

# SCIENTIFIC REPORTS



OPEN

## Contributions of multiple refugia during the last glacial period to current mainland populations of Korean pine (*Pinus koraiensis*)

Received: 17 August 2015  
Accepted: 09 November 2015  
Published: 22 December 2015

Lei Bao, Ayijiamali Kudureti, Weining Bai, Rongzhang Chen, Tianming Wang, Hongfang Wang & Jianping Ge

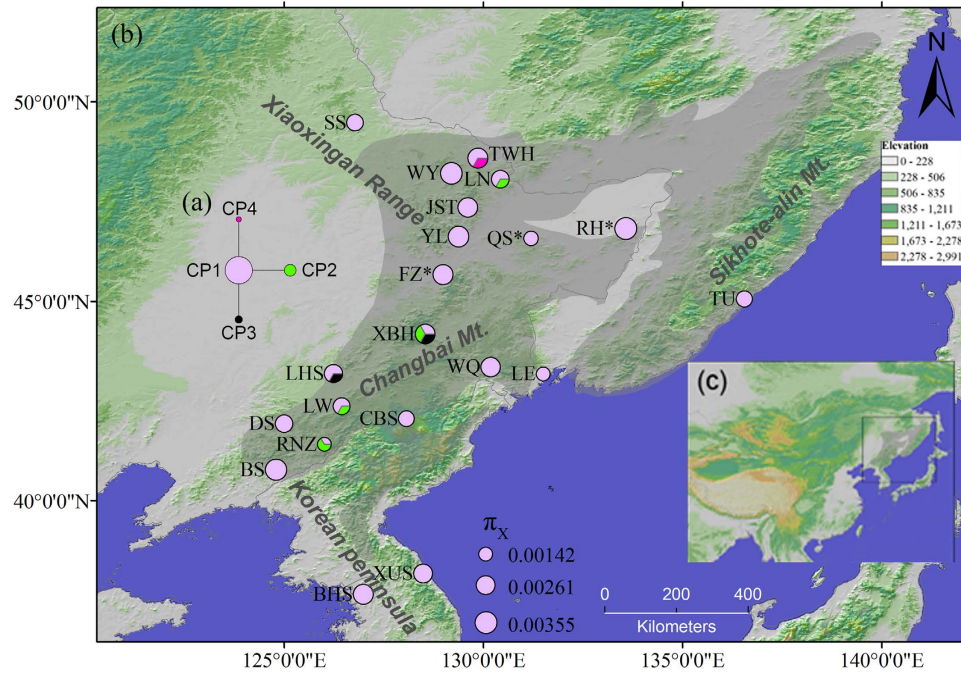
The northern microrefugia that existed during the Last Glacial Maximum (LGM) are a key factor in the demographic history of species. *Pinus koraiensis* has a unique distribution in northeast Asia. The Changbai Mountains and the Korean peninsula (CM/KP) are usually considered to be the LGM refugia for *P. koraiensis*. However, the Xiaoxingan Range (XR), at the northern part of this species' distribution, is another possible refugium. We used chloroplast sequencing and ten nuclear single-copy gene loci to calculate the genetic diversity pattern of *P. koraiensis*. The probabilities of a single LGM refugium and of multiple LGM refugia were calculated based on approximate Bayesian computation. The effect of the latitudinal gradient on genetic diversity was not significant. However, unique alleles occurred at low frequencies in CM/KP and XR. A conservative estimate of the coalescence time between CM/KP and XR is 0.4 million years ago, a time prior to the LGM. Gene flow between CM/KP and XR was estimated to be more than one in per generation, an amount that may be sufficient to limit genetic divergence between the regions. Our study strongly supports the hypothesis that XR was another LGM refugium in addition to CM/KP.

Historical climate change, especially during the last glacial maximum (LGM, ca. 18,000–24,000 years before the present), has affected the demographic history of species and shaped modern distributions. Enlarged ice caps in North America and Europe during the LGM caused significant local extinctions of temperate flora in their northern ranges<sup>1,2</sup>. In East Asia, extensive land ice caps did not develop<sup>3</sup>, but pollen record data show that reduced temperatures (mean reduction = 7–10 °C) and increased aridity caused extensive local extinctions (> 30 °N)<sup>4–6</sup>. For plants currently restricted to high latitudes, it is typically unknown which of their northern populations went extinct during the LGM. Did some populations survive in Northern microrefugia? The discovery and analysis of northern LGM microrefugia is important for better understanding evolution and range shifts under global climate change<sup>7,8</sup>.

Korean pine (*Pinus koraiensis* Siebold & Zuccarini) is distributed in Northeast China and adjacent regions including the Korean peninsula and the Russian Far East (hereafter, NEA) and scattered locations in the Japanese archipelago. Pollen records indicate that mixed forest in NEA was replaced by boreal forest, tundra or steppe during the LGM. Very low amounts of *Pinus* pollen have been found during this period<sup>4,9,10</sup>, which may be due to rare locations with LGM pollen records in this region. Korean pine populations may have declined or disappeared in most of the modern range of this species. Which location(s) were possible refugia for Korean pine population survival during the cold LGM period? Mitochondrial DNA markers indicate clear genetic differentiation between the Japanese archipelago and the NEA populations, suggesting that the Japanese archipelago was a refugium independent from the Asian continent during the LGM<sup>11</sup>. However, because very low genetic diversity has been found in mitochondrial and chloroplast DNA markers, the possible location(s) for NEA population survival during the LGM are unknown.

The Changbai Mountain Range, located in southern NEA, has the only mountains in this region with altitudes > 2000 m (maximum, 2749 m). The Changbai Mountains (CM) can act as a barrier blocking cold air masses from Siberia. Therefore, the southern CM and the Korean peninsula (KP) are areas that potentially supported relict

State Key Laboratory of Earth Surface Processes and Resource Ecology, Ministry of Education Key Laboratory for Biodiversity Science and Engineering & College of Life Sciences, Beijing Normal University, Beijing, China. Correspondence and requests for materials should be addressed to H.W. (email: wanghf@bnu.edu.cn)



**Figure 1.** (a) Median-joining network of chloroplast haplotypes, calculated in Network 4.6.1.2. (b) Chloroplast haplotype distribution over the 21 sampled populations. The colors in the pie chart correspond to the haplotype colors shown in (a). The circle size represents the pairwise nucleotide genetic differences within populations ( $\pi_x$ ). The three populations (FZ, QS and RH) marked with stars were at the northern border of the Changbai Mts. and were not included in the msABC analysis. (c) Map of East Asia. The box indicates our study area. Maps (b,c) were generated in ArcMap 9.3 (ESRI Inc).

populations of Korean pine during the LGM. Existing phylogeographic studies on plants distributed in NEA have found relatively higher genetic diversity in CM and/or KP, suggesting possible LGM refugia in CM and/or KP<sup>12</sup>. Examples of tree species represented by this pattern include *Juglans mandshurica*<sup>13</sup>, *Quercus mongolica*<sup>14</sup>, *Fraxinus mandshurica*<sup>15</sup>, and *Acer mono*<sup>16</sup>. It has been suggested that Korean pine had LGM refugia in CM and/or KP. Based on allozyme and RAPD markers, high genetic diversity was found in Korean pine populations in CM and KP<sup>17–19</sup>. Relatively higher concentrations of *Pinus* pollen in CM may also support the CM/KP refugia hypothesis<sup>9</sup>. Because no modern Korean pine populations occur outside of NEA, it is believed that CM/KP is a Korean pine refugium<sup>11,20</sup>. Continuing debate on LGM refugia of Korean pine has focused on the numbers of LGM refugia<sup>11,19,21</sup> and the search for additional refugia north of CM, such as in the Xiaoxingan Range (XR, Fig. 1).

