

Highly informative SSR genotyping reveals large genetic diversity and limited differentiation in European larch (*Larix decidua*) populations from Romania

Pietro GRAMAZIO^{1,*}, Ioana M. PLESA^{2,**}, Alina M. TRUTA³, Adriana F. SESTRAS³, Santiago VILANOVA¹,
Mariola PLAZAS⁴, Oscar VICENTE⁴, Monica BOSCAIU⁵, Jaime PROHENS^{1,**}, Radu E. SESTRAS²

¹Institute for Conservation and Improvement of Valencian Agrodiversity, Universitat Politècnica de València, Valencia, Spain

²Department of Horticulture and Landscaping, Faculty of Horticulture,

University of Agricultural Sciences and Veterinary Medicine, Cluj-Napoca, Romania

³Department of Forestry, Faculty of Horticulture, University of Agricultural Sciences and Veterinary Medicine, Cluj-Napoca, Romania

⁴Institute of Molecular and Cellular Plant Biology, (UPV-CSIC), Universitat Politècnica de València, Valencia, Spain

⁵Mediterranean Agroforestry Institute, Universitat Politècnica de València, Valencia, Spain

Received: 10.01.2018 • Accepted/Published Online: 02.03.2018 • Final Version: 29.05.2018

Abstract: European larch (*Larix decidua*) is naturally distributed at high elevations in Central Europe. Romanian populations of *L. decidua* are scattered in different areas of the country. In this study, we used 12 highly informative genomic simple sequence repeat (SSR) markers for genotyping seven populations from different areas of Romania. The SSR markers were highly variable, with up to 11 alleles per SSR locus and an average polymorphic information content of 0.713. High values of observed ($H_o = 0.542$) and expected ($H_e = 0.738$) heterozygosities were observed. Cluster analysis of populations did not group them according to geographical distance, but some clusters contained populations from a similar altitudinal range. The partition of genetic variation revealed that two-thirds of the genetic variation was found within individuals (due to high H_o levels), while only one-sixth of the total genetic variation corresponded to variation among populations. A population structure analysis identified four genetic clusters, and in most cases individuals from a single population were assigned to several genetic clusters. Multivariate principal coordinates analysis confirmed the population structure analysis. SSR markers are a powerful tool for evaluating diversity, relationships, and the genetic structure of Romanian *L. decidua* populations, which have high levels of diversity and low genetic structuration.

Key words: Analysis of molecular variance, heterozygosity, genetic diversity, *Larix*, SSR markers

1. Introduction

European larch (*Larix decidua* Mill.) is an important deciduous forest tree species from the family Pinaceae with high economic interest in its timber (Matras et al., 2008). Because of the high quality of its timber and its fast growth, several breeding programs have been established to develop improved materials for establishing cultivated commercial stands of European larch (Mihai and Teodosiu, 2009; Marchal et al., 2017). European larch is used as an ornamental species (Torchik et al., 2010) and, among coniferous species, it is considered one of the best in terms of coexisting with good understory pastures for domestic grazing animals (Mosca et al., 2012). European larch mostly grows in mountainous cold areas in Central Europe (Alps, Carpathians, Sudetes) (Matras et al., 2008)

and, because of this, it is particularly susceptible to climate change (Mosca et al., 2012; Nardin et al., 2015). Due to higher temperatures caused by climate change, European larch can be displaced from its present altitudinal range to higher elevations by other evergreen conifers that are better adapted to warmer temperatures, which in some regions may limit its occurrence and long-term population viability (Kharuk et al., 2007; Matras et al., 2008; Nardin et al., 2015). At the same time, European larch is colonizing new areas where glacier retreat and warmer temperatures allow the tree line to move up to higher elevations (Pluess, 2011).

In Romania, European larch is distributed in the Carpathian Mountains range, in particular in the northern part of the country as well as in some scattered areas in the

* These authors contributed equally to this work.

** Correspondence: jprohens@btc.upv.es

center and west of the country (Matras et al., 2008; Vilcan et al., 2017). European larch populations of Romania are geographically isolated from the other main natural distribution areas in the Alps and the Sudetes, are largely fragmented, and are situated at the southeastern edge of its natural distribution (Mihai and Teodosiu, 2009). This makes Romanian populations of European larch particularly sensitive to climate change (Hamrick, 2010). A similar concern exists about the effects of climate change on other Romanian forest trees situated on the edge of the distribution, like European beech (*Fagus sylvatica* L.) (Budeanu et al., 2016).

