne = 1.939, Ho = 0.340 and He = 0.439), Zhanhe (Ne = 2.009, Ho = 0.356 and He = 0.413), Liangshui (Ne = 1.914, Ho = 0.470 and He = 0.370) and Tieli (Ne = 1.663, Ho = 0.390 and He = 0.310), whereas those with the lowest levels were Jiaohe (Ne = 1.586, Ho = 0.293 and He = 0.275) and Helong (Ne = 1.663, Ho = 0.390 and He = 0.310). Zhanhe had not only the highest genetic diversity but also the largest number of private alleles, identifying it as a unique natural P. koraiensis population. The F value ranged from −0.235 to 0.325 among the populations, with a mean value of 0.02, indicating that there might exist a deficiency of heterozygosity in the natural P. koraiensis populations.
Table 3
Genetic diversity estimates for 16 *P. koraiensis* populations based on fifteen EST-SRRs.

| Population | Size | Na  | Ne  | I   | Ho  | He  | uHe | F (null) | NRA |
|------------|------|-----|-----|-----|-----|-----|-----|---------|-----|
| LZH        | 30   | 3.267 | 1.713 | 0.639 | 0.415 | 0.374 | 0.388 | -0.054 | 7   |
| HL         | 30   | 2.667 | 1.663 | 0.529 | 0.390 | 0.310 | 0.321 | -0.195 | 2   |
| LSH        | 30   | 3.133 | 1.829 | 0.628 | 0.394 | 0.350 | 0.359 | 0.015  | 1   |
| HH         | 30   | 3.133 | 1.939 | 0.746 | 0.430 | 0.394 | 0.449 | 0.325  | 3   |
| LS         | 30   | 3.333 | 1.914 | 0.664 | 0.470 | 0.370 | 0.380 | -0.235 | 6   |
| ZH         | 30   | 4.133 | 2.009 | 0.782 | 0.413 | 0.422 | 0.422 | 0.015  | 12  |
| TL         | 30   | 4.467 | 2.222 | 0.806 | 0.417 | 0.422 | 0.422 | 0.178  | 8   |
| HG         | 30   | 3.533 | 1.702 | 0.634 | 0.401 | 0.354 | 0.361 | -0.088 | 7   |
| LJ         | 30   | 3.667 | 2.257 | 0.860 | 0.488 | 0.502 | 0.018 | 4      |
| JH         | 30   | 3.067 | 1.586 | 0.497 | 0.293 | 0.280 | 0.127 | 1      |
| HLQY       | 30   | 2.733 | 1.877 | 0.523 | 0.324 | 0.294 | 0.300 | -0.025 | 0   |
| BL         | 30   | 3.400 | 1.970 | 0.666 | 0.361 | 0.367 | 0.116 | 2      |
| ML         | 30   | 3.000 | 1.840 | 0.525 | 0.285 | 0.290 | 0.013 | 3      |
| MES        | 30   | 3.000 | 1.784 | 0.539 | 0.293 | 0.300 | 0.186 | 2      |
| FQ         | 30   | 3.000 | 1.711 | 0.531 | 0.286 | 0.292 | 0.033 | 1      |
| WQ         | 31   | 2.800 | 1.913 | 0.579 | 0.327 | 0.328 | 0.337 | 0.148  | 2   |
| Mean       |      | 3.271 | 1.870 | 0.634 | 0.370 | 0.352 | 0.361 | 0.020  | 3.813|

Genetic variation among *Pinus koraiensis* populations

To evaluate the genetic variation among the collected samples, AMOVA was performed, and Fst among natural populations, genetic clusters and geographical regions were calculated; and the results are shown in Table 4. The AMOVA results indicated that 67% of the total genetic variation existed within populations, indicating high genetic diversity within populations. AMOVA of the two genetic clusters identified by the STRUCTURE analysis indicated that 63.79% of the total variation was attributable to differences within populations, and the overall Fst was 0.362 (Fst > 0.25), indicating high genetic differentiation between the 2 clusters. In addition, the AMOVA of two groups classified according to geographical location indicated low genetic variation among populations within each group (2.77%). All of these results indicated high genetic differentiation within populations and groups.

Table 4
Results of analysis of molecular variance (AMOVA) for 16 populations of *Pinus koraiensis* in China

| Type                  | Source of variation | d.f. | Sum of squares | Variance component | Percentage of variation | Fixation index |
|-----------------------|--------------------|------|----------------|--------------------|------------------------|---------------|
| Variance partitiona   | Among populations  | 15   | 2191.019       | 4.559              | 33.00                  | FST = 0.516   |
|                       | Within populations | 464  | 4311.567       | 9.292              | 67.00                  |               |
|                       | Total              | 479  | 6502.585       | 13.851             |                        |               |
| Variance partitionb   | Among groups       | 1    | 115.322        | 0.266              | 26.67                  | FST = 0.362   |
|                       | Among populations within groups | 14   | 88.917       | 0.095              | 9.55                  | FSC = 0.130   |
|                       | Within populations | 942  | 600.429        | 0.637              | 63.79                  | FCT = 0.267   |
|                       | Total              | 957  | 804.668        | 0.999              |                        |               |
| Variance partitionc   | Among groups       | 1    | 23.753         | 0.024              | 2.77                   | FST = 0.264   |
|                       | Among populations within groups | 14   | 180.487       | 0.205              | 23.64                  | FSC = 0.243   |
|                       | Within populations | 942  | 600.429        | 0.637              | 73.59                  | FCT = 0.028   |
|                       | Total              | 957  | 804.668        | 0.866              |                        |               |