Kremenetski *et al.*<sup>21</sup> suggested that the limiting factor for Korean pine survival during the LGM was not simply low temperatures but also aridity. According to this hypothesis, XR might have served as another LGM refugium because it was more humid. Thus far, neither genetic studies nor paleovegetation reconstruction has provided support for the multiple refugia hypothesis. Inconsistent genetic patterns of Korean pine populations have been reported. Significant genetic clines were found by Kim *et al.*<sup>17</sup> using allozyme markers, suggesting a single refugium in CM/KP. Another allozyme study found no genetic clines with populations in the Russian Far East (north of XR) with high genetic diversity equal to CM populations<sup>19</sup>. These studies were conducted > 10 years ago and were limited by available molecular markers and older phylogeographic statistics. In a more recent study, using mitochondrial and chloroplast DNA markers, Aizawa *et al.*<sup>11</sup> found very low genetic diversity in NEA and suggested CM/KP as the LGM refugia for the Asian continental populations. However, because some populations north of CM have distinctive chloroplast haplotypes, the possibility of another LGM refugium in the northern region cannot be excluded. The range of Korean pine in Asia is restricted to NEA. Hence, low genetic diversity in plasmids may limit the ability to recover the demographic history of this species in Asia. Abundant nuclear single-copy DNA (nsc DNA) markers are now available for the genus *Pinus*<sup>22–25</sup>, providing an opportunity to more thoroughly assess the possible location(s) of the refugia.

In this study, we utilized multiple nsc DNA markers and robust statistical phylogeographical approaches. Our goal was to assess the possibility of XR as another LGM refugium. The specific questions asked are as follows: (1) Do NEA populations have a genetic diversity gradient related to latitude? Does the XR population have unique alleles? (2) What were the coalescence times of the continental Asian Korean pine populations before and after the LGM? (3) What was the direction of migration between the CM and the XR populations? If only one LGM refugium existed, in CM/KP, the XR populations probably originated from postglaciation expansion from the south. Hence, we would expect a significant cline in genetic diversity from south to north, and no unique alleles would be expected in XR populations. Based on this hypothesis, the coalescence time of the continental populations would be after the LGM, and a northward migration would be expected. If LGM refugia existed in both CM/KP and XR, a cline in genetic diversity is less likely. Unique alleles might exist in both CM/KP and XR populations.

In a multiple refugia hypothesis, we would expect a long coalescence time for the continental populations, at least before the LGM. Both northward and southward migrations are expected if both CM/KP and XR supported relict populations during the LGM.

## Results

**Genetic diversity of each locus.** The total length of the chloroplast intergenic fragment *psbA-trnH* was 597 bp. This fragment generated four haplotypes in the continental populations, with three polymorphic sites. The grey haplotype, CP1, was most dominant, occurring in all studied populations. The less frequent CP2 haplotype was found in both CM and XR. CP3 was only found in CM (populations LHS and XBH), while CP4 was only found in XR (population TWH) (Fig. 1).

Among the 10 nsc loci we sequenced, the lengths varied from 336 bp to 738 bp, with the exon length varying from 0 bp to 416 bp and the intron length varying from 54 bp to 647 bp (Table 1). Each nsc locus contained 2–19 segregating sites and produced 3–12 haplotypes (Table 2). The average total nucleotide diversity ( $\pi_t$ ) was 0.00295, ranging from 0.00866 (*a3ip2*) to 0.00023 (*CL180Contig1\_03*). The nucleotide diversity of silent sites ( $\pi_s$ ) was 0.00537 (0–0.00866), and that of nonsynonymous sites ( $\pi_a$ ) was 0.00062 (0–0.00285). Detailed information on the genetic diversity of each locus is presented in Table 2. According to statistical results based on Tajima's *D*, Fu and Li's *D\** and *F\**, no nsc loci departed significantly from the neutral hypothesis (Table 2).

Median-joining networks of haplotypes found in every 10 nsc loci indicated that 8 out of 10 loci had two or more dominant haplotypes and a frequency greater than 10%. Loci *0\_12929\_02* and *CL180Contig1\_03* only had one dominant haplotype, with frequencies of 85.8% and 93.2%, respectively. Dominant haplotypes of all 10 nsc loci occurred in all regions, including KP, CM, XR and the Sikhote-Alin Mts. Unique alleles were found in all regions except the Sikhote-Alin Mts. CM regions had unique alleles at all loci except *PK169*, whereas XR had unique alleles at five loci (*IFG8612*, *erd3*, *gatabp1*, *nac*, *PK169*), and KP also had unique alleles at five loci (*IFG8612*, *nac*, *PK20*, *PK47*, *setb*) (Fig. 2).

**Latitudinal gradient of genetic diversity.** The genetic diversity of each population ( $\pi_x$ ) ranged from 0.0014 to 0.0036. Two CM populations, RNZ and TU, had the lowest genetic diversity (0.0014), while a northern marginal population of CM (RH) had the highest genetic diversity (0.0036). Another two populations, BS (in CM) and YL (in XR), had relatively high genetic diversity (0.0033 and 0.0031, respectively) (Table S1, Fig. 1). After adjusting for the effect of sample size, two CM populations (BS and DS) had the highest allelic richness (*AR*) and private allelic richness (*PAR*), while CBS (in CM) had the lowest *AR* and *PAR* (Table S1).

Pearson correlation analyses indicated no significant correlation between latitude and  $\pi_x$  ( $R = 0.127$ ,  $P = 0.583$ ), *AR* ( $R = 0.220$ ,  $P = 0.443$ ), or *PAR* ( $R = -0.475$ ,  $P = 0.073$ ). A Mantel test showed no significant pattern of isolation by distance ( $R = 0.110$ ,  $P = 0.160$ ).

**Historical demography analysis.** The genetic diversity pattern of continental Korean pine populations fit the MRM best, with the posterior probability equal to 1 (Fig. 3). The Bayes factor of the MRM and SRM was much larger than 10, and this result strongly supported the MRM scenario. This model selection result meant that the CM/KP populations and the XR populations had a long coalescence time before the LGM and suggested that both the CM/KP and XR populations existed before the LGM.

