

# Pedigree Reconstruction and Spatial Analysis for Genetic Testing and Selection In A *Larix Kaempferi* (Lamb.) Carrière Plantation

Kyungmi Lee

National Institute of Forest Science

In-Sik Kim

National Institute of Forest Science

Kyu-Suk Kang (✉ [kangks84@snu.ac.kr](mailto:kangks84@snu.ac.kr))

Seoul National University

---

## Research Article

**Keywords:** genetic testing, progeny trial, pedigree reconstruction, spatial autoregression, *Larix kaempferi*

**Posted Date:** September 29th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-879861/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Plant Biology on March 28th, 2022. See the published version at <https://doi.org/10.1186/s12870-022-03530-y>.

## Abstract

**Background:** *Larix kaempferi* is one of the major timber species in Northeast Asia. Demand for the reforestation of the species is rising in South Korea due to an increase in large timber production and utilization. However, progeny trials for the species have not been explored, making it challenging to foster advanced generations of tree improvement. In the present study, genetic testing and selection for diameter growth were conducted using pedigree reconstruction and phenotypic spatial distribution analysis in a plantation of *L. kaempferi*. The aim of the present study was to select the superior larch individuals using the pedigree reconstruction and phenotypic spatial distribution to substitute progeny trials. The plantation of seed orchard crops was established in 1990 and one-hundred and eighty-eight trees were selected as the study material. Genetic variation was investigated first to validate its adequacy as breeding material. Genetic testing was carried out using a model considering pedigree information and spatial autoregression of the phenotypes. The major environmental effects on growth were assessed to understand the site characteristics where the breeding materials were located.

**Results:** The expected heterozygosity of the mother trees and offspring were 0.672 and 0.681 presenting the corresponding level of genetic variation between two groups. The pedigree reconstruction using maternity analysis assigned one to six progenies to ninety-two candidate mothers. The accuracy of genetic testing was exceedingly increased with the animal model considering AR1 ⊗ AR1 structure compared to the animal model only. The estimated genetic variance of the former was 9.086 whereas that of the latter was 4.9E-5. The predicted breeding values of the offspring were ranged from -5.937 to 5.655 and the estimated heritability of diameter growth was 0.344. Shifts in the effects of the factors were detected using geographically weighted regression analysis, which provided information on the post-sectioning of the plantation.

**Conclusions:** The genetic testing approach based on pedigree reconstruction and phenotypic spatial distribution analysis was considered a useful analytical scheme that could replace or supplement progeny trials. Subdivision of the plantation as blocks of the progeny testing was accomplished *a posteriori* based on the changes of the environmental effects.

## Background

The genus *Larix*, a deciduous tree species in the family Pinaceae, provides timber with high economic value as well as great ecological value in forests. Larch tree improvement programs have been carried out for the major timber species and hybrids in the genus [1–4]. *Larix kaempferi* (Lamb.) Carrière has been one of the major timber species in Korea since its introduction in 1904. Large scale plantations of the species were cultivated until the 1970s [5]. However, the plantation areas had since been reduced due the low timber production and difficulties encountered in larch wood processing for a time [6]. Conversely, demand for larch timber and seeds has been increasing since 2000s consistent with increase in timber production and developments in wood processing technologies [6]. Today, forestry policies that could expand the larch plantation and revitalize their use in Korea are being pursued in response to increasing demand. Over the past five years, the plantation area under *L. kaempferi* has increased considerably, contrary to the area under other major species. It reached 273,000 ha of forest area, which is the second largest after the area under *Pinus densiflora* (2015) [7].

Several studies have been conducted only on partial stages of the *L. kaempferi* tree improvement program in Korea. Provenance tests and adaptability tests had been performed to introduce the superior provenance of larch until 1987 [6]. 145 plus trees were selected during 1959–1980 and 306.2 ha of seed orchard has been established to produce improved seeds of the species since 1968. [6]. In addition, studies to promote flowering and selection of multi-flowering clones have been carried out to address a key constraint in the unstable larch seed production [8, 9]. However, the tree improvement via progeny testing of the species was suspended due low levels of historic utilization, as well as a putatively limited gene pool caused by the introduction of only a small number of individuals to Korea.

Tree improvement is one of the essential activities in forest management, which is used to enhance the economic value of forests [10]. It aims to increase genetic gain in a gradual manner with recurrent testing and selection of advanced generations [11, 12]. The selection of advanced generation by genetic testing of larch has become necessary to ameliorate economic value of the species which has been actively used in plantation forestry in Korea. To shorten the time-consuming progeny trial approach of genetic testing, the alternative methods to the approach are required to meet current demand.

Pedigree reconstruction in plantations using seed orchard crops is one of the feasible approaches that could substitute the progeny trial method. Pedigree reconstruction analysis was first applied in tree improvement activities to detect the paternity arising from polymix breeding [13], followed by Breeding without Breeding [14], quasi-field trials [15], ad hoc breeding [16], and *in situ* genetic evaluation [17]. Pedigree reconstruction analysis is carried out to support existing tree improvement programs and to improve the accuracy of genetic parameter and breeding value estimates by substituting artificial crossing or wind pollination [18–22].

The accurate distinction of environmental effects is essential for proper genetic testing. Genetic testing considering spatial effects on the target traits has been attempted in previous studies in progeny trials or experimental forests [23]. Analytic models accounting for spatial structure could reduce the standard error associated with typical interfamily differences by 5–20% when compared to random complete block design experiments in genetic testing for tree height [24]. In addition, the fitness of the analysis was improved using the spatial model when compared to the incomplete block model when testing the growth properties of the Maritime pine (*Pinus pinaster*) [25]. The model has also been demonstrated to increase the accuracy of the log-likelihood and genetic value predictions based on restricted maximum likelihood [26]. The variation due to the spatial factors could be modelled effectively complementing the spatial autocorrelation. The supplementation of the spatial autocorrelation factor was more effective in the modeling of the spatial variation, significantly reducing the residual variation in a progeny trial of *Cunninghamia lanceolata* (Chinese fir) [23]. The genetic factor was efficiently explained considering the residuals caused by spatial autoregression of the target traits even in progeny trials based on a block design [26–30].

The objective of the present study was to perform genetic testing aided by pedigree reconstruction analysis and phenotypic spatial distribution analysis. The distinctive challenges of the present study to previous studies were the absence of *a priori* information on the family and the experimental design of the study

site. The study was conducted in phases with three detailed subjects to analyze the genetic and environmental effects at the individual level, including 1) genetic variation analysis to determine the adequacy of the individuals as breeding material followed by pedigree reconstruction analysis, 2) genetic testing based on pedigree information and phenotypic spatial autoregression, and 3) analysis of the environmental factors influencing phenotypic spatial distribution. The overall aims of the present study were to present schematic of genetic testing in plantations that could be applied 1) to acquire basic genetic variation data and pedigree information on breeding materials of seed orchard crops, 2) to ameliorate the accuracy of genetic testing in plantations while considering phenotypic spatial distribution, and 3) to understand the environmental gradients in a study site in which breeding materials are found.

The following hypotheses were formulated: 1) the plantation of the offspring from the plus trees in the seed orchard would retain a significant amount of the genetic variation even though information about the introduced population is missing, 2) the genetic and environmental effects would be detectable from growth differences within and between the local groups in the plantation, and 3) the site characteristics and *a posteriori* subdivision of the plantation would be discernible through the geographically weighted regression analysis.

## Results

### Genetic variation in plantation of seed orchard crops of *Larix kaempferi*

The number of alleles and levels of heterozygosity in the three groups of seed orchard (SO) and offspring (HC and HS) were estimated using seven microsatellite markers. The average number of alleles ( $N_A$ ) and effective alleles ( $N_E$ ) were 9.7 and 3.6, respectively (Table 1). The estimated genetic parameters of the mother tree group were analyzed by the National Forest Seed and Variety Center [8]. Genetic variation in the *L. kaempferi* was slightly lower than that in the Japanese natural population (0.717–0.762) based on the average expected heterozygosity ( $H_E$ ) [31]. However, the difference was considered to be due to the low number of polymorphic markers used, as the estimates were in the 0.518–0.860 range. The inference was supported by the even lower  $H_E$  in the introduced 140 plus trees and the offspring in China with 17 microsatellite markers, at 0.525 and 0.557, respectively [32]. The plantation of the seed orchard crops, with the considerable genetic variation compared to that in the group of mother trees, was considered the appropriate base population.

$N_A$  and  $N_E$  were similar between SO and HC, with averages of 10.0 and 3.7, respectively (Table 1). The differences in  $N_A$  and  $N_E$  between SO and offspring (HC or HS) were not significant, with HS having slightly lower estimates of 9.1 and 3.3, respectively. Furthermore, neither  $H_O$  nor  $H_E$  were significantly different between SO and offspring.  $H_O$  in HC and HS were 0.683 and 0.669, respectively, showing little difference from the 0.677 in SO.  $H_E$  in the two offspring groups were 0.681 and 0.656, with no significant difference from 0.672 in SO. In a comparison between the two offspring groups (HC and HS), the differences in genetic variation were not significant, although some of the estimates were slightly lower in HS. The latter had lower  $N_A$ ,  $N_E$ ,  $H_O$ ,  $H_E$  values in except in the analysis with bcLK033 and LK4178.  $N_A$  in bcLK033 was similar between the two groups, but  $N_E$  was somewhat higher in HS. Both estimates of LK4178 were higher in HS than in HC, exhibiting distinctiveness from the other markers.