Note: a The analysis included all collected populations as one hierarchical group. b The analysis included two geographical groups (G1 and G2). c The analysis included two genetic clusters (Clusters 1 and Clusters 2).
The Nei's genetic distance and pairwise Fst values are shown in Table 5. The Fst-value was considered the main genetic parameter for evaluating genetic differentiation among populations. In this study, the pairwise Fst values ranged from 0.014 to 0.348, and most of the *P. koraiensis* population pairs exhibited high values (Fst > 0.15), indicating high levels of genetic diversity. The greatest level of differentiation was observed between populations Helong and Liangshui, and the lowest was observed between Jiaohe and Hulin. The highest genetic distance was observed between populations Helong and Liangshui (0.813), consistent with the pairwise Fst values and indicating pronounced differentiation between these two populations. The relative migration network among the sixteen *P. koraiensis* populations was constructed using relative migration rate with the divMigrate function in R software. Analysis of gene flow between populations suggested a biased geographic distribution, and gene flow was not uniform among all populations (Fig. 2). A high degree of gene flow was observed among three populations located close to each other (Muleng, Maoershan and Fangzheng), consistent with the principal coordinate analysis and dendrogram analysis. In addition, one genetically isolated population (Boli) displayed high levels of gene flow with the three nearby populations Muleng, Maoershan and Fangzheng. Moreover, a moderate level of gene flow was found among three admixed populations, and two genetically distinct populations (Zhanhe and Wangqing) exhibited distinct segregation from the other populations.

![Table 5](image)

**Population structure**

Analysis of population structure of 16 natural *P. koraiensis* populations was performed based on a Bayesian approach using STRUCTURE software. The number of clusters within the range of 1 to 10 was evaluated for 10 repetitions in each run. In the structure plot, the maximum delta K value appeared at K = 2, with a noticeable peak apparent at this value. Thus, this value was considered the optimal genetic cluster number for all of the EST-SSR markers (Fig. 3b, c). The 480 sampled individuals of *P. koraiensis* were divided into two genetic groups (Group 1 and Group 2) at K = 2. Group 1 comprised of 149 individuals from five populations (Heihe, Liangshui, Zhanhe, Tieli and Hegang), and Group 2 comprised of a higher number of individuals (331) from eleven populations (Liangzihe, Helong, Lushuihe, Linjiang, Jiaohe, Hulin, Boli, Muleng, Maoershan, Fangzheng and Wangqing). Group 1 comprised almost all of the *P. koraiensis* plant materials from Xiaoxinganling Mountains, whereas Group 2 comprised almost all of the individuals from Changbaishan Mountains, suggesting a relationship between genetic structure and geographical distribution of the populations.

To further analyze clustering patterns, principal component analysis (PCA) based on the pairwise genetic distance matrix of fifteen EST-SSRs was performed; and the results are shown in Fig. 5a. The 480 individuals from the sixteen populations were roughly divided into two clusters according to the first two axes in the PCA plot. Principal axes 1 and 2 accounted for 22.99% and 12.46%, respectively, of the total genetic variation among the individuals, together accounting for 35.45% of the total genetic variation (Fig. 4a). Five populations (Heihe, Liangshui, Zhanhe, Tieli and Hegang) were grouped into cluster 1, and the remaining populations were grouped into cluster 2. The same clustering was obtained in the STRUCTURE analysis using the same dataset, indicating marked genetic differentiation. Furthermore, the Neighbor-joining (NJ) dendrogram based on Nei's genetic distance clustered the 480 *P. koraiensis* individuals from the 16 populations into two clusters, consistent with the above results (Fig. 4b, Fig. 5).

**Correlations between genetic distance and geographic distance**
In this study, the geographic distance and genetic distance values ranged from 37.72 km to 825.45 km and from 0.02 to 0.83, respectively. We examined whether the genetic distance estimated based on molecular markers may be related to the distribution of the species under study and the geographic distance between individuals or populations. To investigate the correlations between genetic distance and geographic distance, the Mantel test was carried out. The results showed that genetic distance was not significantly correlated with the geographic distance among *P. koraiensis* populations (p = 0.26, R² = 0.01), indicating a lack of association between geographical distance and the genetic differentiation of *P. koraiensis* (Fig. 6). Liangzihe and Hegang exhibited the lowest geographic distance and were not grouped in the same cluster. Therefore, there was not apparent isolation by genetic and geographical distance among the sampled populations.

**Discussion**

To understand the genetic differentiation of forest tree populations and contribute to the development of effective breeding strategies, comprehensive evaluations of natural germplasm resources of individual species are essential. Such evaluations can accelerate breeding-strategy and industrial development [34–35]. Naturally, *P. koraiensis* mainly grows in the cold temperate zone, especially in northeast China, and natural forests of this species have been shown to be sensitive to climate factors. Thus, to conserve genetic resources of this species, it is important to obtained data on its genetic diversity and population structure. In present study, we conducted a population genetic analysis using codominant molecular markers, representing the first such analysis in *P. koraiensis*. The results can help guide the genetic improvement and resource conservation of this important conifer species.

**Genetic diversity**

Genetic diversity has been increasingly evaluated in species lacking a reference genome, including some conifers [36], endemic species [37] and endangered plants [38]. Studies of genetic diversity can provide insight into speciation and genetic variation within and among populations and can aid the development of conservation strategies. *P. koraiensis* is a valued woody plant that provides high-quality timber and pine nuts and possesses high economic, ecological and ornamental value. However, transcriptome data and molecular markers remain lacking for *P. koraiensis*. The available genetic data provide few markers suitable for the study of population genetics in this species. Evaluating the germplasm resources of this species represents the first step towards understanding the genetics of natural *P. koraiensis* populations.