The long-term effective population size of the CM/KP region was nearly 50 times larger than that of the XR region (median  $N_s = 4.34 \times 10^4$  and  $N_N = 0.09 \times 10^4$ , Table 3, Fig. 4). The long-term migration between CM/KP and XR was large in both directions. The migration from CM/KP to XR per generation ( $N_e m$ ) was slightly larger than 1 individual (1.08, 95% HPD: 0.16–2.08), while the contrasting direction of migration was less than 1 individual (0.24, 95% HPD: 0.06–0.55). The coalescence time of the CM/KP and XR populations was  $3.96 \times 10^4$  generations ago (95% HPD:  $0.50 \times 10^4$ – $7.58 \times 10^4$ ). The ancestral population was slightly larger than the population of the CM/KP region (median  $N_A = 5.39 \times 10^4$ ).

## Discussion

Based on ten nsc DNA loci and one chloroplast intergenic fragment, we found low genetic diversity within mainland populations of Korean pine. This result was consistent with another report using mitochondria and chloroplasts by Aizawa *et al.*<sup>11</sup>. We found four chloroplast haplotypes, of which one was dominant and widespread in all populations. The distribution pattern and chloroplast diversity of Korean pine were similar to those of other sympatric trees, such as *Pinus densiflora* (four chlorotypes), *P. sylvestris* var. *mongolica* (six chlorotypes)<sup>25</sup>, *Larix sibirica* (one chlorotype)<sup>26</sup>, *Juglans manschurica* (one chlorotype)<sup>13</sup>, *Quercus mongolica* (four chlorotypes)<sup>14</sup>, and *Acer mono* (four chlorotypes)<sup>16</sup>. The nucleotide diversities of the ten nsc loci were 0.00295 ( $\pi_t$ ), 0.00062 ( $\pi_a$ ), and 0.00537 ( $\pi_s$ ), which were consistent with levels of genetic diversity reviewed by Leffler *et al.*<sup>27</sup> but lower than the genetic diversity in four sympatric *Pinus* species<sup>25</sup>. For example, among the four pine species studied by Ren *et al.*<sup>25</sup>, *P. sylvestris* var. *mongolica* had the lowest diversity based on seven nsc loci ( $\pi_t = 0.00579$ ,  $\pi_a = 0.00172$ ,  $\pi_s = 0.00759$ ), but this was greater than that of Korean pine. Populations of Korean pine from the CM/KP and XR regions had unique alleles in the chloroplast and nuclear genome, but the latitudinal gradient of population genetic diversity was not significant. ABC simulations strongly supported the theory that CM/KP and XR regions had an ancestral coalescence time long before the LGM. These findings suggest that the continental Korean pine probably had multiple refugia populations during the LGM.

We found no significant gradients related to latitude and no pattern of isolation by distance ( $P > 0.05$ ). Many populations located in CM/KP had high genetic variation and uniqueness (e.g., population RH, BS and YL, Fig. 1). Both a single refugium and a multiple-refugia scenario could produce such a pattern. Due to possible genetic drift along the leading edge, population expansion would generally be accompanied by a loss of genetic variation during the expansion period<sup>28</sup>. We would therefore expect postglacial expansion from a single refugium located in

Locus	PCR primers (5'-3')	Annealing temperature (°C)	Alignment length (bp)	Putative function	Exon	Intron	Sources
IFG8612	F: TGTTAGCATGCAATCAATCAC	59	434	Late embryogenesis abundant protein	333-434	1-332	[1]
	R: CTGACAGAGTGGCAGCTTCAT						
a3ip2	F: AATGCCAGGTTGGTGTTA	60	545	ABI3-interacting protein 2		1-545	[2]
	R: CAGCCTCAATTTGCTTTCC						
erd3	F: GAACGGGTCCGTACATTTCTG	65	657	Early responsive to dehydration 3	1-165; 262-338; 484-657	166-261; 339-483	[2]
	R: TGCCAGATTGATTGGCATAGAA						
nac1	F: CTTGGCTGTGGATGTATTGGA	63	738	Regulatory element	648-738	1-647	[3]
	R: ATCGTTTCAACTGCCTTGGCTT						
gatabp1	F: CGACCTTTGTGGAGGATTCTTTTATG	60	336	Regulatory element	7-288	1-6; 289-336	[3]
	R: CTTGGGCTTCTTGCTATGGGTTTT						
set-like-b	F: TTATTACATCCAACAGCGCATTT	63	489	Regulatory element	253-304; 482-489	1-252; 305-481	[3]
	R: GAAAGTATGGATTGCCAACTGCAC						
0_9389_01	F: CAACTCCCTCTCTTC	48	480	Nucleic acid binding protein	1-47, 264-328, 424-480	48-263, 329-423	[4]
	R: CATTGCCACAGTGGTTAC						
CL180Contig1_03	F: AAGTGAACAACGGCTG	57	579	Polygalacturonate 4-alpha-galacturonosyltransferase	1-31, 435-576	32-434	[4]
	R: TGGAGAAGGGAGAAGTG						
0_1688_02	F: GACTGACTATAACAGCC	54	628	Leucine-rich repeat family protein	1-154, 469-628	155-468	[4]
	R: CCCAAAATCACCAACAAC						
0_12929_02	F: GTTACAACATCAGCAATCAG	57	707	Protein kinase family protein	1-255, 606-707	256-605	[4]
	R: AAGCCAATCCACCAGTTATATACAG						

**Table 1.** Information about the ten nuclear single-copy loci used in this study. [1]<sup>24</sup>; [2]<sup>25</sup>; [3]<sup>23</sup>; [4]<sup>38</sup>.