Table 1  
Genetic variation across populations of *Larix kaempferi* based on  
microsatellite marker analysis

Group	Locus	$N_A$	$N_E$	$H_O$	$H_E$	$F_{IS}$
Plus trees <sup>1)</sup>	<b>Total</b>	<b>10.0</b>	<b>3.7</b>	<b>0.677</b>	<b>0.672</b>	<b>-0.010</b>
		<b>(1.7)</b>	<b>(0.7)</b>	<b>(0.059)</b>	<b>(0.058)</b>	<b>(0.027)</b>
Offspring (Hongcheon)	<b>Total</b>	<b>10.0</b>	<b>3.7</b>	<b>0.683</b>	<b>0.681</b>	<b>-0.006</b>
		<b>(1.9)</b>	<b>(0.7)</b>	<b>(0.055)</b>	<b>(0.050)</b>	<b>(0.035)</b>
	bcLK033	12.0	4.0	0.754	0.754	-0.002
	bcLK224	12.0	5.0	0.802	0.804	-0.001
	bcLK235	17.0	3.2	0.659	0.694	0.048
	bcLK241	4.0	2.1	0.572	0.526	-0.092
	LK4146	6.0	2.2	0.591	0.537	-0.104
	LK4170	14.0	7.0	0.908	0.860	-0.059
	LK4178	5.0	2.4	0.492	0.592	0.166
Offspring (Hwaseong)	<b>Total</b>	<b>9.1</b>	<b>3.3</b>	<b>0.669</b>	<b>0.656</b>	<b>-0.026</b>
		<b>(1.4)</b>	<b>(0.5)</b>	<b>(0.053)</b>	<b>(0.051)</b>	<b>(0.033)</b>
	bcLK033	12.0	4.6	0.723	0.785	0.075
	bcLK224	12.0	4.8	0.785	0.795	0.010
	bcLK235	11.0	3.1	0.645	0.683	0.052
	bcLK241	4.0	1.8	0.452	0.439	-0.034
	LK4146	6.0	2.1	0.577	0.529	-0.093
	LK4170	13.0	4.0	0.881	0.752	-0.177
	LK4178	6.0	2.6	0.621	0.612	-0.018
<b>Overall</b>	<b>Total</b>	<b>9.7</b>	<b>3.6</b>	<b>0.676</b>	<b>0.671</b>	<b>-0.014</b>
		<b>(0.9)</b>	<b>(0.3)</b>	<b>(0.031)</b>	<b>(0.029)</b>	<b>(0.018)</b>

#### Pedigree reconstruction of individuals in the *L. kaempferi* plantation

The average polymorphic information content (PIC) was 0.638 and the non-exclusion probability of identity was as low as 4.1E-07, based on the seven microsatellite markers. Each individual was identifiable as none of them showed the coincident genotypes in the seven loci. The non-exclusion probability of the first parent was 0.055. Ninety trees corresponding to 27% of the overall trees were assigned at relaxed confidence level (Table 2). The assignment of the other 244 trees was accomplished but did not fall within the confidence level. The proportions of assignments within the relaxed confidence level were 23% and 25% in HC and HS, respectively (Table 2). The low confidence level in the present study was considered to be derived from the characteristics of the used markers in the form of similar proportions of assignments at confidence levels in the two groups. Both exclusion probability and likelihood are considered in the maternity analysis by CERVUS. The low confidence level of assignment is caused by the negative score of the logarithm-of-odds (LOD) in the case of frequent major alleles or discordant genotypes in relation to the mother candidates.

The maternity analysis assigned the offspring to 92 and 79 candidate mothers in HC and HS respectively (Table S1). The total number of the candidate mothers was 112. The number of mutually exclusive mother trees was 33 in HC and 20 in HS, displaying the diverse constitution of the mother trees. The results potentially indicated that the genetic constitution would be distinguished by the reproduction year considering the samples of the seed orchard crops were produced in different years. The number of progenies by each mother tree were 1–6 trees in HC. The mother trees with six progenies were CN05, JPR-4, and CN09, and the 42 mother trees had only the one offspring. The number of progenies by each mother tree were 1–5 trees in HS. The assignment revealed that the mother trees GB12 and GB27 each had five progenies, and 36 mother trees had one offspring in HS. The offspring by mother trees were relatively evenly dispersed in the plantation (Fig S1–S5). Similarly, majority of the mother trees had one offspring according to the results of pedigree reconstruction analyses for *Abies nordmanniana* and *Larix occidentalis* [15, 17].

Table 2  
Maternity analysis of individual trees by population of *Larix kaempferi*

Level	Confidence (%)	Observed assignments (%)	
		Hongcheon	Hwaseong
Strict	95	10 ( 5%)	6 ( 4%)
Relaxed	80	53 ( 23%)	37 ( 25%)
Low	< 80	135 ( 71%)	109 ( 75%)
Total		188 (100%)	146 (100%)

### Genetic testing in the plantation of *L. kaempferi*

The prediction of breeding values was performed for each mother tree and the individual tree in the plantation although the confidence level of the maternal assignment was rather low. The estimated variance of the genetic effect was infinitesimal, at 4.9E-05, and the estimate for the residual was 25.07. Consequently, differences among the individuals were imperceptible (Table S2). The maximum predicted breeding value was 0.00003 for hc1485, with a DBH 29.7 cm, and the minimum estimates was - 0.00002 for hc1442, with a DBH of 6.6 cm. The correlation coefficient of the observed and predicted value was 0.94, which could be explained with predictions considering not only observations but also maternal information.

To separate the genetic effects with increased accuracy, the environmental effects were investigated to be accounted in the model of genetic testing. The spatial autocorrelation of phenotype was significant using Moran's I index ( $p < 7.4E-12$ ). The results of seven types of spatial models, including exponential, gaussian, spherical, linear, rational, AR1, AR1⊗AR1, and the basic animal model were compared to assess the model that would most adequately explain the phenotypic spatial structure of the plantation. The distribution of the phenotype was rasterized for use in the spatial analysis with AR1⊗AR1 (Fig. 1). The rasterization was cautiously processed to prevent data loss due to adjacency between individuals, considering their irregular distribution.

The integration of the animal and AR1⊗AR1 model (animal + AR1⊗AR1) yielded the highest fitness followed by the basic animal model, based on the log likelihood and AICc (Table 3). The AICc of the animal + AR1⊗AR1 model was largely distinct from those of the other models, highlighting its superior fitness. The fitness of the model with spatial AR1 was very low, suggesting the need for the use of a two-dimensional first-order autoregression model. The fitness and residual of the animal model and animal + AR1⊗AR1 model were confirmed in the present study (Fig. 2, Fig. 3). The residual was reduced in the animal + AR1⊗AR1 model when compared with in the animal model. In the meanwhile, the fitness was relatively low for the individuals with both extreme observations (individuals with either large or small DBH). The results implied that the trees with poor growth were more inferior than predicted, and trees with strong growth were more superior. The reason for this phenomenon was potentially the exceeded spatial autocorrelation for those individuals. However, no distortion of the selection was expected due to over or underestimation of the superior (high DBH) and inferior (low DBH) individuals.

Table 3  
Comparison of the fitness of spatial autoregression models of DBH distribution in *Larix kaempferi* stand

Spatial model	Intercept	Degree of freedom	Log likelihood	AICc <sup>1)</sup>
AR1⊗AR1	18.07	3	-553.6	1113.1
Animal model	17.751	2	-568.9	1141.7
Rational	17.8	5	-568.1	1146.5
Gaussian	17.76	5	-568.1	1146.6
Spherical	17.76	5	-568.3	1147.0
Exponential	17.8	5	-568.4	1147.2
Linear	17.76	5	-568.4	1147.2
AR1	18.18	4	-603.9	1216.0
<sup>1)</sup> Second-order Akaike information criterion				

The variance component of genetic factor was 9.086, and the heritability was estimated as 0.344 (Table 4). The accuracy of the genetic testing improved following the elimination of the environmental effects, with a heritability estimate of 0.224 for *L. kaempferi* DBH [1]. The predicted breeding value (PBV) of each individual also reflected the enhanced power of explanation of the animal + AR1⊗AR1 model (Table S2). The maximum and minimum PBV were 5.655 and - 5.937, respectively, showing the apparent difference between the animal model and the animal + AR1⊗AR1 model. The Spearman correlation coefficient ( $r$ ) between the observation and the prediction using the animal + AR1⊗AR1 model was 0.87 (Fig. 4). The weakened correlation was considered to be due to the consideration of the environmental effects when evaluating the phenotypes. The  $r$  between the prediction by the animal model and the animal + AR1⊗AR1 model was 0.89, revealing the dissimilar estimates obtained by the two models. The breeding values predicted by the two models were compared to confirm the frequent changes in rank (Table S2) reflecting the practicality of spatial analysis approaches in genetic testing.

Table 4  
 Variance components based on animal + AR1 ⊗ AR1  
 model analysis of DBH growth of *Larix kaempferi*

	Estimated variances	Standard error
Genetic	9.086	7.634
Spatial	12.478	3.894
Residual	4.885	7.393

PBV of mother trees were estimated, with the highest value obtained in KW02, followed by JB10, NO05, NO03, and GB25 (Table S3). The PBV of GB23 was the lowest, at -3.90, followed by JB08, CN15, and KW56. In the case of KW02, which had the highest PBV, five trees were assigned to its family in HC (Table S1, Table S3). All the five individuals, including hc1327, hc1427, hc1485, hc1538, and hc1549, had positive PBVs of 1.507, 4.846, 3.621, 2.171, and 0.117, respectively. JB10, which had the second highest PBV, was the mother tree of hc1335, which showed the highest PBV among the offspring, at 5.655. The progenies of NO05 were hc1436 and hc1552, with PBVs of 3.971 and 2.604, which were the 6th and 24th highest PBVs among the 188 trees. NO03 had two progenies (hc1437 [PBV: 3.795], hc1463 [PBV: 2.442]), and GB25 had five progenies (hc1423 [PBV: 2.390], hc1468 [PBV: -1.163], hc1519 [PBV: 0.875], hc1520 [PBV: 3.900], and hc1559 [PBV: 2.390]). GB23, the mother tree with the lowest PBV, had two progenies, hc1509 and hc1516, with PBVs of -3.812 and -5.937, respectively. The latter had the lowest breeding value among the 188 trees. The PBVs of mother trees were considered to reflect the estimates of their progenies (Fig. 5), however the accuracy of analysis was limited due to the insufficient number of mother trees and their progenies.