In this study, we evaluated the genetic diversity of natural *P. koraiensis* resources in China using 15 polymorphic EST-SSR markers. High levels of genetic diversity in natural *P. koraiensis* populations were detected, with mean values of 10.33 and 0.521 for Na and He, respectively. High genetic diversity was observed in the Heihe population in the northern Xiaoxinganling Mountains, possibly due to less human disturbance in this region than in other areas. According to a previous study, a PIC value equal to or more than 0.5 indicates high genetic information for genetic markers. In the present study, the PIC values obtained for the multiallelic EST-SSR markers ranged from 0.142 to 0.833, with a mean value of 0.461, indicating high level of genetic information among the 480 *P. koraiensis* individuals from the 16 natural populations. The genetic diversity of *P. koraiensis* obtained in the present study is higher than that reported in *Pinus bungeana* (Na = 3.70, He = 0.36) [39], *Pinus dabeshanensis* (He = 0.36) [40] and *Pinus yunnanensis* (Na = 4.10, He = 0.43) [41] but lower than that reported in *Pinus tabuliformis* (Na = 6.52, He = 0.68) [42].

The genetic diversity of a species may vary with characteristics such as adaptability, pollination mechanism and population size [43–45]. The observed genetic diversity in the present study might be attributed to the genetic background, life history, and population dynamics of *P. koraiensis*. Previous reports have found that *P. koraiensis* has a large population size, long life cycle, strong adaptability, long pollination distance and large genome size, and it has a complex genetic background, which allows it to generate high genetic diversity [46–48]. Over a long period of natural selection and succession, the natural distribution range and population number of *P. koraiensis* have greatly decreased, while those of man-made forests have increased in recent years. Furthermore, under the increasing influences of human activity and climate change, seed yield, plant resistance to stresses and other traits may be reduced and may hinder the maintenance of high levels of genetic diversity. Natural selection under the changing environment conditions is likely to lead to differences in genotype frequency among populations. In addition, previous studies have suggested that evaluations of genetic diversity are limited by low number of populations and molecular markers [49–50].

Studies on genetic diversity of natural *P. koraiensis* populations have been conducted using a variety of molecular-marker techniques. Kim et al. [51] analyzed allozyme loci variation and found a moderate level of genetic diversity among natural *P. koraiensis* populations in Korea. The genetic diversity of natural *P. koraiensis* populations detected by allozyme markers (He = 0.18) in Russia [52] was much lower than that identified using EST-SSRs markers in this study. These findings indicate that *P. koraiensis* maintains high genetic diversity worldwide. The level of genetic diversity detected in this study is similar to that detected based on nine EST-SSR markers in seven natural populations of *P. koraiensis* in northeast China (He = 0.610) [53]. Xiaoxinganling Mountain of China was considered the distribution center for *P. koraiensis*, possessing abundant germplasm resources and ancient founding stocks and maintaining considerable numbers of individuals. In addition, the genetic diversity of *P. koraiensis* populations from Xiaoxinganling Mountains was higher than that of the Changbaishan Mountains populations, with high expected heterozygosity and abundant private alleles found for the former populations (Fig. 5). All of these results indicate that Xiaoxinganling Mountains may be the center of genetic diversity of *P. koraiensis* in China.

**Population genetic differentiation**

Detection of genetic differentiation is a key process in the genetic improvement of forest trees. Regarding the estimation of genetic differentiation, previous studies have considered an Fst value higher than 0.15 but lower than 0.25 to indicate significant divergence [54–56]. In the present study, the genetic differentiation assessed by EST-SSRs among *P. koraiensis* populations ranged from 0.014 to 0.348, with a mean value of 0.177, indicating significant differentiation among populations in China. However, previous studies reported low genetic differentiation among populations assessed by
alozyne loci variation in Korea Far East (Fst = 0.06) [51] and Russia (Fst = 0.015) [52] and by EST-SSRs in China (Fst = 0.02) [53]. In addition, Kim et al. [29] studied the genetic variation of *P. koraiensis* in Korea, Russia and China using allozymes and RAPDs and detected small differences among the three regions. Different degrees of genetic differentiation were observed in natural *P. koraiensis* populations in these countries, with low Fst values. The main reason for these differences is that only limited numbers of natural populations and molecular markers were analyzed.

The genetic differentiation index (Fst) is correlated with gene flow (Nm). Generally, the greater the degree of differentiation, the weaker the gene flow, i.e., a lower gene migration rate among populations [57–59]. Thus, the genetic differentiation is facilitating gene flow. Natural *P. koraiensis* forest originated in Siberia in Northeast Asia and has undergone regeneration, succession and migration over millions of years [60–61]. After the Quaternary glaciation, many species died out, but the *P. koraiensis* forests persisted into the present and underwent a range of changes and varying degrees of differentiation. In natural *P. koraiensis* populations, low levels of genetic differentiation have been observed in Korea [29], whereas high genetic differentiation has occurred in Northeast China, which may have contributed to the rich *P. koraiensis* germplasm resources (representing approximately 60% of the world’s total) and broad distribution area (more than 3000 hectares) in this country.

The mean He (0.521) across all loci was greater than Ho (0.374), indicating a high heterozygosity among the sampled populations of *P. koraiensis*. This high heterozygosity is attributable to the fact that Pinus species exhibit cross-pollination and wind pollination. Furthermore, the AMOVA suggested that most of the genetic variation (more than 60%) in *P. koraiensis* exists within populations, with a small proportion occurring among populations; suggesting that genetic differentiation between populations exists.