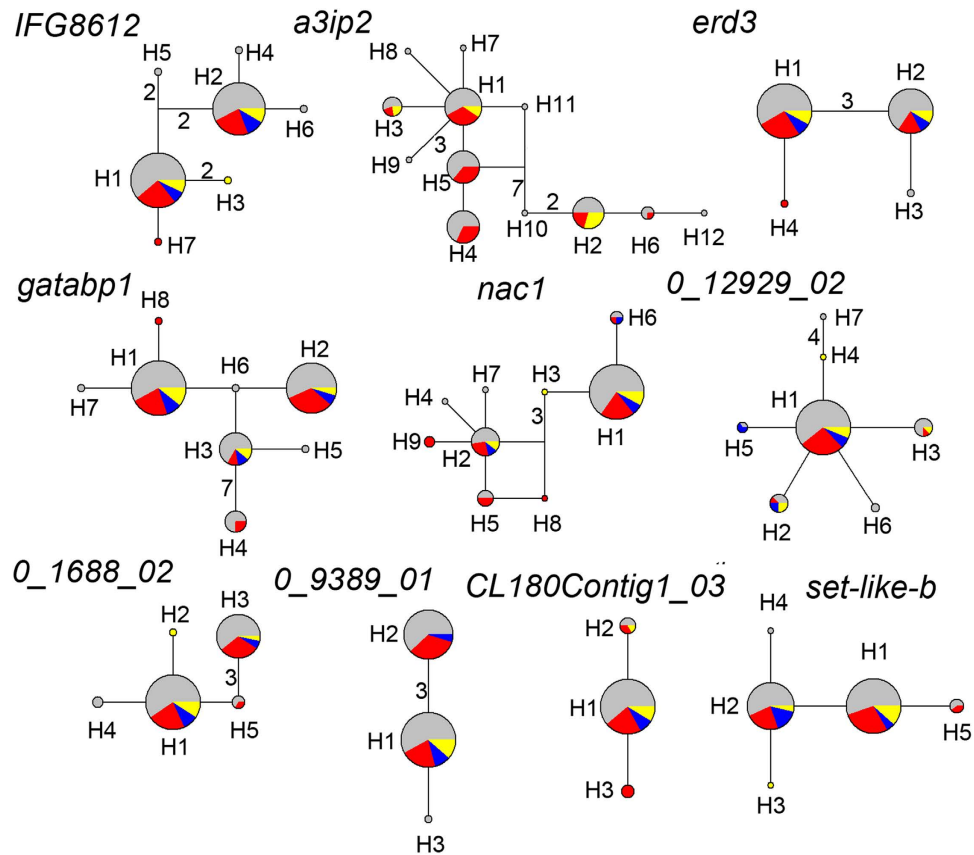
CM/KP to generate a latitudinal gradient of genetic diversity and isolation by distance pattern with the refugium population containing the highest genetic variation and uniqueness (such as for *F. manschurica*<sup>15</sup>, *Q. mongolica* (Zeng *et al.* unpublished data), and *J. mandshurica*<sup>13</sup> in the same region). If postglacial expansion evolved from multiple refugia, genetic variation loss during expansion would not be correlated with latitude and there would not be a clear pattern of isolation by distance. Single or multiple genetic diversity center(s) are both possible in the multiple-refugia scenario. However, a single diversity center in this multiple-refugia scenario would have originated from a mixture of migration events from multiple refugia, and this would reduce the uniqueness of genetic diversity<sup>29</sup>. Based on our results, the diversity pattern of Korean pine on the Asian continent supports the notion that there were refugia in addition to the existing CM/KP area. However, we should be cautious when inferring multiple refugia based on the absence of a latitude gradient and the diversity center locations, because in some situations, single refugium might generate a similar pattern. Pines are usually considered to be species with a high dispersal ability because pollen can, in some instances, travel more than 1000 kilometers<sup>30,31</sup>. Hence, for pine species, significant gene flow via pollen might prevent latitude gradients. For example, *Picea chihuahuana*, naturally distributed in Northwestern Mexico, shows two spatially separated mitotypes, which disperse via seeds. In contrast, no phylogeographic structure was found from chloroplast markers, which are dispersed via pollen in conifer species<sup>32</sup>. The distribution of Korean pine on the Asian continent extends less than 2000 kilometers between its southern and northern margins. Therefore, we cannot rule out the possibility of a single refugium if our inferences are based only on patterns of latitude gradient and isolation by distance.

Fortunately, multiple nuclear gene markers and the development of statistical phylogeography can provide more details of the demographic history of Korean pine. ABC simulations suggest that the coalescence time of CM/KP and XR populations was approximately  $3.96 \times 10^4$  generations ago (95% HPD:  $0.50 \times 10^4$ – $7.58 \times 10^4$ ) (Table 3, Fig. 4). The average reproductive age of Korean pine is approximately 200 years<sup>33,34</sup>, which can be considered a reasonable estimate of Korean pine's generation time. In this scenario, the coalescence time of the two regions was 7.92 million years ago (MYA), with the lower bound estimated as one MYA. Even using the initial reproductive age of Korean pine (80 years) as the generation time<sup>35</sup>, which usually underestimates the actual generation time, the coalescence time of CM/KP and XR populations was 3.17 MYA, with a lower bound estimated as 0.40 MYA, long before the last glaciation. Hence, it is certain that the two mountain regions sustained two separate populations for at least several glacial-interglacial cycles.



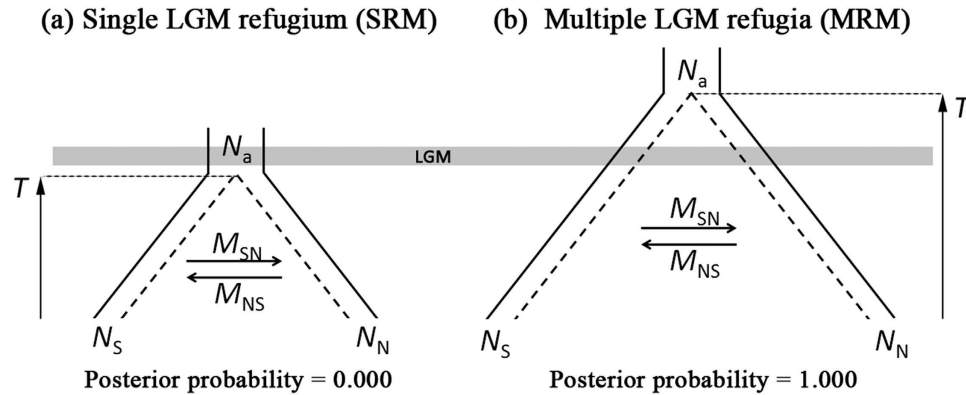
Locus	S	H	$\pi_t$	$\pi_s$	$\pi_a$	$\theta_w$	$\theta_{ws}$	Hd	Tajima's D	Fu and Li's D*	Fu and Li's F*
<i>IFG8612</i>	10	7	0.00329	0.00397	0.00018	0.00413	0.00453	0.478	-0.492	0.669	0.549
<i>a3ip2</i>	19	12	0.00866	0.00866	0.00000	0.00644	0.00644	0.816	0.959	-0.707	-0.088
<i>erd3</i>	5	4	0.00172	0.00239	0.00000	0.00136	0.00190	0.380	0.520	-1.223	-0.758
<i>nac1</i>	10	9	0.00254	0.00944	0.00072	0.00243	0.00698	0.411	0.112	-0.144	-0.063
<i>gatabp1</i>	13	8	0.00586	0.02030	0.00285	0.00693	0.01900	0.644	-0.395	-0.419	-0.490
<i>set-like-b</i>	4	5	0.00119	0.0000	0.00185	0.00149	0.00000	0.533	-0.385	-1.576	-1.401
<i>0_9389_01</i>	4	3	0.00260	0.00355	0.00000	0.00150	0.00205	0.419	1.356	-0.348	0.249
<i>CL180Contig1_03</i>	2	3	0.00023	0.00030	0.00000	0.00062	0.00081	0.131	-0.906	0.658	0.196
<i>0_1688_02</i>	7	5	0.00286	0.00461	0.00000	0.00200	0.00323	0.406	0.944	0.241	0.574
<i>0_12929_02</i>	8	7	0.00051	0.00045	0.00062	0.00203	0.00206	0.279	-1.710	0.377	-0.420
Average	8.2	6.3	0.00295	0.00537	0.00062	0.00289	0.00470	0.450	0.000	-0.247	-0.165