The individual trees and their mother trees were evaluated with PBVs based on the animal + AR1 ⊗ AR1 model. Considering the spatial autoregression as the products of innumerable biotic and abiotic factors proved useful in the plantation genetic testing, in combination with the pedigree reconstruction analysis.

#### Regression analysis of growth and environmental factors of *L. kaempferi*

The environmental factors of individual trees in the plantation of *L. kaempferi* assessed were elevation, slope, solar irradiance, and soil characteristics (Table S4, Fig S6). The average elevation was 390 m, and mostly distributed in the 374–418 m range. The aspects were east, south, and southwest, with an average slope of 30.7°, and mainly in the 17°–36° range. The wetness index of the geographic environment for each individual was 3.51 on average, ranging from 1.41 to 16.85, indicating that the trees were located near the valley section with low humidity. The average percentages of sand, silt, and clay were 69.7%, 21.8%, and 8.5%, respectively, and were within the 67.8–70.9%, 21.1–22.6%, and 8.0–11.1% ranges. The average pH was 5.4 and the average organic matter, total N, and available P contents, and CEC were 8.5%, 0.36%, and 65.6 mg/kg, and 13.6 cmol<sub>c</sub>/kg respectively. Among the exchangeable cations, K<sup>+</sup>, Na<sup>+</sup>, and Ca<sup>2+</sup> were within the optimal ranges for tree growth, while Mg<sup>2+</sup> was relatively low.

Geographically weighted regression (GWR) analysis was conducted using total N, aspect, wetness, clay content, and Na<sup>+</sup>, which were selected as the key factors influencing growth based on the correlation and regression analysis results. The AICc was sequentially reduced as the inclusion of the five factors showing the adequacy of factors (Fig. 6). The fitness of the GWR model was higher than that of the global regression model, indicating the significance of local spatial differences within the plantation (Table 5).

Table 5  
 Comparison between global regression model and geographically weighted regression (GWR)  
 model results

Variable	Global model		GWR model		
	Coefficient	Probability (> t )	mean of coefficient	Probability(>) [33]	
F <sub>1</sub>	F <sub>3</sub>				
(Intercept)	-40.88	0.006	-767.27	0.051	0.000
Total Nitrogen	163.34	0.000	2303.95		0.000
Aspect <sup>1)</sup>	-1.57	0.001	-1.23		0.000
Wetness	1.25	0.000	0.72		0.001
Clay	16.61	0.232	-28.10		0.000
Na <sup>+</sup>	-2699.57	0.218	3613.8		0.000
AIC <sup>2)</sup>	1109.1		1065.7		
AICc <sup>3)</sup>	1109.8		1087.8		
R <sup>2</sup>	0.20		0.38		
adjusted R <sup>2</sup>	0.18		0.30		

The ranges of GWR coefficients of the environmental factors were estimated (Table 6) and the changes were visualized by color to indicate the appropriate interval (Fig. 7). The changes in regression coefficients by grade were expressed through the comprehensive sectioning of the plantation. The GWR coefficients were distinguished among the north, south, and the valley section within the south, indicating the underlying disparity. The change of the effect, not the change of the factor *per se*, was thought to be the product of the interaction among the numerous environmental factors. The GWR analysis, conceived from the existing spatial autoregression, was useful in detecting the environmental gradients and facilitating a comprehensive understanding of the plantation characteristics. The plantation information could in turn facilitate systematic management and exploitation, and address some limitations of genetic testing with plantation.

Table 6  
Range of the coefficient of independent variables of geographically weighted regression model

Variable	Minimum	1st Quantile	Median	3rd Quantile	Maximum
Intercept	-5536.0	-47.9	28.3	70.3	139.6
Total N	60.2	86.6	118.8	293.2	15214.1
Aspect	-2.3	-1.8	-1.3	-0.6	-0.2
Wetness	-0.1	0.3	0.7	1.1	1.5
Clay	-88.6	-59.0	-28.7	-0.3	43.5
Na <sup>+</sup>	-7476.3	-150.2	3881.0	7578.5	13644.9

## Discussion

### Genetic variation in plantation of seed orchard crops of *Larix kaempferi*

Forest trees, which are long-lived plants that undergo outcrossing frequently, generally have high genetic diversity, excluding in the cases of endangered or rare species. The *L. kaempferi* seed orchard offspring was confirmed to retain considerable genetic variation in the present study. High genetic diversity was detected in the limited natural distribution range of *L. kaempferi* in Mt. Fuji even on fragmented habitats [31]. Frequent gene flow between populations was inferred from the low level of genetic differentiation despite the geographical distance in both Japanese larch and European larch [31, 34].

Studying the change in the genetic variation using seed orchard crops is necessary to detect and preclude decline in genetic diversity with increased inbreeding [35, 36, 37]. Similar levels of genetic diversity in seed orchards and their crops are frequently reported, dispelling concern over narrowed genetic base in tree improvement programs. The genetic diversity of the progeny trial corresponded to that of a naturally regenerated stand in *Picea sitchensis* and *Picea abies* [35, 36]. In addition, the genetic diversity of a plantation with seed orchard crops was not affected by a tree improvement program in *Cryptomeria japonica* [37].

However, genetic variation in larch should be assessed cautiously. Genetic variation in larch is influenced by the mating system, reproductive characteristics, and changes in descendants. Deficit of heterozygosity has been detected in *Larix sibirica* and *Larix gmelinii*, and attributed to inbreeding [38]. Furthermore, in a study on the larch mating system, the distance of pollen dispersal was reported to be relatively short [32, 39, 40], and the genetic structure of European larch over its entire distribution was more obvious than those of other conifers [41].

In summary, larch generally retain considerable genetic variation, similar to most forest trees. However, the species could be experience decreased genetic variation due to inhibited gene flow or severe fluctuation in flowering and seed production activities [6, 8, 42].

The first prerequisite in tree improvement is understanding the genetic variation in the natural population and in the selected breeding population [43]. Genetic variation is the key factor influencing natural and artificial selection activities [44] related with the coefficient of variance of the genetic factor. Genetic variation determines the genetic gain and the potential of improvement by influencing adaptability and resistance to biotic or abiotic stress [45]. Studying tree genetic variation is also essential for sustainable forest management and conservation [46], which could enhance the understanding of variation in adaptable traits including growth, wood characteristics, and resistance to disease or drought [47]. Understanding genetic variation is particularly essential prior to genetic testing in *L. kaempferi* plantation considering the reproduction and mating characteristic of the species. The additional introduction to enrich the gene pool of larch would assure the genetic diversity of breeding materials in long-term tree improvement program.

### Pedigree reconstruction of individuals in the *L. kaempferi* plantation

Using an increased number of polymorphic markers would be necessary to enhance the confidence level of the maternity assignment in *L. kaempferi*. A study investigated the accuracy of pedigree reconstruction using a microsatellite marker as the principal marker in the field of research [48]. The success rate of pedigree reconstruction was improved with an increase in the number of markers in the pedigree reconstruction analysis for European sturgeon (*Acipenser sturio*) [49]. The low confidence level of parental assignment was thought to be related with allele frequency in a *Abies procera* study, where the microsatellite marker revealed insufficient variation [48]. The separate experiment and genotyping of the SO and the offspring was the other plausible reason of a decrease in the confidence in this study. The simultaneous analysis of the overall sample would reduce the scoring error of genotyping [50].

The application of molecular biology technologies in tree improvement can be classified into two approaches [12, 48], including association between target traits and markers, such as the genetic dissection or genomic selection, and enhancing tree improvement efficiency using DNA markers, which includes pedigree reconstruction analysis [51, 52]. The development of the DNA markers with high resolutions increases the usefulness of the pedigree reconstruction [53]. Furthermore, it is expected that the confidence of assignment would be enhanced by the rapid development and commercialization of automated and high-throughput genotyping approaches, including SNP marker analysis [48].

The small number of assigned progenies by parent has been discussed as one of the main challenges of genetic testing with pedigree reconstruction [15], and it impedes researchers' ability to obtain sufficient numbers of samples required to establish statistical significance in genetic testing. Therefore, forward selection is considered desirable compared to the backward selection in the case. The latter has been recommended considering the relative ease of propagation following tree improvement acceleration [16]. However, forward selection was suggested in this study due to the possibility of deterioration of accuracy when performing genetic testing on mother trees.