**Population structure and gene flow**

Analyses of population structure can provide insight into population size, breeding system, extent of isolation and population migration or gene flow [62–63]. Furthermore, such analyses can help reveal the relationships between genetic variation and environmental stresses and enhance our understanding of evolution. Evaluating population structure is a key component of genome-wide association analysis (GWAS) and marker-assisted selection (MAS) [64]. *P. koraiensis* is mainly distributed in Xiaoxinganling Mountains and Changbaishan Mountains in northeast China, areas with a humid climate. Due to the environmental conditions, the germplasm resources of *P. koraiensis* from different locations display high phenotypic and genetic variation. The STRUCTURE analysis of population structure identified two groups (optimal K = 2) from the sixteen natural populations, with five populations in one group and the remainder in another group. Similar results were obtained in the PCA and dendrogram (neighbor-joining tree) analysis, indicating genetic differentiation of *P. koraiensis* in China.

Interestingly, individuals from Xiaoxinganling Mountains were clustered into one group, occupying a northern area, which makes them more like an ancestral group. Furthermore, the samples from Changbaishan Mountains and adjacent ridge region were clustered into the other group; the populations corresponding to these samples are distributed in a southern area and exhibit different degrees of genetic differentiation and gene flow. However, some of the individuals from Xiaoxinganling Mountains were clustered into cluster 2, although the majority were clustered into cluster 1 (Fig. 5a, Fig. 5b). The main potential reasons for this finding are as follows: (1) these two mountain regions are close to each other, and some hybridization events may occur; (2) for populations separated by a short spatial distance, the probability of gene flow is high, which will affect population genetic structure; and (3) pollen and seed dispersal occurs over long distances in this species, which promote gene flow. The genetic structure of the natural *P. koraiensis* populations in China determined in this study is consistent with the current geographical distribution of these populations. Furthermore, the findings are consistent with previous studies showing that populations in similar geographical locations or environments tend to cluster into the same group [65–66].

Gene flow among populations is closely related to geographical distance and effective population size and can generate new genetic combinations, potentially enhancing species resilience and persistence [67–69]. In plants, migration or gene flow is achieved via seeds, pollen and other propagules, and influences the genetic diversity and differentiation among independent evolutionary units [70–71]. We found that two genetically distinct populations (Zhanhe and Wangqing) exhibited segregation from other populations, which may be related to their geographic distance from those other populations (approximately 565 km), limiting the level of gene flow between them. These independent units play an important role in maintaining the genetic diversity of this species. This interpretation is consistent with previous studies demonstrating that isolated populations of plants with long-distance pollination may have higher levels of genetic diversity than large contiguous populations [72–74].

Moreover, high levels of gene flow were found among Helong, Maorheshan and Fangzheng populations. High levels of gene flow can reduce the effects of artificial selection or genetic drift and promote the maintenance of genetic information. Similar results were obtained for *Oryza sativa* accessions [61]. Extensive gene flow can alter the gene frequencies in populations to affect genetic diversity and structure. In our study, although a strong correlation between gene flow and geographic distance between populations was observed, some degree gene flow was also evident between geographically distant populations. In addition, geographic distance was not correlated with genetic distance in the natural *P. koraiensis* populations in this study, suggesting that geographic distribution may not be a determinant factor for the genetic structure of populations.

**Suggestions for conservation**

Evaluations of germplasm resources are needed to maintain abundant genetic variation and high levels of genetic diversity of some species of interest and establish sound conservation strategies. Our population genetic analysis revealed that the populations distributed in the Xiaoxinganling Mountains area (Zhanhe, Heihe, Liangshui, Tieli and Hegang) exhibit high levels of genetic diversity and moderate levels of gene flow (Fig. 4). These populations represent the core populations and have stronger environmental adaptability and evolutionary potential than the other populations, and they can be considered independent genetic units. Hence, measures such as in situ conservation should be implemented for conserving natural *P. koraiensis* resources. In addition, the marginal populations represent special germplasm resources; and they are characterized by low genetic diversity but have high levels of
genetic differentiation relative to the other natural populations. Habitat fragmentation can reduce gene flow among populations, leading to a loss of genetic diversity. In this study, the Helong population, which occurs in a marginal area, should be targeted for conservation measures, such as ex situ measures. In addition, the greatest level of differentiation observed between populations was between Helong and Liangshui, indicating that these populations can be considered independent units. Therefore, regulations and management strategies must be established to protect the natural habitat of this species and prohibit harvest. More importantly, a national-level core germplasm resources library of *P. koraiensis* should be established by the government, with the objectives of maintaining genetic variation, improving plant adaptability to environmental changes, and developing new breeding materials. Under these measures, the existing natural *P. koraiensis* populations in China can be protected and be better used as a source of material for genetic improvement.

### Conclusion

This study investigated the genetic diversity and population structure of natural *P. koraiensis* populations in northeast China, which is the first comprehensive report of the genetic diversity of natural *P. koraiensis* populations in China. We found that the existing *P. koraiensis* populations in China maintain high levels of genetic diversity, which provide a foundation for germplasm innovation and genetic improvement of *P. koraiensis*. The population genetic analysis in this study identified two independent genetic units (Liangshui and Helong) that exhibit high degrees of genetic differentiation. The populations distributed in the Xiaoxinganling Mountains area are highly genetically diverse and may represent the central population of natural *P. koraiensis* in China. Our findings provide genetic information useful for future genome-wide association studies (GWAS) and marker-assisted selection (MAS) and genomic selection (GS) studies. Furthermore, the genetic structure of *P. koraiensis* populations identified in this study is consistent with the geographical distribution of these populations in China. These results have significant implication for the protection of natural *P. koraiensis* germplasm resources in China. Thus, we suggest that appropriate in situ and ex situ conservation measures should be taken to preserve the germplasm resources.