**Table 2. Nucleotide diversity of the ten nuclear single copy loci.** All summary statistic were calculated in DNAsp v. 5.10.01. For each locus, we calculated the observed number of haplotypes ( $H$ ), haplotype diversity ( $Hd$ ), the number of segregating sites ( $S$ )<sup>41</sup>, nucleotide polymorphism at total sites ( $\theta_w$ ), silent sites ( $\theta_{ws}$ ), and the nucleotide diversity at total sites ( $\pi_t$ ), silent sites ( $\pi_s$ ), and nonsynonymous sites ( $\pi_a$ )<sup>42</sup>, Tajima's  $D$ <sup>43</sup>, Fu and Li's  $D^*$  and  $F^{*44}$ .



**Figure 2. Median-joining network of haplotypes found in each of the ten nuclear single-copy loci.** Color scale represents haplotypes found in different regions: grey, Changbai Mts.; red, Xiaoxingan Range; yellow, Korean peninsula; blue, Sikhote-Alin Mts.

Although CM/KP and XR probably had separate refugia during the glaciation period, the populations from the two mountain ranges shared almost all dominant haplotypes in both nuclear and chloroplast markers (Fig. 1b and 2), which may be due to gene flow and ancestral polymorphism. It is possible that the CM/KP and XR refugia shared most genetic variation for several glacial cycles. However, we still observed new, unique mutants in each refugium. For example, haplotype H3 of *CL180Contig1\_03* was unique to XR, and H4 of *0\_1688\_02* was unique to CM/KP (Fig. 2). From another perspective, homogenizing gene flow may obscure differentiation between the populations of the two mountain ranges. Our msABC simulation estimated that the number of migrants per generation from CM/KP to XR was slightly larger than one individual but that it was less than one individual in



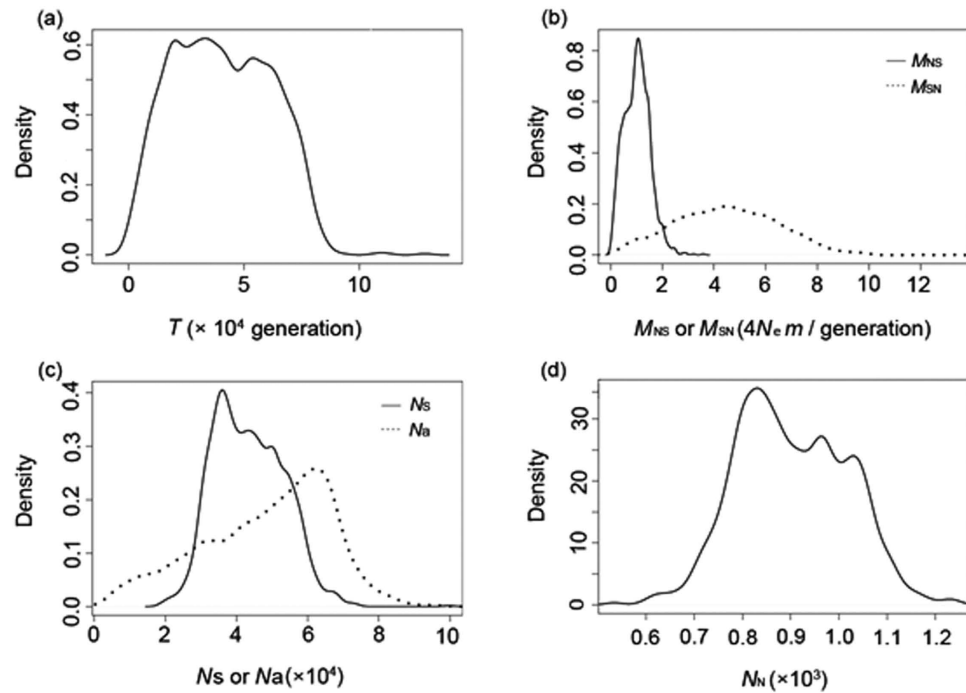
**Figure 3. Two demographic models of the coalescence time of two regional Korean pine populations, Changbai Mts. (including adjacent regions) and Xiaoxingan Range.** Their corresponding posterior probability estimated with the ABC model selection procedure is shown below the model. (a) Single LGM refugium model (SRM): the coalescence time of the two regional populations was previous to the LGM. (b) Multiple LGM refugia model (MRM): the coalescence time of the two regional populations was subsequent to the LGM.  $N_N$  and  $N_S$  denote the long-term effective population size of the Changbai Mts. (including the adjacent regions) and Xiaoxingan Range, respectively.  $N_a$  is the effective population size of the most common recent ancestor of the two regional populations.  $M_{SN}$  and  $M_{NS}$  are the long-term average gene flow from the Changbai Mts. to the Xiaoxingan Range and from the Xiaoxingan Range to the Changbai Mts., respectively.  $T$  denotes the coalescence time of the two regional populations.

	$N_S$	$N_N$	$N_a$	$T$	$M_{SN}$	$M_{NS}$
2.5% HPD	26854	719	9525	5032	0.16	0.06
Median	43413	869	53881	39568	1.08	0.24
Mean	43311	894	50573	39608	1.07	0.26
Mode	46494	830	63703	39044	1.28	0.24
97.5% HPD	62897	1107	85222	75836	2.08	0.55

**Table 3. Results of parameter estimation based on multiple LGM refugia ABC model in Fig. 3.**  $N_N$  and  $N_S$  denoted the long-term effective population size of Changbai Mt. (including its adjacent regions) and Xiaoxingan Range, respectively.  $N_a$  was the effective population size of the most common recent ancestor of the two regional populations.  $M_{SN}$  and  $M_{NS}$  were the long-term average gene flow from Changbai Mt. to Xiaoxingan Range and from Xiaoxingan Range to Changbai Mt., respectively.  $T$  denoted the coalescent time of the two regional populations.

the reverse migration direction (Table 3, Fig. 4). Simplification of the ABC models results in a gene flow estimate that is a long-term average for the time since the separation of the two populations. When the two populations contracted during the LGM, gene flow might have been reduced due to population isolation. During the interglacial period, gene flow might have exceeded one individual ( $N_e m$ ) due to population expansion and the strong dispersal capability of pine pollen (discussed above). Wright<sup>36</sup> suggested that when gene flow between populations exceeds one individual per generation, genetic differentiation will be eliminated.