#### Genetic testing in the plantation of *L. kaempferi*

Genetic gain following selection is expected in the genus *Larix* considering previous genetic testing [2, 17, 54–57]. The heritability of *L. decidua* height was 0.23 in the 25–35 age group [17], and the heritability values for height and DBH estimates of *L. kaempferi* were 0.549 and 0.224, respectively in the genetic testing of artificial crosses [1]. The infinitesimal genetic variance in the analysis with the animal model highlighted the need for isolating the environmental factor. The plantation was located in a highly heterogeneous site, which hampered the analysis of environmental effects. Prior pedigree reconstruction studies have mainly been carried out in experimental forests under intensive management, with a focus on the development of analytical methodologies [14–17, 21, 22, 58]. In addition, few substantive genetic testing studies have been conducted in the genus *Abies* [15, 16, 48], and the genetic resources of *L. decidua* have been evaluated based on locations [17]. The effective analysis of environmental effects is a major obstacle in genetic testing. Notably, the application of pedigree reconstruction in practical genetic testing has been accomplished in the study of genus *Abies* by supplementation with polymorphic SNP marker and spatial analyses.

The inferior fitness of the spatial model except AR1 $\otimes$ AR1 highlighted the importance of selecting the adequate spatial structure model beyond the consideration of the existence of spatial autocorrelation. The enhancement of genetic testing via spatial analysis with AR1 $\otimes$ AR1 has been reported [26–30]. Genetic testing with the model considering AR1 $\otimes$ AR1 structure in pedigree reconstruction analysis has been used in family selection in a *Abies bornmülleriana* progeny trial [16].

It is impossible to control the number of samples by family in genetic testing in plantations, contrary to the case in traditional progeny testing. Only one progeny was assigned to the multiple family in the present study as in the pedigree reconstruction analysis in *L. occidentalis* and *A. nordmanniana* [15, 17]. The spatial analysis was considered to be essential in the selection without the experimental design including the replication of the family. Spatial analysis is also considered to be useful in the progeny trials [23, 27, 29], in particular, under high environmental heterogeneity, for example in mountainous sites in Korea [59].

#### Regression analysis of growth and environmental factors of *L. kaempferi*

Few studies have investigated the effects of nitrogen (N) and other environmental factors on the growth of trees in the genus *Larix* [60]. Total N was generally positively correlated with the site index estimates in a *L. kaempferi* plantation [61]. N also promoted the biomass and photosynthetic rates of *Larix gmelinii* var. *japonica*  $\times$  *Larix kaempferi* seedlings under photosaturation [62]. Available N was essential for *Larix* growth in a region located in the north latitude 70° [63]. However, N load did not necessarily promote larch growth in another study [60]. Other soil characteristics have also exhibited inconsistent effects on larch growth. Organic matter and available P contents exhibited negative correlations with larch growth in Korea [61]; however, P did not influence larch seedling growth [62]. The inconsistencies reported could arise from the highly diverse conditions of the study sites, with complex interactions among site factors. Conversely, relatively consistent effects of wetness have been reported. Water availability has been reported to be a key factor influencing *L. sibirica* growth [64, 65], and the water stress has negative effects on *L. decidua* growth and respiration [66].

Understanding the major environmental factors influencing target traits is necessary for the ameliorating genetic gain and sustainable forestry [60]. Abiotic stress factors influence tree growth, and in turn forest productivity [67], so that site selection is vital [68]. In particular, *L. kaempferi* is highly sensitive to biotic and abiotic stress [62]. *L. kaempferi* is one of the larch species considered vulnerable to climate change, decreased distribution projected in future [69, 70]. Another study reported a reduction in the distribution range of the genus *Larix* in the southern hemisphere under drought stress, in contrast with expansion in Siberia and North America in the northern hemisphere [71]. Furthermore, *L. kaempferi* distribution is considered to have reduced in mountainous areas, shrinking to approximately 40,000 ha in Korea under RCP 4.5 [69]. The environmental effects had been studied in accordance with the analysis of variance component as one of the major topics in the tree improvement. However, few studies have assessed the effects of specific site characteristics at the individual tree level. Investigation of site characteristics using geographical information systems (GIS) was considered useful in the management and exploitation of plantations following genetic testing in this study.

## Conclusions

Genetic testing was conducted in a *L. kaempferi* plantation to establish a systematic scheme that could supplement progeny trials in tree improvement activities. Pedigree reconstruction and phenotypic spatial distribution were applied in the analysis of genetic and environmental effects on tree growth. The pedigree reconstruction was performed for individual trees following the assessment of genetic variation to identify the adequacy of breeding population. The genetic testing was implemented considering the spatial autoregression to dissect the environmental factors influencing the growth characteristics of a

plantation without experimental design. A detailed investigation of the environmental factors was executed using GWR to provide information to facilitate further management of breeding materials.

Several factors should be taken into account in the course of interpreting the results of the present study. The planting and management records of the study site would save the additional labor for the verification of the data. In addition, assessing the genetic diversity is necessary for preventing a decrease in genetic variation and inbreeding depression in advanced generations. Furthermore, appropriate selection strategies should be adopted based on the limited sample sizes obtained in the progeny. Forward selection was suitable in the present study considering the small progeny size per mother tree. The effective application of the pedigree reconstruction in genetic testing requires the combined analysis of phenotypic spatial distribution as well as the influence of numerous environmental factors. The extended practical application of such an analytical scheme is expected in practice considering projected advancements and developments in relevant technologies, including molecular markers and GIS.

## Materials And Methods

### Study sites and sampling

The *L. kaempferi* plantation of seed orchard crops, owned by the state, was selected as the study site based on plantation history. The reforestation was undertaken on the preexisting larch plantation in Hongcheon-gun (HC), Gangwon-do (37°35'24.0"N, 127°53'24.0"E) with 24,000 larch seedlings over an area of 8.0 ha in 1990. The remaining number of larch trees at present was 271 including the preexisted old trees. Therefore, the study materials were selected through screening to take the field circumstances into consideration preceding the analysis and in accordance with the national legislation. The growth data and the empirical yield table were inspected in the screening. The average height and DBH of 271 trees in the plantation were 16.0 m and 23.7 cm. The respective average values were 14.4 m and 18.7 cm under the lowest site index (16) and 19.5 m and 21.8 cm under the highest site index (22) based on the empirical yield [72]. The outlying 83 trees with DBH > 30 cm were excluded from further analyses. The overlapping existence of different age group of trees was sustained by deviation from the normal distribution pattern in of the growth data (Fig. 8). The rest of the 188 were considered to be the subsequent generation from the seed orchard based on the scarce chance of larch seedling occurrence in the larch forest (Fig. 9). The screening based on DBH was possible in the present study because the differences in growth between the excluded and remnant trees were rather distinct. One-hundred and forty-six trees from a five-year-old seed orchard planted in Hwaseong-si (HS), Gyeonggi-do, were additionally included in the analysis for the comparison with the HC plantation.

Three types of data were collected for the phenotype, genotype, and spatial distribution analyses of the individuals. Tree height and DBH, cambium or leaf samples, and GPS coordinates (Trimble Geo 7X, Trimble Inc., Sunnyvale, CA, USA) were collected simultaneously. Soil samples were collected at three points showing the discrete tree growth or environments.

#### Genetic variation and pedigree reconstruction of *L. kaempferi* plantation

The isolated DNA of 334 *L. kaempferi* trees were genotyped using seven microsatellite markers, which were used in a previous analysis on the mother trees in the seed orchard [73, 74] (Table 7). The genotype data of the mother trees used in the present study were assessed by the National Forest Seed and Variety Center (NFSV). The polymerase chain reaction (PCR) was carried out using FAM-labelled forward primers of each marker followed by genotype analysis using an ABI 3730xl Genetic analyzer and GeneMapper 4.0 (Applied Biosystems, Foster City, CA, USA).

Table 7  
Microsatellite markers of *Larix kaempferi* used in the present study

Marker	Primer name	Primer Sequence (5'→3')
bcLK033	bcLK033-F	GGAAATGTAGAGATGAGCAATAA
	bcLK033-R	AGGTGCGGTAGTACAAAGTGA
bcLK224	bcLK224-F	GGAGAGGCCACTACTATTATTAC
	bcLK224-R	ATGCGTTCCTTCATTCCTCT
bcLK235	bcLK235-F	TTCACCTGTGATCCTAGAGTTAGA
	bcLK235-R	AACCCCTAACCATATAATATCCA
bcLK241	bcLK241-F	TGAGGTTAGGAGCATCTCGT
	bcLK241-R	GTCCTTCATCGCCTCTTCTT
LK4146	LK4146-F	CAACATGTTTCTCCTACCACCA
	LK4146-R	TCAGACATTCCCAAACATGC
LK4170	LK4170-F	TTTTCCAAAGCCAAAATTCTACA
	LK4170-R	TATGAGCCCGACCCTATTTG
LK4178	LK4178-F	TCCACCTTAGCACTCCCCT
	LK4178-R	GGGCCTTTATAGGTTGGTT

Genetic variation in the larch plantation of seed orchard crops was estimated based on the number of alleles ( $N_A$ ), the number of the effective alleles ( $N_E$ ), observed heterozygosity ( $H_O$ ), expected heterozygosity ( $H_E$ ), and fixed index (GenAlEx 6.5) [75]. The data on the mother trees were provided by the NFSV [6].

The maternity analysis was conducted for the pedigree reconstruction of the *L. kaempferi* plantation. The individuals in the plantation were assigned to the mother trees in the seed orchard using CERVUS 3.0 based on the maximum likelihood approach [76, 77]. The confidence levels were set at 95% (restricted) and 80% (relaxed), with default values in the program. The error rate of genotype identification was specified as 0.01 and the confidence level was decided based on 10,000 simulations.