### Methods

#### Collection of plant materials and genomic DNA extraction

The plant materials used in this study were obtained from the wild and permission was obtained to collect samples. The collection of plant materials also complied with institutional, national, or international guidelines. The formal identification of the samples used in this study was performed by Xi-Yang Zhao. Voucher specimens were deposited in the herbarium of Northeast Forestry University. We investigated 16 populations of *P. koraiensis* from Jilin Province (J) and Heilongjiang Province (H) in the current study. A total of 480 samples were collected throughout the natural distribution areas in northeastern China (Table 1, Fig. 1). To obtain representative samples of populations, sampling was conducted throughout the distribution range, with at least 200 m between sampled individuals. The number of sampled individuals per population was on average number 30. The average altitude of the sampled individuals was 320 m. The sampled populations were selected to represent the main distribution region and included the Xiaoxinganling Mountains group (G1, including populations P4, P5, P6, P7 and P8) and the Changbai Mountains group (G2, including P1, P2, P3, P9, P10, P11, P12, P13, P14, P15 and P16). The populations were distributed across the northernmost (Heihe, 49°24'47", 126°36'47.16"), southernmost (Linjiang, 42°0'36", 127°13'12"), and easternmost regions (Hulin, 45°46'12", 132°58'48") of natural *P. koraiensis* populations in Northeast China.

The fresh needle samples of *P. koraiensis* with no signs of pests or disease were immediately frozen in liquid nitrogen and stored at -80 °C for subsequent genomic DNA extraction and PCR amplification. In addition, nucleic acids were extracted from needles using the improved cetyltrimethyl Ammonium Bromide (CTAB) method described by Li et al. [75]. WDNA quality and concentration were evaluated using 1.0% agarose gel electrophoresis and the K5500 Plus microspectrophotometer (KAIAO Technology Development Co., Ltd., Beijing, China), respectively.

#### PCR amplification and SSR analysis

To detect polymorphisms in the 16 sampled *P. koraiensis* populations, 15 highly polymorphic and reproducible EST-SSR markers of *P. koraiensis* developed in our laboratory were selected in this study. The primers of *P. koraiensis* was developed as described by Li et al. [22]. Eight capillary electrophoresis templates were amplified with fifteen primers synthesized by Sangon Biotech (Shanghai, China), and universal M13 sequence (5’-TGTAAAACGACGGCCAGT3’) labeled with four fluorescent dyes (TAMRA, FAM, HEX and ROX) was added at the 5’ end of the forward primers. DNA was diluted to a working concentration of 25 ng/µL. To detect SSR loci, Polymerase Chain Reaction (PCR) was performed in a total volume of 20 µL containing 10 µL 2x Super PCR Mix (Beijing Genomics Institute Tech Solutions (Beijing Liuhe) Co., Ltd., Beijing, China), 2 µL template DNA, 0.8 µL forward primer (1 µM), 3.2 µL reverse primer (1 µM), 1 µL M13 primer with fluorescent label and 3 µL ddH2O. The PCR amplification conditions were as follows: 94 °C for 5 min; followed by 30 cycles at 94 °C for 30 s, 57 °C for 30 s, and 72 °C for 30 s; followed by 8 cycles at 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s; followed by final extension at 72 °C for 10 min. The PCR products were subjected to 1.0% agarose gel electrophoresis and then analyzed by high performance capillary electrophoresis (HPCE) using an ABI 3730XL DNA Sequencer (Applied Biosystems, Foster City, CA, USA) to detect fragment size. The original sequence data were analyzed using GeneMapper (version 4.1) software.

#### Data analysis

GeneMapper was used to obtain the microsatellite allele data, and the Microsatellite toolkit v 3.1.14 was used to convert the data into the necessary format for analysis. The genetic diversity analysis was conducted using GENALEX software version 6.50 [76] with the following parameters: number of alleles (Na), effective number of alleles (Ne), observed (Ho) and expected (He) heterozygosity, number of rare alleles (NRA), Shannon diversity index (I), Hardy–Weinberg equilibrium (HWE), F-statistics (Fis, Fit and Fst) and Nei's genetic distance. The TBtools software [77] was used to plot the map of heatmap of expected heterozygosity (He). In addition, we calculated the polymorphism information content (PIC) values of each SSR primer using the
PICcalc program [78]. Gene flow (Nm) was calculated as $Nm = (1-Fst)/4 \times Fst$ and used to measure the degree of gene exchange among or within the 16 populations. ALREQUIN software (version 3.5) [79] was used to analyze the level and sources of molecular genetic variation via AMOVA based on the evolutionary distances among and within the sampled populations and the observed genetic clusters. The total genetic variation was divided into three components: among groups, among populations within groups and within populations.

To evaluate the population genetic structure of *P. koraiensis*, a Bayesian clustering algorithm was performed in STRUCTURE software (version 2.3) [80] with the following settings: K-values from 1 to 10, with ten runs per K value and a burn-in period and number of Markov chain Monte Carlo (MCMC) reps after burn-in of 100,000 iterations and 100,000, respectively. The optimal K value for the number of populations was based on the delta-K values calculated by the Evanno method [81], using an algorithm of the online tool of STRUCTURE HARVESTER [82]. A clear peak was observed in the plot of delta K. In addition, principal component analysis (PCA) was performed to evaluate the genetic relationships among different populations using GENALEX software version 6.50. Based on the Nei's genetic distance (1983), a Neighbor-joining (NJ) phylogenetic tree of the populations was constructed using PowerMarker software (version 3.25) [83] and annotated and visualized using the online tool interactive Tree Of Life (iTOl) [84]. Geographic distance among populations was calculated as described in Li et al.'s study [75]. Finally, to detect the gene flow among the sixteen populations, a relative migration network was constructed using the ‘diveRsity’ [85] package of R software (version 3.5.0) [86].