The ABC simulation results suggested that the northern refugium located in XR would have been much smaller than the southern refugium in CM/KP. The effective population size of XR was less than 1200, nearly 50 times smaller than that of CM/KP (Table 3, Fig. 4). If both the CM/KP and XR regions sustained LGM refugia, the XR environment was probably harsher than that of CM/KP during the cold period. Scattered locations with suitable local environments (such as humid valleys) in XR might have provided habitats for relict populations in XR<sup>21</sup>. An ice cap was absent in NEA during the LGM, so the Korean pine might not be the only example of a species with northern microrefugia in the XR region. For example, a population of *Fraxinus mandshurica* is found in XR (near population TWH in our study, Fig. 1). This population possesses genetic variation that is divergent from other populations in NEA based on nuclear microsatellite data<sup>15</sup>. A population of *Quercus mongolica* near location FZ harbored a unique chlorotype with a very high frequency (Zeng *et al.* unpublished data), and this location has been considered a northern microrefugium. However, we again note that long-term average migration from XR to CM/KP is quite low, less than one individual per generation ( $N_e m$ , Table 3, Fig. 4). We cannot exclude the possibility that gene flow during the interglacial period would be much larger, as discussed above. As the potential microrefugium in XR was located peripherally, it is possible that population expansion was constrained, after the LGM, compared to the southern macrorefugium in CM/KP. This is probably due to the suboptimal environment in XR and potential genetic constraints in small populations of refugia<sup>7</sup>. In addition, the flowering season of Korean pine is from mid-to late June<sup>37</sup> and is under the influence of the summer monsoon in NEA. Hence, the northward monsoon may promote migration via pollen from CM/KP to XR while limiting the north-to-south pollen movement.



**Figure 4. Probability densities of six demographic parameters for the multiple LGM refugia models for Korean pine.** (a)  $T$ ; (b)  $M_{SN}$  (dotted line) and  $M_{NS}$  (solid line); (c)  $N_s$  (solid line) and  $N_a$  (dotted line); (d)  $N_N$ . The legend is the same as that of Fig. 3.

In summary, based on nuclear sequencing markers and ABC simulation statistics, we concluded that both the Changbai Mts. (including the Korean peninsula and the Sikhote-Alin Mts.) and the Xiaoxingan Range had relict populations of Korean pine during the LGM. The coalescence time between the populations of the two mountain ranges was at least 5000 generations ago, i.e., an estimated 0.4 million years ago. Although the two mountain ranges shared most of the total genetic variation, unique alleles exist in both regions for all studied markers. In contrast to the macrorefugia in the Changbai Mts. and its adjacent area, the microrefugium in the Xiaoxingan Range, combined with strong gene flow between regions, may have prevented genetic differentiation and eliminated the pattern of genetic variation over the species' range.

Due to the limited samples from the Korean peninsula and the Sikhote-Alin Mts., we treated those two regions and the Changbai Mts. as a single unit. Although the samples from the Korean peninsula were limited, we found unique alleles occurring with low frequency there (Fig. 2). Some studies have reported that the Korean peninsula might be a refugium separate from the Changbai Mts<sup>16</sup>. Additional plant samples and details of refugia locations in the Changbai Mts. and the adjacent areas would provide more insight into the postglacial evolutionary history of Korean pine.

## Methods

**Sample Collection.** During 2009–2012, we collected leaf samples from 75 adult trees in 21 natural populations covering the entire range of Korean pine in the NEA. From each population, 2–9 individuals were sampled, and the sampled individuals were at least 30 m apart from each other (Fig. 1, Table S1). The needles were dried and preserved in silica gel until DNA extraction.

**DNA extraction, amplification, and sequencing.** Total DNA was extracted from needles using a Plant Genomic DNA Kit (Tiangen Biotech, Beijing, China). We selected ten polymorphic nsc loci based on 44 loci initially developed for other pine species<sup>23,24,38</sup>. These loci could be stably amplified with a single polymerase chain reaction (PCR) band via agarose gel electrophoresis. The primer sequences, annealing temperatures, size of each PCR product, and the putative function and structure of these loci are shown in Table 2. Three intergenic chloroplast fragments, *psbA-trnH*, *trnS-trnG* and *trnL-trnF*<sup>39</sup>, were examined in preliminary studies, and little genetic diversity was found in loci *trnS-trnG* and *trnL-trnF* (one of 100 individuals had a different haplotype at locus *trnS-trnG*). Therefore, we only report the results from *psbA-trnH*, which was polymorphic. PCR products were directly sequenced using an ABI 3730 automated sequencer (Applied Biosystems) from BGI Tech Co., Ltd., Beijing, China. Each haplotype found was also verified by a cloning method. A pGEM-T Easy Vector (Promega, Fitchburg, WI, USA) was used in the cloning procedure, and at least five clones were sequenced. All sequence data have been deposited in GenBank, with accession nos. KP182424–KP182921, KT993573–KT995096.

**Genetic diversity analysis.** Sequences were aligned using the CodonCode Aligner 3.6.1 Program (<http://www.codoncode.com/aligner/>) with the “Muscle” module. Each individual sequence was independently checked by two researchers, and heterozygous sites were marked by hand. Sequences of all 10 nsc loci were phased in

DNA SP 5.10.01 using default settings<sup>40</sup>. Among the 10 nsc loci, only *IFG8612* had two indels (3 bp and 55 bp). We replaced the two deletions of *IFG8612* with almost the same sequence as the inserts, with only one base pair difference. Hence, each indel was treated as a nucleotide substitution and assumed to evolve by nucleotide substitution.

Population genetic variation was assessed by the following parameters: the observed number of haplotypes ( $H$ ), haplotype diversity ( $Hd$ ), the number of segregating sites ( $S$ )<sup>41</sup>, nucleotide polymorphism at total sites ( $\theta_w$ ), silent sites ( $\theta_{ws}$ ), and the nucleotide diversity at total sites ( $\pi_t$ ), silent sites ( $\pi_s$ ), and nonsynonymous sites ( $\pi_a$ )<sup>42</sup>. All of the loci were tested for departures from neutrality using Tajima's  $D$ <sup>43</sup>, Fu and Li's  $D^*$  and  $F^{*44}$ . The significance of each test was determined using 1000 coalescence simulations. The aforementioned statistics were computed using the DnaSP program v. 5.10.01<sup>40</sup>.

Median-joining networks were reconstructed in Network 4.6.1.2 for haplotypes found in each of the 10 nsc loci and the cpDNA<sup>45</sup>.

**Latitudinal gradient analysis of genetic diversity.** Pairwise genetic differences within populations ( $\pi_{xi}$ ) and between populations ( $\pi_{xyi}$ ) were calculated for each sampled population and each locus ( $i$ ) using Alequin 3.5 software<sup>46</sup>. The total pairwise genetic differences over ten nsc loci within populations ( $\pi_x$ ) and between populations ( $\pi_{xy}$ ) were calculated by the sum  $\pi_{xi}$  or  $\pi_{xyi}$  of each locus divided by the total sequence length of the ten nsc loci.