## Genetic testing using pedigree reconstruction and spatial distribution of phenotype

The growth correlations among adjacent individuals were observed in the study site. Spatial autocorrelation was tested based on DBH using Moran' I prior to genetic testing. As the existence of the spatial autocorrelation of DBH was identified, the spatial model was explored using seven spatial structures, including AR1, exponential, Gaussian, spherical, linear, rational, and AR1 $\otimes$ AR1. The best fitted model was selected by comparing the log-likelihood values and the second-order Akaike Information Criterion (AICc) of the eight linear mixed models, including the animal model [78, 79]. The accuracy of the genetic testing was improved using the AR1 $\otimes$ AR1 model, which is the Kronecker product of the autoregression matrix for rows and columns in several studies [80, 81].

$$y = Xb + Zu + \xi + \eta$$

where  $y$  is the vector of data;  $b$  is a vector of fixed effects;  $u$  is a vector of random effects;  $\xi$  is dependent spatial residuals and  $\eta$  is independent spatial residuals (nugget effects)

$$u \sim N(0, \sigma_s^2 R_{AR})$$

where  $\sigma_s^2$  is the spatial variance parameter;  $R_{AR}$  is a correlation structure of AR1( $\rho_x$ ) $\otimes$ AR1( $\rho_y$ ); AR1( $\rho$ ) is the first-order autoregressive correlation matrix.

The breeding values of mother trees and the individuals in the plantation were predicted by the best fitted model using R program [82] with ape, nlme, breedR, and MuMIn packages [83–86].

### Regression analysis of growth and environmental factors in the *L. kaempferi* plantation

The site and soil characteristics were investigated to analyze the influence of the environmental factors on *L. kaempferi* growth (Table 8). The site characteristics were extracted at a 5 × 5 m raster resolution from a digital topographic map at a 1:5000 scale. The soil characteristics are presented in Table 9, including pH, organic matter, total nitrogen, available phosphate, cation exchange capacity (CEC), K<sup>+</sup>, Na<sup>+</sup>, Ca<sup>2+</sup>, Mg<sup>2+</sup>, and electrical conductivity. The aspect (the direction the slope on the mountain was facing) of each raster cell was used to determine the transformed value between 0 and 2: northeast closer to 2, southwest closer to 0. The soil characteristics were obtained by kriging the results of the soil analysis of the three samples collected. The analysis of the site characteristics was carried out using ArcGIS 10.4 (ESRI, Redlands, CA, USA).

Table 8  
Site characteristics analyzed from soil samples and in *Larix kaempferi* stand in Hongcheon

Category	Site characteristics
Topography	Elevation (m), slope (degree), aspect <sup>1)</sup>
Solar	Solar radiation (MJ/cm <sup>2</sup> per day)
Water	Wetness <sup>2)</sup>
Soil	Sand, silt, clay, pH, organic matter, total nitrogen, available phosphoric acid, cation exchange capacity, K <sup>+</sup> , Na <sup>+</sup> , Ca <sup>2+</sup> , Mg <sup>2+</sup> , electrical conductivity
<sup>1)</sup> Linear transformed aspect (southwest [= 0] to northeast [= 2])	
<sup>2)</sup> Value of topographic wetness index	

Table 9  
Soil properties surveyed from soil samples in the study site

Sample no.	Tree no.	Texture			pH	Organic matter (%)	Total N (%)	Avail. P <sup>1)</sup> (mg/kg)	CEC <sup>2)</sup> (cmol/kg)	K <sup>+</sup> (cmol/kg)	Na <sup>+</sup> (cmol/kg)	Ca <sup>2+</sup> (cmol/kg)	Mg <sup>2+</sup> (cmol/kg)	EC <sup>3)</sup> (dS/m)
		Sand (%)	Silt (%)	Clay (%)										
1	1301	67.8	21.1	11.1	5.2	6.29	0.273	56.4	13.13	0.29	0.07	2.15	0.29	0.22
2	1403	69.5	22.6	7.9	5.8	9.43	0.443	66.6	13.35	0.43	0.05	5.31	0.70	0.19
3	1481	70.9	21.1	8.0	5.3	8.43	0.360	68.5	14.01	0.32	0.05	2.52	0.42	0.25

The local variations in the environmental effects on *L. kaempferi* growth were examined using the geographically weighted regression analysis. The method was used to test the variabilities of the local regression coefficients in the measurements with diverse spatial relationships.

$$y_i = \beta_{i0} + \sum_{k=1}^m \beta_{ik}x_{ik} + e_i$$

where  $y_i$  is dependent variable of location  $i$ ,  $x_{ik}$  is independent variable of location  $i$ ,  $m$  is the number of independent variables;  $\beta_{i0}$  is the intercept parameter of location  $i$ ,  $\beta_{ik}$  is the local regression coefficient of the  $k$ -th independent variable of location  $i$ ,  $\{e_i\}$  is the random error of location  $i$ .

The five environmental factors (clay, TN, aspect, wetness, and Na<sup>+</sup>) were used as the independent variables, and were selected based on the correlation and regression analysis. The bisquare kernel was applied and the value of bandwidth was specified as 132 in the geographically weighted regression analysis.

## Declarations

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

### Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

### Competing interest

The authors declare that they have no competing interests

### Funding

K L and I-S K have a grant from the National Institute of Forest Science (NIFoS), Korea and K-S has a financial support (2020182B10-2022-BB01) from the Korea Forest Service.

### Authors' contributions

K L and I-S K collected data and analyzed, and K-S K had role in study design. All authors prepared the manuscript together.

### Acknowledgements

Authors greatly appreciate for the permission of data collecting and utilizing to the National Institute of Forest Science in the Republic of Korea.

## References

1. Fukatsu E, Hiraoka Y, Matsunaga K, Tsubomura M, Nakada R. Genetic relationship between wood properties and growth traits in *Larix kaempferi* obtained from a diallel mating test. *J Wood Sci.* 2015;61:10-18.
2. Diao S, Hou Y, Xie Y, Sun X. Age trends of genetic plus tree group, early selection and family by site interactions for growth traits in *Larix kaempferi* open-pollinated families. *BMC Genet.* 2016;17:104.
3. Larsson-Stern M. Aspects of hybrid larch (*Larix × eurolepis* Henry) as a potential tree species in southern Swedish forestry. *Alnarp:Sveriges lantbruksuniversitet;2003.*
4. Jaquish BC. Western larch tree improvement in British Columbia. In: *Integrated research activities for supply of improved larch to tree planting: tree improvement, floral biology and nursery production. LARIX 2007: International Symposium of the IUFRO Working Group S2.02.07 (Larch Breeding and Genetic Resources).* Quebec city, Canada;2007.
5. National Institute of Forest Science (NIFoS). *Economic species of Larix kaempferi.* Research Book No. 55. Seoul, Korea;2012a
6. National Institute of Forest Science (NIFoS). *Tree improvement and establishment of breeding system in Larix kaempferi.* Research Reports no. 19-12. Seoul, Korea;2019.
7. Korea Forest Service (KFS). *Statistical yearbook of forestry (no. 49).* Daejeon, Korea;2019.
8. Kim I-S, Kim J-H, Kan J-T, Lee B-S. Clonal variation in female flowering of *Larix leptolepis* L. *Korean J Plant Res.* 2008;21(1):1-4.
9. National Institute of Forest Science (NIFoS). *Promotion of flowering and seed production of larch.* Research Data Collection No. 457. Seoul, Korea;2012b.