**List Of Abbreviations**

Number of alleles: Na

Effective number of alleles: Ne

Number of rare alleles: NRA

Observed heterozygosity: Ho

Expected heterozygosity: He

Shannon diversity index: I

Hardy–Weinberg equilibrium: HWE

Gene flow: Nm

Expressed sequence tag-simple sequence repeat: EST-SSR

**Declarations**

**Ethics approval and consent to participate:** not applicable

**Consent for publication:** Not applicable

**Availability of data and material:** The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Competing interests:** there is no competing interests

**Funding:** This research study was supported by the Fundamental Research Funds for the Central Universities (Northeast Forestry University) (No. 2572017DA02). The funding agency has no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

**Authors’ contributions:** XZ and XL conceived the study, developed the experimental design and provided the suggestions and comments for this manuscript. XL, MZ and YX carried out fieldwork. XL and YL performed the data analysis. XL wrote the original draft and all authors contributed in preparing the manuscript in its final form. MT revised the original manuscript. All authors have read and approved the submission of the manuscript.

**Acknowledgements:** We thank members of the State Key Laboratory of Tree Genetics and Breeding for their assistance during laboratory works and for fruitful discussions.

**References**

1. Choi DS, Jin HO, Lee CH, Kim YC, Kayama, M. Effect of soil acidification on the growth of Korean pine (*Pinus koraiensis*) seedlings in a granite-derived forest soil. Environ. Sci. 2005;12:33-47.

2. Jia Y, Zhu, J, Wu Y, Fan WB, Zhao GF, Li ZH. Effects of geological and environmental events on the diversity and genetic divergence of four closely related Pines: *Pinus koraiensis*, *armandii*, *P. griffithii*, and *P. pumila*. Front. Plant. Sci. 2018;9:1264.
3. Wang F, Chen S, Liang DY, Qu GZ, Chen S, Zhao XY. Transcriptomic analyses of Pinus koraiensis under different cold stresses. BMC Genomics. 2020;21:10.

4. Ma JL. The evolution of Korea Pine forest. J. Northeast For. Univ. 1997;67-71.

5. Wang H, Shao XM, Jiang Y, Fang XQ, Wu SH. The impacts of climate change on the radial growth of Pinus koraiensis along elevations of Changbai Mountain in northeastern China. For. Ecol. Manage. 2013;289:333-340.

6. Zhu JJ, Wang k, Sun YR, Yan QL. Response of Pinus koraiensis seedling growth to different light conditions based on the assessment of photosynthesis in current and one-year-old needles. J. Forestry Res. 2014;25:53-62.

7. Fan Y, Moser W, Cheng YX. Growth and Needle Properties of Young Pinus koraiensis et Zucc. Trees across an Elevational Gradient. Forests. 2019;10:54.

8. Li X, Liu XT, Wei JT, Li Y, Tigabu M, Zhao XY. Genetic improvement of Pinus koraiensis in China: current situation and future prospects. Forests. 2020;11:148.

9. Zhang J, Zhou Y, Zhou G, Xiao C. Composition and structure of Pinus koraiensis mixed forest respond to spatial climatic changes. PloS One. 2014;9:e97192.

10. Wang XC, Zhang MH, Ying J, Li ZS, Zhang YD. Temperature signals in tree-ring width and divergent growth of Korean pine response to recent climate warming in northeast Asia. Trees. 2016;31:1-13.

11. Seo YW, Balekoglu S, Choi Jk. Growth pattern analysis by stem analysis of Korean white pine (Pinus koraiensis) in the central northern region of Korea. For. Sci. Technol. 2014;10:220-226.

12. Zhang ST, Zhang LG, Wang L, Zhao YH. Total phenols, flavonoids, and procyanidins levels and total antioxidant activity of different Korean pine (Pinus koraiensis) varieties. Forestry Res. 2018;30:1743-1754.

13. Liu M, Mao ZJ, Li Y, Sun T, Li XH, Huang W, Liu RP, Li YH. Response of radial growth of Pinus koraiensis in broad-leaved Korean pine forests with different latitudes to climatical factors. J. Appl. Ecol. 2016;27:1341-1352.

14. Kang KS, Harju AM, Lindgren D, Nikkanen T, Almqvist C, Suh GU. Variation in effective number of clones in seed orchards. New. Forest. 2001;21:17-33.

15. Han SU, Kang KS, Kim CS, Kim TS, Song JH. Effect of top-pruning in a clonal seed orchard of Pinus koraiensis. Ann. For. Res. 2008;51:155-156.

16. Park JM, Kwon SH, Lee HJ, Na SJ, El-Kassaby YA, Kang KS. Integrating fecundity variation and genetic relatedness in estimating the gene diversity of seed crops: Pinus koraiensis seed orchard as an example. Can. J. Forest. Res. 2016;47:366-370.

17. Guan, GY. Effects of different scion types on grafting survival rate and growth potential of Pinus koraiensis. Protection Forest Science and Technology. 2019;44-45.

18. Xia D, Yang SW, Yang CP, Lv QY, Liu GF, Zhang PG. Study of Pinus koraiensis provenance test (1):Provenance regionalization. J. Northeast. Univ. 1991;122-128.

19. Wang YX, Dong YH, Yang H, Wang HZ, Xu JM, Shan MG, Zhang GL. Open pollination progeny test of clonal seed orchard of Pinus koraiensis. Forestry Sci. Technol. 2005;34:20-21.