To verify the potential effect of sample size on genetic diversity within populations, rarefied allelic richness and private allelic richness were calculated using HP-Rare<sup>47,48</sup>. Haplotypes of each locus were coded as numbers and imported into the program. Only populations with 3 or more samples were used in this analysis. The two samples from population XUS were combined into BHS due to the approximate sample location. Six genes were set during the rarefied calculation.

The Pearson correlation coefficient between latitude and the genetic diversity indices was calculated and tested using R 3.1.0 (<http://www.r-project.org/>). Mantel tests of the correlation between genetic distance ( $\pi_{xy}$ ) and geographic distance (kilometers) were calculated using GenAlEx 6.1<sup>49</sup>.

**Historical demography analysis.** To test the single refugium vs. multiple-refugia hypotheses for Korean pine, two ABC models were established. The single LGM refugium ABC model (SRM) hypothesized that CM/KP populations and XR populations coalesced after the LGM, while the multiple LGM refugia (MRM) ABC model hypothesized that CM/KP and XR populations coalesced before the LGM. In both models, the migration direction and magnitude were unrestricted. We had limited samples for KP and the Sikhote-Alin Mts. (three populations and 12 individuals total), so we treated populations in CM, KP and the Sikhote-Alin Mts. as a single southern population (marked as CM/KP), while we treated populations in XR as a northern population (marked as XR). Populations FZ, QS and RH were located far from the center of CM and near the XR region (Fig. 1). Hence, we did not include these three boundary populations in the ABC simulation. We performed the simulation with the program msABC<sup>50</sup>.

In the ABC models, we set the average mutation rate for the ten nsc loci as  $1.31 \times 10^{-9}$  according to a previous estimate for *Pinus* species<sup>51</sup>. The initial effective population size was set as 10,000, which was close to the estimate based on the 10 loci. The generation time for Korean pine is unknown. Korean pine reaches maturity around 80–140 years<sup>34</sup>, whereas the data from a 5 ha clearcut plot of Korean pine in the NEA suggested that the average age of reproduction of Korean pine was approximately 200 years<sup>33</sup>. To avoid model selection bias due to the long generation time, we used a generation time of 100 years when we set the priors. In the SRM, we set the prior for the coalescence time of Korean pine as no more than the LGM (0–24,000 years before the present; the msABC-required transformation is 0–0.006). In the MRM, the coalescence time was set to be greater than the LGM. Korean pine is a Tertiary relict species<sup>52</sup>, so a large upper limit for the coalescence time was set as 10 million years before the present. Hence, the msABC-required prior transformation for the MRM was 0.0061–2.5. We set the northward and southward migration in both models as unlimited ( $4N_m$ : 0–1000). Details of the models and the msABC executable statements are shown in Fig. 3 and Table S2.

For each model, we simulated one million steps. Model selection was conducted in the statistics package abc in R 3.1.0. The “mnlogistic” method was used for model selection, with a tolerance rate of 0.001. We used 16 summary statistics that were closely related to gene flow and effective population size, including the mean and variance of segregating sites within populations and between populations  $S$ <sup>41</sup>, pairwise nucleotide differences within populations and between populations  $\pi$ <sup>42</sup>, genetic divergence ( $F_{ST}$ ) and shared allelic richness. The Bayes factor was used to determine the best supported model using the observed data<sup>53</sup>. To avoid random error, each model was simulated independently twice, using an independent model selection procedure.

For the best supported model, an additional four million steps were simulated. With the five million simulations, we estimated parameters such as the coalescence time of the CM/KP and XR populations, effective population size, and the migration rate between the CM/KP and XR populations. To ensure that the estimated range was located in the prior range, a log transformation was applied during the procedure. Neuronet's method with  $nnet = 50$  was used in the parameter estimation.

## References

- Overpeck, J. T., Webb, R. S. & Webb, T. Mapping eastern North-American vegetation change of the past 18 ka: No-analogs and the future. *Geology* **20**, 1071–1074, doi: 10.1130/0091-7613(1992)020<1071:menavc>2.3.co;2 (1992).
- Bennett, K. D., Tzedakis, P. C. & Willis, K. J. Quaternary refugia of North European trees. *J Biogeogr* **18**, 103–115, doi: 10.2307/2845248 (1991).
- Shi, Y., Cui, J. & Su, Z. *The quaternary glaciations and environmental variation in China*. (Hebei Science and Technology Publishing House, 2005).