10. Fox TR, Jokela EJ, Lee AH. The evolution of pine plantation silviculture in the Southern United States. In: General Technical Reports SRS-75 (U.S. Department of Agriculture, Forest Service, Southern Research Station). North Carolina, USA; 2004.
11. Namkoong G, Kang HC, Brouard JS. Tree Breeding: Principles and strategies. New York, NY: Springer Verlag; 1988.
12. Grattapaglia D, Silva-Junior OB, Resende RT, Cappa EP, Müller BSF, Tan B, Isik F, et al. Quantitative genetics and genomics converge to accelerate forest tree breeding. *Front Plant Sci.* 2018;9:1693.
13. Lambeth C, Lee B-C, O'Malley DO, Wheeler N. Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing. *Theor Appl Genet.* 2001;103:930-943.
14. El-Kassaby YA, Lstibůrek M. Breeding without breeding. *Genet Res (Camb).* 2009;91(2):111-120.
15. Hansen OK, McKinney LV. Establishment of a quasi-field trial in *Abies nordmanniana* - test of a new approach to forest tree breeding. *Tree Genet Genomes.* 2010;6:345-355.
16. Xu J, Nielsen UB, Hansen OK. Ad hoc breeding of *Abies bornmülleriana* for Christmas tree production using a combination of DNA markers and quantitative genetics—a case study. *Tree Genet Genomes.* 2018;14:5.
17. Lstibůrek M, Schueler S, El-Kassaby YA, Hodge GR, Stejskal J, Korecký J, et al. In Situ Genetic Evaluation of European Larch Across Climatic Regions Using Marker-Based Pedigree Reconstruction. *Front Genet.* 11:1-8.
18. Doerksen TK, Herbinger CM. Impact of reconstructed pedigrees on progeny-test breeding values in red spruce. *Tree Genet Genomes.* 2010;6(4):591-600.
19. El-Kassaby YA, Cappa EP, Liewlaksaneeyanawin C, Klápště J, Lstibůrek M. Breeding without breeding: Is a complete pedigree necessary for efficient breeding?. *PLoS One.* 2011;6(10):e25737.
20. Klápště J, Suontama M, Telfer E, Graham N, Low C, Stovold T, et al. Exploration of genetic architecture through sib-ship reconstruction in advanced breeding population of *Eucalyptus nitens*. *PLoS One.* 2017;12:e0185137.
21. Vidal M, Plomion C, Harvengt L, Raffin A, Boury C, Bouffier L. Paternity recovery in two maritime pine polycross mating designs and consequences for breeding. **Tree Genet Genomes.** 2015;11:1-13.
22. Bouffier L, Klápště J, Suontama M, Dungey HS, Mullin TJ. Evaluation of forest tree breeding strategies based on partial pedigree reconstruction through simulations: *Pinus pinaster* and *Eucalyptus nitens* as case-studies. *Can J For Res.* 2019;49(12):1504-1515.
23. Bian L, Zheng R, Su S, Lin H, Xiao H., Wu HX, et al. Spatial analysis increases efficiency of progeny testing of Chinese fir. *J For Res (Harbin).* 2017;29(3):445-452.
24. Magnussen S. Application and comparison of spatial models in analyzing tree-genetics field trial. *Can J For Res.* 1990;20:536-546.
25. Kusnandar D, Galwey N. A proposed method for estimation of genetic plus tree group on forest trees without raising progeny. *Silvae Genet.* 2000;49:15-21.
26. Costa e Silva J, Dutkowski GW, Gilmour AR. Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. *Can J For Res.* 2001;31:1887-1893.
27. Chen Z, Helmersson A, Westin J, Karlsson B., Wu HX. Efficiency of using spatial analysis for Norway spruce progeny tests in Sweden. *Ann For Sci.* 2018;75:2
28. Dutkowski GW, Costa e Silva J, Gilmour AR, Lopez GA. Spatial analysis methods for forest genetic trials. *Can J For Res.* 2006;32:2201-2214.
29. Ye TZ, Jayawickrama KJS. Efficiency of using spatial analysis in first-generation coastal Douglas-fir progeny tests in the US Pacific Northwest. *Tree Genet Genomes.* 2008;4:677-692.
30. Fukatsu E, Hiraoka Y, Kuramoto N, Yamata H, Takahashi M. Effectiveness of spatial analysis in *Cryptomeria japonica* D. Don (sugi) forward selection revealed by validation using progeny and clonal tests. *Ann For Sci.* 2018;75-96.
31. Nishimura M, Setoguchi H. Homogeneous genetic structure and variation in tree architecture of *Larix kaempferi* along altitudinal gradients on Mt. Fuji. *J Plant Res* 2011;124:253-263.
32. Chen X, Sun X, Dong L, Zhang S. Mating patterns and pollen dispersal in a Japanese larch (*Larix kaempferi*) clonal seed orchard: a case study. *Sci China Life Sci.* 2018;61(9):1011-1023.
33. Leung Y, Mei C-L, Zhang W-X. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environ Plan A.* 2000;32:9-32.
34. Nardin M, Musch B, Rousselle Y, Guérin V, Sanchez L, Rossi JP, et al. Genetic differentiation of European larch along an altitudinal gradient in the French Alps. *Ann For Sci.* 2015;72: 517-527.
35. Chaisurisri K, El-Kassaby YA. Genetic diversity in a seed production population vs. natural populations of Sitka spruce. *Biodivers Conserv* 1994;3(6):512-513.
36. Ruņģis D, Luguza S, Bāders E, Šķipars V, Janson Ā. Comparison of genetic diversity in naturally regenerated Norway spruce stands and seed orchard progeny trials. *Forests.* 2019;10:926.
37. Iwasaki H, Uchiyama K, Kimura M, Saito Y, Hakamata T, Ide Y. Impact of a tree improvement program on the genetic diversity of sugi (*Cryptomeria japonica* D Don) plantations. *For Ecol Manag.* 2019;448:446-473.
38. Oreshkova NV, Belokon MM, Jamiyansuren S. Genetic diversity, population structure, and differentiation of Siberian larch, Gmelin larch, and Cajander larch on SSR-marker data. *Plant Genet Res.* 2013;49:178-186.
39. Hansen OK. Mating patterns, genetic composition and diversity levels in two seed orchards with few clones—Impact on planting crop. *For Ecol Manag.* 2008;256:1167-1177.

40. Burczyk J, Sandurska E, Lewandowski A. Linking levels of background pollination with pollen dispersal kernels. *Forests*. 2019;10:1139.
41. Wagner S, Liepelt S, Gerber S, Petit RJ. Within-Range translocations and their consequences in European larch. *PLoS One*. 2015;10:e0127516.
42. Colas F, Perron M, Tousignant D, Parent C, Pelletier M, Lemay P. A novel approach for the operational production of hybrid larch seeds under northern climatic conditions. *For Chron*. 2008;84(1):95-104.
43. Isabel N, Holliday J, Aitken SN. Forest genomics: Advancing climate adaptation, forest health, productivity, and conservation. *Evol Appl*. 2019;13:3-10.
44. Bouffier L, Raffin A., Kremer, A. Evolution of genetic variation for selected traits in successive breeding populations of maritime pine. *Heredity (Edinb)*. 2008;101:156-165.
45. Jin Y, Ma Y, Wang S, Hu XG, Huang LS, Li Y, Wang XR, Mao JF. Genetic evaluation of the breeding population of a valuable reforestation conifer *Platycladus orientalis* (Cupressaceae). *Sci Rep*. 2016;6:34821.
46. Nechaeva YS, Julianov AA, Boronnikova SV, Prishnivskaya YV. Nucleotide polymorphisms of candidate genes of adaptive significance in the Ural populations of *Larix sibirica* Ledeb. *Russ J Genet*. 2017;53(5):587-595.
47. Eckert AJ, Wegrzy JL, Pande B, Jermstad KD, Lee JM, Liechty JD, et al. Multilocus patterns of nucleotide diversity and divergence reveal positive selection at candidate genes related to cold hardiness in coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*). *Genetics*. 2009;183:290-298.
48. Hansen OK, Lamour K, Whetten R, Xu J. Ad hoc breeding of a genetically depauperate landrace of noble fir (*Abies procera* Rehder) using SNP genotyping via high-throughput targeted sequencing. *Tree Genet Genomes*. 2020;16(5):63.
49. Roques S, Berrebi P, Ch?vre P, Rochard E, Acolas ML. Parentage assignment in the critically endangered European sturgeon (*Acipenser sturio*) based on a novel microsatellite multiplex assay: a valuable resource for restocking, monitoring and conservation programs. *Conservation Genet Resour*. 2016;8(3):313-322.
50. Jones AG, Small MC, Paczolt KA, Ratterman NL. A practical guide to methods of parentage analysis. *Mol Ecol Resour*. 2010;10:6-30.
51. Lindgren and Wei. Low-input tree breeding strategies. In Proceedings of the IUFRO Division 2 Joint Conference: Isik F (de) Low input breeding and conservation of forest genetic resources, Antalya, Turkey, 9-13 October 2006, p124-138. [http://www4.ncsu.edu/~fisik/IUFRO%20Antalya %20Conference-Preceedings.pdf](http://www4.ncsu.edu/~fisik/IUFRO%20Antalya%20Conference-Preceedings.pdf).
52. El-Kassaby YA, Lindgren D. Increasing the efficiency of breeding without breeding through phenotypic pre-selection in open pollinated progenies. In: Byram ED (ed) Proceeding of 29th STFIC/WFGA meeting. Texas, USA, 2008.
53. Wang XR., Torimaru T, Lindgren D, Fries A. Marker-based parentage analysis facilitates low input 'breeding without breeding' strategies for forest trees. *Tree Genet Genomes*. 2010;6: 227-235.
54. Mihai G, Teodosiu M. Genetic diversity and breeding of larch (*Larix decidua* Mill.) in Romania. *Ann For Res*. 2009;52:97-108.
55. Pâques LE, Millier F, Rozenber P. Selection perspectives for genetic improvement of wood stiffness in hybrid larch (*Larix × eurolepis* Henry). *Tree Genet Genomes*. 2010;6:83-92.
56. Lai M, Sun X, Chen D, Xie Y, Zhang S. Age-related trends in genetic plus tree group for *Larix kaempferi* and their implication for early selection. 2014. *BMC Genet*. 2014;15(Suppl 1):S10.
57. Pan Y, Li S, Wang C, Ma W, Xu G. Shao L, et al. Early evaluation of growth traits of *Larix kaempferi* clones. *J For Res (Harbin)*. 2018;29:1031-1039.
58. El-Kassaby YA, Funda T, Liewlaksaneeyanawin C. Increasing breeding without breeding (BwB) efficiency: Full- vs. partial-pedigree reconstruction in lodgepole Pine. *SOJ Genet Sci*. 2015;2(1):1-6.
59. Park S, Park H, Im J, Yoo C, Rhee J, Lee B, Kwon C. Delineation of high -resolution climate regions over the Korean Peninsula using machine learning approaches. *PLoS One*. 2019;14(10):e0223362.
60. Ryu K, Watanabe M, Koike T. Ecophysiological responses of the larch species in northern Japan to environmental changes as a basis for afforestation. *Ecol Eng*. 2009;5:99-106.
61. Park NC, Lee KS, Par MS, Shin HC, Jun KS, Jung SY. Relation of the physico-chemical properties of forest soil to site indices of *Larix leptolepis* stands. *J For Res*. 2008;97(6):589-596.
62. Mao QA, Watanabe M, Makoto K, Kita K, Koike T. High nitrogen deposition may enhance growth of a new hybrid larch F-1 growing at two phosphorus levels. *Ecol Eng*. 2014;10(1):1-8.
63. Liang M, Sugimoto A, Tei S, Bragin IV, Takano S, Morozumi T, et al. Importance of soil moisture and N availability to larch growth and distribution in the Arctic taiga-tundra boundary ecosystem, northeastern Siberia. *Polar Sci*. 2014;8:327-341.
64. Gradel A, Ganbaatar B, Nadaldorj O, Dovdondemberl B, Kusbach A. Climate-growth relationships and pointer year analysis of a Siberian larch (*Larix sibirica* Ledeb.) chronology from the Mongolian mountain forest steppe compared to white birch (*Betula platyphylla* Sukaczew). *For Ecosyst*. 2017;4:22.
65. Kharuk VI, Ranson KJ, Im ST, Petrov IA. Climate-induced larch growth response within the central Siberian permafrost zone. *Environ. Res. Lett*. 2015;10:125009.
66. Obojes N, Armin M, Newesely C, Tasser E, Oberhuber W, Mayr S, et al. Water stress limits transpiration and growth of European larch up to the lower subalpine belt in an inner-alpine dry valley. *New Phytol*. 2018;220:460-475.
67. Harfouche A, Meilan R, Altman A. Molecular and physiological responses to abiotic stress in forest trees and their relevance to tree improvement. *Tree Physiol*. 2014;34:1181-1198.
68. Aguirre-Salade CA, Aldéz-Lazalde JR, Sánchez-Díaz G, Miranda-Aragón L, Aguirre-Salado AI. *J Trop For Sci*. 2015;27(3):298-313.
69. Kang JT, Son YM, Yim JS, Jeon JH. Estimation of carbon stock and uptake for *Larix kaempferi* Lamb. *Clim Res*. 2016;7(4):499-506.