20. Liang DY, Ding CJ, Zhao GH, Leng WW, Zhang M, Zhao XY. Variation and selection analysis of Pinus koraiensis clones in northeast China. J. Forestry Res. 2018;29:611-622.

21. Chen MM, Feng FJ, Sui X, Li MH, Zhao D, Han SJ. Construction of a framework map for Pinus koraiensis et Zucc. using SRAP SSR and ISSR markers. Trees. 2010;24:685-693.

22. Li X, Liu XT, Wei JT, Li Y, Tigabu M, Zhao XY. Development and transferability of EST-SSR markers for Pinus koraiensis from cold-stressed transcriptome through Illumina sequencing. Genes. 2020;11:500.

23. Li JQ, Wang LH, Zhan QW, Liu YL, Yang XC. Transcriptome characterization and functional marker development in Sorghum sudanense. PloS One. 2016;11:e0154947.

24. Zhu HY, Song PY, Koo DH, Guo LO, Li YM, Sun SR, Wen YQ. Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis. Bmc Genomics. 2016;17:557.

25. Zhao XB, Li CJ, Wan SB, Zhang TT, Shan SH. Transcriptomic analysis and discovery of genes in the response of Arabis hypogaea to drought stress. Mol. Biol. Rep. 2018;45:119-131.

26. Liu YL, Geng YP, Song ML, Zhang PF, Wang WJ. Genetic structure and diversity of glycyrrhiza populations based on transcriptome SSR markers. Plant. Mol. Biol. Rep. 2019;37:401-412.

27. Gouy M, Rousselle Y, Roussel Y, Anglade A, Royaert S, Nibouche S, Costet L. Genome wide association mapping of agro-morphological and disease resistance traits in sugarcane. Euphytica. 2015;202:269-284.

28. Huang Z, Zhao N, Qin MF, Xu AX. Mapping of quantitative trait loci related to cold resistance in Brassica napus L. J. Plant Physiol. 2018;231:147-154.

29. Kim ZS, Hwang JW, Lee SW, Yang C, Gorovoy PG. Genetic variation of Korean Pine (Pinus koraiensis et Zucc.) at allozyme and RAPD markers in Korea, China and Russia. Silvae. Genet. 2005;54:235-246.

30. Acheampong A, Leveque N, Tchapla A, Heron S. Application of SRAP in the genetic diversity of Pinus koraiensis of different provenances. Afr. J. Biotechnol. 2010;8:1000-1008.