4. Harrison, S., Yu, G., Takahara, H. & Prentice, I. Palaeovegetation (Communications arising): diversity of temperate plants in east Asia. *Nature* **413**, 129–130 (2001).
5. Yu, G. *et al.* Palaeovegetation of China: a pollen data-based synthesis for the mid-Holocene and last glacial maximum. *J Biogeogr* **27**, 635–664, doi: 10.1046/j.1365-2699.2000.00431.x (2000).
6. Winkler, M. G. & Wang, P. K. in *Global Climates: since the Last Glacial Maximum* (eds Wright, H. E. *et al.*) (University of Minnesota Press, 1994).
7. Edwards, M. E., Armbruster, W. S. & Elias, S. E. Constraints on post-glacial boreal tree expansion out of far-northern refugia. *Global Ecol Biogeogr* **23**, 1198–1208, doi: 10.1111/geb.12213 (2014).
8. Tzedakis, P. C., Emerson, B. C. & Hewitt, G. M. Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends Ecol Evol* **28**, 696–704, doi: http://dx.doi.org/10.1016/j.tree.2013.09.001 (2013).
9. Ren, G. & Zhang, L. A preliminary mapped summary of Holocene pollen data for Northeast China. *Quaternary Sci Rev* **17**, 669–688 (1998).
10. Cao, X., Herzschuh, U., Ni, J., Zhao, Y. & Böhmer, T. Spatial and temporal distributions of major tree taxa in eastern continental Asia during the last 22,000 years. *The Holocene* **25**, 79–91 (2015).
11. Aizawa, M., Kim, Z.-S. & Yoshimaru, H. Phylogeography of the Korean pine (*Pinus koraiensis*) in northeast Asia: inferences from organelle gene sequences. *J Plant Res* **125**, 713–723 (2012).
12. Qiu, Y., Fu, C. & Comes, H. P. Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of quaternary climate and environmental change in the world's most diverse temperate flora. *Mol Phylogenet Evol* **59**, 225–244, doi: 10.1016/j.ympev.2011.01.012 (2011).
13. Bai, W., Liao, W. & Zhang, D. Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. *New Phytol* **188**, 892–901 (2010).
14. Zeng, Y., Liao, W., Petit, R. J. & Zhang, D. Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. *Mol Ecol* **20**, 4995–5011 (2011).
15. Hu, L. *et al.* Nuclear DNA microsatellites reveal genetic variation but a lack of phylogeographical structure in an endangered species, *Fraxinus mandshurica*, across north-east China. *Ann Bot-london* **102**, 195–205 (2008).
16. Guo, X. *et al.* Evolutionary history of a widespread tree species *Acer mono* in East Asia. *Ecol Evol* **4**, 4332–4345 (2014).
17. Kim, Z., Hwang, J., Lee, S., Yang, C. & Gorovoy, P. Genetic variation of Korean Pine (*Pinus koraiensis* Sieb. et Zucc.) at allozyme and RAPD markers in Korea, China and Russia. *Silvae Genet* **54**, 235–245 (2005).
18. Feng, F., Han, S. & Wang, H. Genetic diversity and genetic differentiation of natural *Pinus koraiensis* population. *Journal of Forestry Research* **17**, 21–24 (2006).
19. Potenko, V. V. & Velikov, A. V. Genetic diversity and differentiation of natural populations of *Pinus koraiensis* (SIEB. et ZUCC.) in Russia. *Silvae Genet* **47**, 202–208 (1998).
20. Ma, J. The evolution of Korean pine forest. *Journal of Northeast forestry university* **25**, 66–70 (1997).
21. Kremenski, C. V., Liu, K.-b. & M. Macdonald, G. in *Ecology and Biogeography of Pinus* (ed David M. Richardson) Ch. 4, 95 (Cambridge University Press, 1998).
22. Eckert, A. J. *et al.* Patterns of population structure and environmental associations to aridity across the range of Loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* **185**, 969–982, doi: 10.1534/genetics.110.115543 (2010).
23. Ersoz, E. S., Wright, M. H., González-Martínez, S. C., Langley, C. H. & Neale, D. B. Evolution of disease response genes in loblolly pine: insights from candidate genes. *PLoS One* **5**, e14234 (2010).
24. Syring, J., Willyard, A., Cronn, R. & Liston, A. Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. *Am J Bot* **92**, 2086–2100 (2005).
25. Ren, G. *et al.* Genetic divergence, range expansion and possible homoploid hybrid speciation among pine species in Northeast China. *Heredity* **108**, 552–562 (2012).
26. Zhang, H., Zhang, M. & Williams, D. M. Genetic evidence and species distribution modelling reveal the response of *Larix sibirica* and its related species to quaternary climatic and ancient historical events. *Biochem Syst Ecol* **54**, 316–325 (2014).
27. Leffler, E. M. *et al.* Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biol* **10**, e1001388, doi: 10.1371/journal.pbio.1001388 (2012).
28. Excoffier, L., Foll, M. & Petit, R. J. Genetic consequences of range expansions. *Annu Rev Ecol Evol* **40**, 481–501, doi: 10.1146/annurev.ecolsys.39.110707.173414 (2009).
29. Petit, R. J. *et al.* Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* **300**, 1563–1565 (2003).
30. Williams, C. G. Long-distance pine pollen still germinates after meso-scale dispersal. *Am J Bot* **97**, 846–855, doi: 10.3732/ajb.0900255 (2010).
31. Robledo-Arnuncio, J. J. Wind pollination over mesoscale distances: an investigation with Scots pine. *New Phytol* **190**, 222–233, doi: 10.1111/j.1469-8137.2010.03588.x (2011).
32. Jaramillo-Correa, J. P., Beaulieu, J., Ledig, F. T. & Bousquet, J. Decoupled mitochondrial and chloroplast DNA population structure reveals Holocene collapse and population isolation in a threatened Mexican-endemic conifer. *Mol Ecol* **15**, 2787–2800, doi: 10.1111/j.1365-294X.2006.02974.x (2006).
33. Ge, J., Guo, H. & Chen, D. Study on age structure and spatial pattern of old-growth Korean Pine forest in lesser xingan mountain. *Journal of Northeast Forestry University* **18**, 26–31 (1990).
34. Xu, H. *Natural forests of Pinus koraiensis in China*. (Chinese Forestry Press, 2001).
35. Zhou, B. & Chen, Y. *Seeds of woody plants in China*. (Chinese Forestry Press, 2001).
36. Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97–259 (1931).
37. Institute of Applied Ecology, C. A. o. S. *Forests of Pinus koraiensis*. (Chinese Forestry Press, 1982).
38. Eckert, A. J. *et al.* Multilocus analyses reveal little evidence for lineage-wide adaptive evolution within major clades of soft pines (*Pinus* subgenus *Strobus*). *Mol Ecol* **22**, 5635–5650, doi: 10.1111/mec.12514 (2013).
39. Shaw, J. *et al.* The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am J Bot* **92**, 142–166 (2005).
40. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
41. Watterson, G. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7**, 256–276 (1975).
42. Nei, M. *Molecular Evolutionary Genetics*. (Columbia Univ. Press, 1987).
43. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
44. Fu, Y. & Li, W. Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709 (1993).
45. Bandelt, H.-J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**, 37–48 (1999).
46. Excoffier, L., Laval, G. & Schneider, S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50 (2005).
47. Kalinowski, S. Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conservation Genetics* **5**, 539–543, doi: 10.1023/B:COGE.0000041021.91777.1a (2004).
48. Kalinowski, S. HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Mol Ecol Notes* **5**, 187–189 (2005).

49. Peakall, R. O. D. & Smouse, P. E. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* **6**, 288–295, doi: 10.1111/j.1471-8286.2005.01155.x (2006).
50. Pavlidis, P., Laurent, S. & Stephan, W. msABC: a modification of Hudson's ms to facilitate multi-locus ABC analysis. *Mol Ecol Resour* **10**, 723–727 (2010).
51. Willyard, A., Syring, J., Gernandt, D. S., Liston, A. & Cronn, R. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Mol Biol Evol* **24**, 90–101 (2007).
52. Zhou, Y. & Zu, Y. *Geography of the vegetation in Northeast China*. (Science Press, 1997).
53. Jeffreys, H. *Theory of Probability*. (Clarendon Press, 1961).

## Acknowledgements

We thank Da-Yong Zhang, Shou-Hsien Li, Victoria Sork, Yan-Ping Guo, Jianqiang Zhang and two anonymous reviewers for their useful discussions and insightful comments. We thank to Xidi Guo, Sheng-Hong Wang and Tao Jiang, who assisted with sample collection. This work was supported by grants from the Natural Science Foundation of China (31210103911, 31421063, 31200328). Lei Bao is also supported by “the Fundamental Research Funds for the Central Universities”.

## Author Contributions

All listed authors are involved in the initial study design. L.B., A.K. and H.W. conducted the experiment and analysis. J.G., T.W., W.B. and R.C. conducted field samples and data analysis. H.W. and L.B. wrote the manuscript. All authors have reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Bao, L. *et al.* Contributions of multiple refugia during the last glacial period to current mainland populations of Korean pine (*Pinus koraiensis*). *Sci. Rep.* **5**, 18608; doi: 10.1038/srep18608 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>