70. Lim JH, Sung JH, Chun JH, Shin MY. Effect of climate factors on tree-ring growth of *Larix leptolepis* distributed in Korea. J Korean For Soc. 2016;105(1):122-131.
71. Mamet SD, Brown CD, Trant, AJ. and Laroque CP. Shifting global *Larix* distributions: Northern expansion and southern retraction as species respond to changing climate. J Biogeogr. 2018;46:30-44.
72. National Institute of Forest Science (NIFoS). Empirical Yield Table. Research Data Collection No. 677. Seoul, Korea;2016b.
73. Isoda K, Watanabe A. Isolation and characterization of microsatellite loci from *Larix kaempferi*. Mol Ecol. 2006;6:664-666.
74. National Forest Seed and Variety Center (NFSV). Development of microsatellite DNA marker for study on forest variety species. Chungju, Korea. 2019.
75. Peakall R, Smouse PE. GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching research-an update. Bioinformatics. 2012;28:2537-2539.
76. Marshall TC, Slate J, Kruuk LE, Pemberton JM. Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol. 1998;7:639-655.
77. Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol. 2007;16:1099-1106.
78. Henderson CR. Estimation of variance and covariance components. Biometrics. 1953;9:226-252.
79. Wilson AJ, Réale D, Clements N, Morrissey MM, Postma E, Walling, CA, et al. An ecologist's guide to the animal model. J Anim Ecol. 2009;79:13-26.
80. Cullis B, Gogel B, Verbyla A, Thompson R. Spatial analysis of multi-environment early generation variety trials. Biometrics. 1998;54(1):1-18.
81. Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R. ASReml user guide release 4.1 functional specification. VSN International Ltd, Hemel Hempstead, HP1 1ES. 2015. UK www.vsn.co.uk.
82. R core team. R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. 2020. URL <https://www.R-project.org/>.
83. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analysis in R. Bioinformatics. 2019;35:526-528.
84. Pinheiro J, Bates D, DebRoy S, Sarkar D, R core team. nlme: Linear and nonlinear mixed effects model. R package version 3.1-142, 2019. URL <https://CRAN.R-project.org/package=nlme>.
85. Munoz F, Sanchez L. breedR: Statistical methods for forest genetic resources analysts. R package version 0.12-5. 2020. <http://github.com/famuvie/breedR>
86. Barton K. MuMIn: Multi-model inference. R package version 1.43.17. 2020. URL <https://CRAN.R-project.org/package=MumIn>.

## Figures

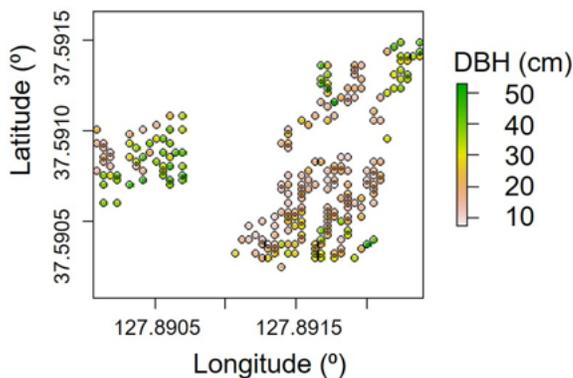


Figure 1

Rasterization of DBH distribution for the spatial analysis using AR1⊗AR1 in *Larix kaempferi* stand

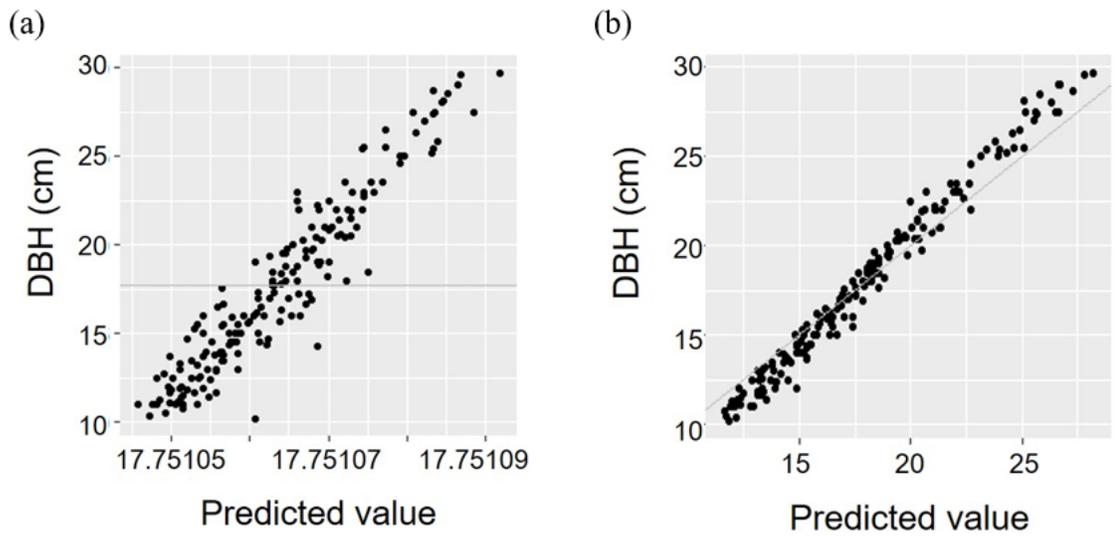


Figure 2  
 Fitness plot based on (a) animal model and (b) animal+AR1⊗AR1 model analysis

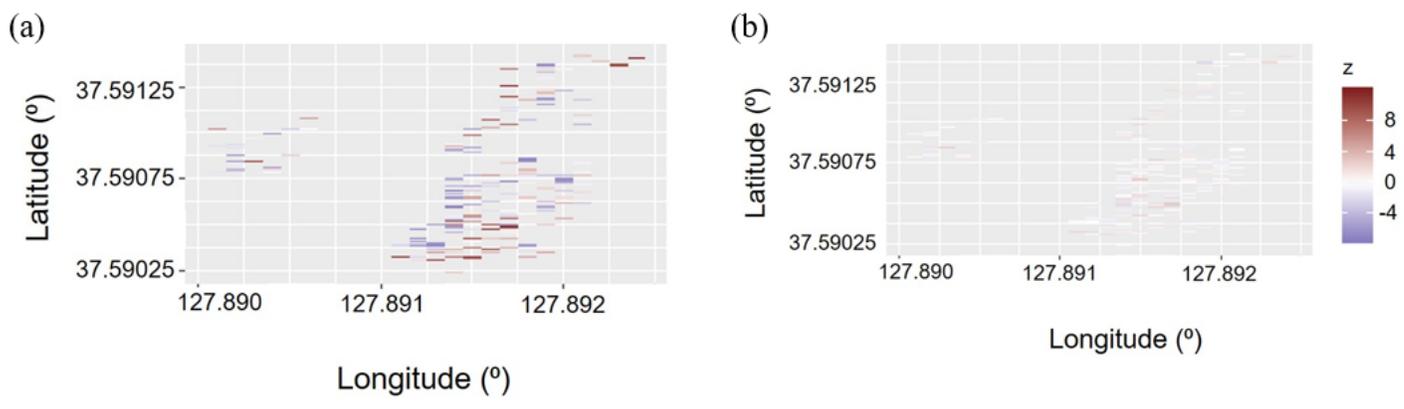


Figure 3  
 Comparison of (a) residual of animal model and (b) residual of animal+AR1⊗AR1 model