31. Chen JY. Analysis of Genetic Diversity of Pinus koraiensis Plantation in CaoHekou Forest Farm by ISSR Marker. Bulletin of Botanical Research. 2009;29:633-636.
32. Feng FJ, Han SJ, Wang HM. Genetic diversity and genetic differentiation of natural Pinus koraiensis J. For. Res. 2006;17:21-24.
33. Tong YW, Lewis BJ, Zhou WM, Mao CR, Wang Y, Zhou L, Yu DP, Dai LM, Qi L. Genetic diversity and genetic differentiation of natural Pinus koraiensis population. Forests. 2020;11:39.
34. Halbauer EM, Bohinec V, Wittenberger M, Hansel-Hohl K, Sehr EM. Genetic diversity of flax accessions originating in the Alpine region: a case study for the selection of genetically superior genotypes based on molecular marker. Euphytica. 2017;213:120.
35. Krishnan AG, John R, Cyriac A, Sible GV. Estimation of genetic diversity in nutmeg (Myristica fragrans Houtt) selections using RAPD markers. Int. J. Plant. Sci. 2012;172:102-107.
36. Thomas L. Genetic variation in Turkish Red Pine (Pinus brutia) seed stands as determined by RAPD markers. Silvae Genet. 2017;53:169-175.
37. Chung MY, Moon MQ, López-Pujol J, Chung JM, Chung MG. Genetic diversity in the two endangered endemic species Kirengeshoma koreana (Hydrangeaceae) and Parasenecio pseudotaimingasa (Asteraceae) from Korea: Insights into population history and implications for conservation. Syst. Ecol. 2013;51:60-69.
38. Sanaa A, Bouilla A, Boussaid M, Fadhel NB. Genetic diversity, population structure and linkage disequilibrium analysis in the endangered Tunisian Pancratium maritimum Linnaeus (Amaryllidaceae) populations. Afr. J. Ecol. 2016;54:379-382.
39. Duan D, Jia Y, Yang J, Li ZH. Comparative transcriptome analysis of male and female conelets and development of microsatellite markers in Pinus bungeana, an Endemic Conifer in China. Genes. 2017;8:393.
40. Zhang ZY, Wang H, Chen W, Pang XM, Li YY. Genetic diversity and structure of native and non-native populations of the endangered plant Pinus dabeshanensis. Genet. Mol. Res. 2016;15:gmr.15027937.
41. Xu YL, Cai NH, Woeste K, Kang XY, He CZ, Li GQ, Chen S, Duan AA. Genetic diversity and population structure of Pinus yunnanensis by simple sequence repeat markers. Sci. 2016;62:38-47.
42. Wang MB, Hao ZZ. Rangefreeze genetic diversity in natural populations of Chinese pine (Pinus tabulaeformis). Biochem. Genet. 2010;48:590-602.
43. Ferrer MM, Eguiarte LE, Montana C. Genetic structure and outcrossing rates in Flourensia cernua (Asteraceae) growing at different densities in the South-western Chihuahuan Desert. Ann. Bot. 2004;94:419-426.
44. Hellmann JJ, Pineda-Krch M. Constraints and recombination mechanisms of population designation under climate change: selection of genetically correlated traits. Biol. Conserv. 2007;137:599-609.
45. Wang Z, Kang M, Liu HB, Gao J, Zhang ZD, Li YY, Wu RL, Pang XM. High-level genetic diversity and complex population structure of Siberian apricot (Prunus sibirica) in China as revealed by nuclear SSR markers. PloS One. 2014;9:e87381.
46. Xu YG, Yu JH, Wang AM. Study on pollination characteristics of natural population of Pinus koraiensis. Acta Ecol. Sin. 2000;20:430-433.
47. Zhang Z, Zhang HG, Mo C, Zhang L. Transcriptome sequencing analysis and development of EST-SSR markers for Pinus koraiensis. Sci. Silvae Sin. 2015;51:114-120.
48. Omelko A, Ukhvatkina O, Zhmerenetsky A. Disturbance history and natural regeneration of an old-growth Korean pine-broadleaved forest in the Sikhote-Alin mountain range, Southeastern Russia. For. Ecol. Manage. 2016;360:221-234.
49. Parsons BJ, Newbury HJ, Jackson MT, Ford-Lloyd BV. Contrasting genetic diversity relationships are revealed in rice (Oryza sativa) using different marker types. Mol. Breed. 1997;3:115-125.
50. Xiong G, Wang JQ, Liu HB, Gao J, Zhang ZD, Li YY, Wu RL, Pang XM. High-level genetic diversity and complex population structure of Siberian apricot (Prunus sibirica) in China as revealed by nuclear SSR markers. PloS One. 2014;9:e87381.
51. Kim ZS, Lee SW, Lim JH, Hwang JW, Kwon W. Genetic diversity and structure of natural populations of Pinus koraiensis (Sieb. et Zucc.) in Korea. For. Genet. 1994;1:41-49.
52. Potenko VV, Velikov AV. Allozyme Variation and mating system of coastal populations of Pinus koraiensis et zucc. in Russia. Silvae Genet. 2001;50:117-122.
53. Tong YW, Lewis BJ, Zhou WM, Mao CR, Wang Y, Zhou L, Yu DP, Dai LM, Qi L. Genetic Diversity and Population Structure of Natural Pinus koraiensis Forests. 2020;11:39.
54. Wright S. The interpretation of population structure by F-Statistics with special regard to systems of mating. Evolution. 1965;19:395-420.
55. Holsinger KE; Weir BS. Genetics in genetically structured populations: defining, estimating and interpreting Fst. Nat. Rev. Genet. 2009;10:639-650.
56. Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P. Calculations of population differentiation based on GST and D: Forget GST but not all of statistics. Mol. Ecol. 2010;19:3845-3852.
57. Waples RS. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. J. Hered. 1998;59:438.
58. Krajeveld-Smit FJL, Beebee TJC, Griffiths RA, Moore RD, Schley L. Low gene flow but high genetic diversity in the threatened Mallorcan midwife toad Alytes muletensis. Mol. Ecol. 2005;14:3307-15.
59. Schmidt AM, Jacklyn P, Korb J. Isolated in an ocean of grass: low levels of gene flow between termite subpopulations. Mol. Ecol. 2013;22:2096-2105.
60. Zhao FQ, Yang J, He HS, Dai LM. Effects of natural and Human-assisted regeneration on landscape dynamics in a Korean Pine forest in Northeast China. PloS One. 2013;8:e82414.
61. Bao L, Kudureti A, Bai WM, Chen RZ, Wang TM, Wang HF, Ge JP. Contributions of multiple refugia during the last glacial period to current mainland populations of Korean pine (Pinus koraiensis). Sci. Rep. 2016;5:8608.
Schoettle AW, Goodrich BA, Hipkins VD, Richards C, Kray J. Geographic patterns of genetic variation and population structure in Pinus aristata, Rocky Mountain bristlecone pine. Can. J. For. Res. 2011;42:23-37.

Shamseldeen E, Sallam A, Belamkar V, Emara HA, Nower AA, Salem KF, Poland J, Baenziger PS. Genetic diversity and population structure of F_2.6 nebraska winter wheat genotypes using genotyping-by-sequencing. Front. Genet. 2018;9:76.

Luo ZN, Brock J, Dyer JM, Kutchan T, Augustin M, Ge YF, Fahlgren N, Abdel-Haleem H. Genetic diversity and population structure of a camelina sativa spring panel. Front. Plant Sci. 2019;10:184.

Wang HR, Vieira FG, Crawford JE, Chu CC, Nielsen R. Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. Genome Res. 2017;27:1029-1038.

Peng LP, Cai CF, Zhong Y, Xu XX, Xian HL, Cheng FY, Mao JF. Genetic analyses reveal independent domestication origins of the emerging oil crop Paeonia ostii, a tree peony with a long-term cultivation history. Sci. Rep. 2017;7:5340.

Luo ZN, Brock J, Dyer JM, Kutchan T, Augustin M, Ge YF, Fahlgren N, Abdel-Haleem H. Genetic diversity and population structure of a camelina sativa spring panel. Front. Plant Sci. 2019;10:184.

Peng LP, Cai CF, Zhong Y, Xu XX, Xian HL, Cheng FY, Mao JF. Genetic analyses reveal independent domestication origins of the emerging oil crop Paeonia ostii, a tree peony with a long-term cultivation history. Sci. Rep. 2017;7:5340.

Chung MY, Chung MG. Large effective population sizes and high levels of gene flow between subpopulations of Lilium cernuum (Liliaceae). Biochem. Syst. Ecol. 2014;54:354-361.

Bartkowska MP, Wong AYC, Sagar SP, Zeng L, Eckert CG. Lack of spatial structure for phenotypic and genetic variation despite high self-fertilization in Aquilegia canadensis (Ranunculaceae). Heredity. 2018;121:605-615.