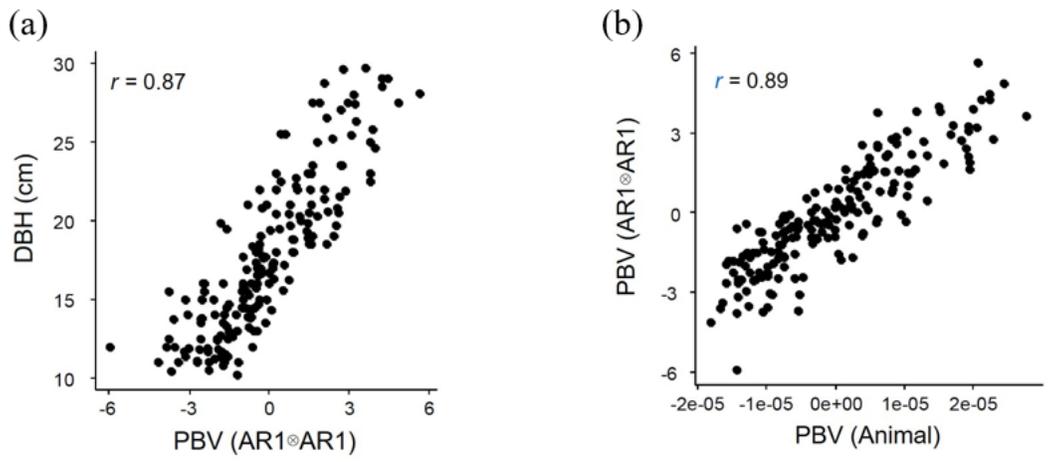


Figure 4  
 Correlation between (a) DBH and the predicted breeding value (PBV) based on animal+AR1⊗AR1 model, (b) predicted breeding values based on animal model and animal+AR1⊗AR1 model

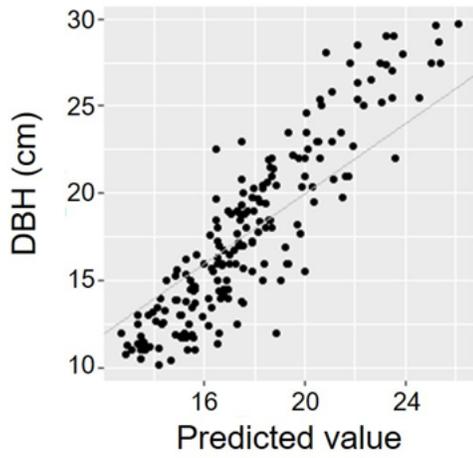
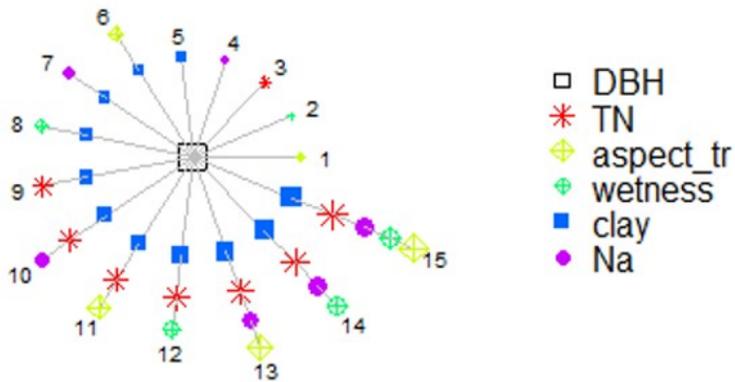


Figure 5

Fitness plot by animal+ AR1⊗AR1 model analysis with breeding values of mother tree

(a)



(b)

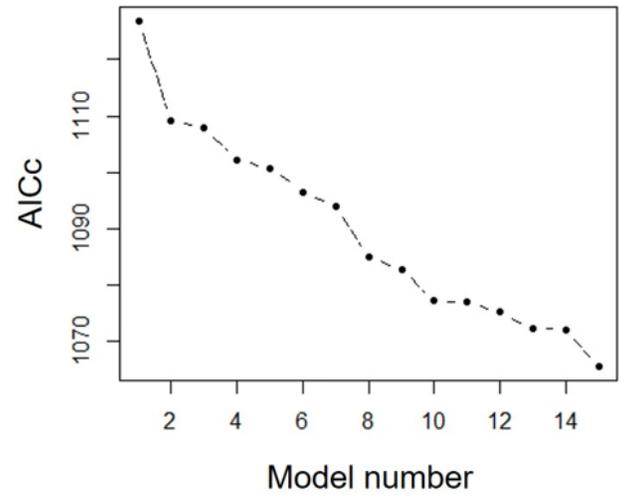
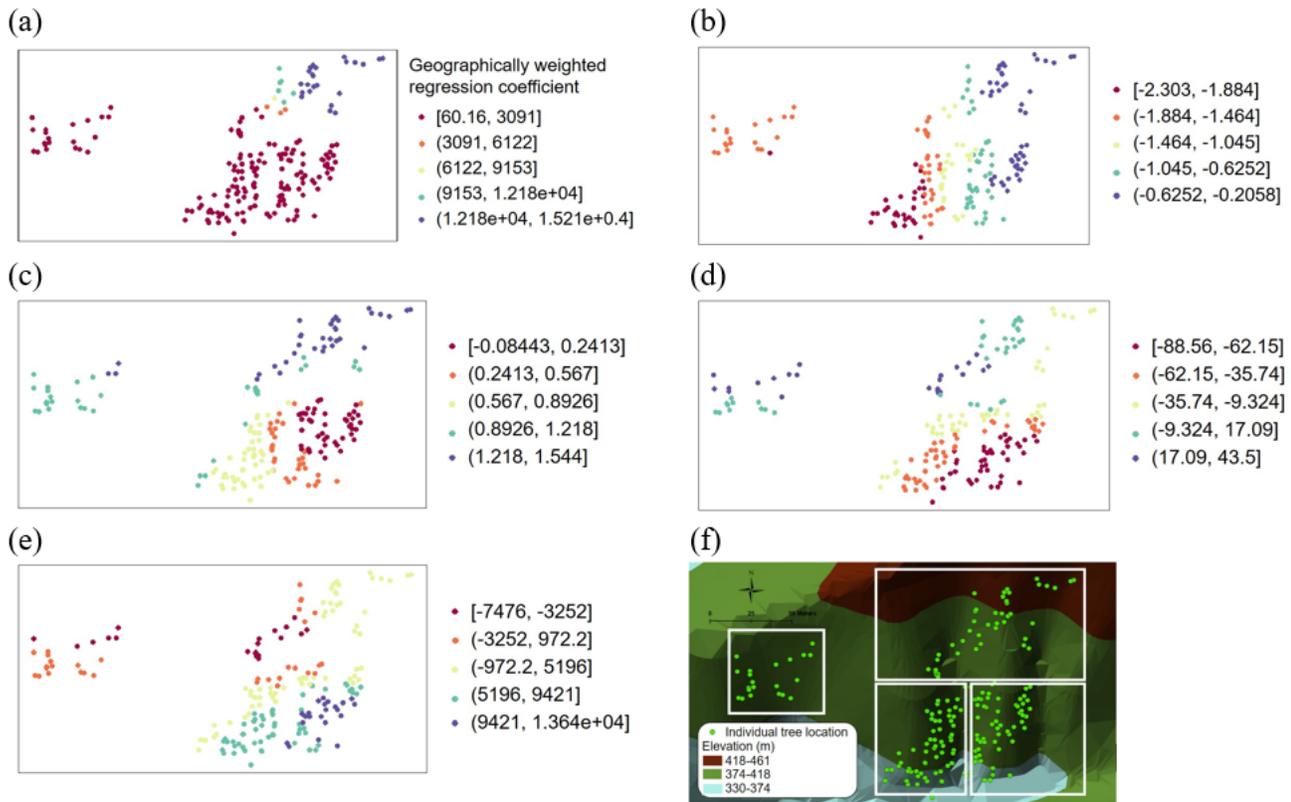
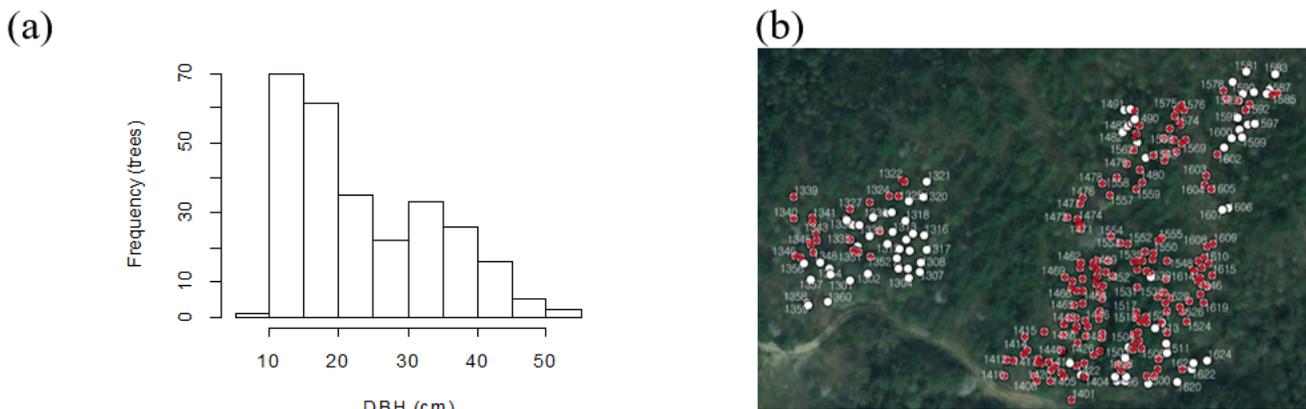


Figure 6

The procedure of model selection by geographically weighed regression model; AICc represents the second-order Akaike information criterion



**Figure 7**  
 Changes of the estimates of geographically weighted regression (GWR) coefficient of environmental factor with DBH of *Larix kaempferi* (a) total N, (b) aspect, (c) wetness, (d) clay, (e) Na+, (f) Site delineation inferred from the changes of GWR coefficients



**Figure 8**  
 (a) Histogram of the diameter at breast height (DBH) of *Larix kaempferi* individuals in the studied stand (b) The 188 individual trees selected for analysis (red marks) and the excluded trees (white) in the *Larix kaempferi* plantation

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [3.Supplementslarch0830BMC.docx](#)