Knowledge of the genetic diversity and structure of the Romanian populations of European larch is important in order to devise strategies for their conservation and assess their potential for adaptation to climate change, as well as for selection and breeding (Young, 2000; Mihai and Teodosiu, 2009; Marchal et al., 2017; Vilcan et al., 2017). Not much information exists on the genetic structure and differentiation of European larch populations. In this respect, Maier (1992), using allozymes, found low genetic differentiation between populations from the Alps, Poland, the Czech Republic, and Slovakia. In a more recent study, Nardin et al. (2015), using simple sequence repeat (SSR) markers, found low genetic differentiation between altitudinal gradients in a continuous natural population of European larch from the Alps. Little information is available on the genetic diversity and structure of Romanian populations of European larch. In this respect, Mihai and Teodosiu (2009) used isozymes to characterize five Romanian populations of *L. decidua* and found levels of diversity similar to those reported in other studies from other areas using similar methodologies (Lewandowski et al., 2000). More recently, Vilcan et al. (2017), using random amplified polymorphic DNA (RAPD) markers, found reduced genetic differentiation among Romanian populations of *L. decidua*, although some provenances could be distinguished at the molecular level.

Genetic characterization of Romanian populations with highly informative molecular markers could provide relevant information on their genetic diversity and structure (Matras et al., 2008). SSR markers present several advantages over other markers, as they are codominant, abundant, scattered all over the genome, robust, highly polymorphic, and amenable to multiplexing; they also have high inter- and intralaboratory reproducibility (Kalia et al., 2011). Among SSRs, genomic SSRs are mostly associated with noncoding regions and are generally more polymorphic than EST-SSRs, which are derived from expressed regions of the genome (Kalia et al., 2011; Muñoz-Falcón et al., 2011). SSRs are widely used for evaluating genetic diversity and relationships in many forest tree species (González-Martínez et al., 2006), and being heterozygous they allow estimation of the real heterozygosity of each of the individuals sampled (Kalia et al., 2011). Genomic SSR markers have been developed and tested in *L. decidua* (Isoda and Watanabe, 2006; Wagner et al., 2012). These same markers have been used in other European larch genetic variation studies (Pluess, 2011; Nardin et al., 2015), confirming their utility.

In the present study, we used genomic SSRs developed by Isoda and Watanabe (2006) and Wagner et al. (2012) to characterize European larch individuals from seven Romanian populations from different areas. The information on the genetic diversity and structure and genetic relationships of these populations will be useful to devise measures for the present and future conservation of genetic diversity of Romanian *L. decidua* as well as for designing strategies for selection and breeding of this timber forest species.

2. Materials and methods

2.1. Plant material

Seven populations of *Larix decidua* from different parts of Romania were sampled for the present study (Table 1; Figure 1). Five of the populations are situated in the central

Table 1. Geographical data of the seven *Larix decidua* populations studied. Three individuals were sampled per population. Trees were aged between 85 and 125 years old.

Population	Provenance	Stand area (ha)	Latitude	Longitude	Altitude (m asl)
1	Anina	1.8	45°05'N	21°51'E	680–710
2	Latorita	7.8	45°23'N	23°50'E	1000–1200
3	Gura Humorului	9.6	47°33'N	25°53'E	520–570
4	Brasov Valea Cetatii	4.8	45°38'N	25°36'E	800
5	Sacele	5.0	45°37'N	25°41'E	870–940
6	Brasov Valea Popii	6.4	45°25'N	25°19'E	670–700
7	Sinaia	3.0	45°19'N	25°33'E	750



Figure 1. Map of Romania displaying the location of each of the seven populations of *Larix decidua* sampled.

part of Romania (Latorita, Brasov Valea Cetatii, Sacele, Brasov Valea Popii, and Sinaia), one in the western part of Romania (Anina), and one in the north of Romania (Gura Humorului). The populations had stand areas of between 1.8 ha (Anina) and 9.6 ha (Gura Humorului) and altitudinal ranges between 520–570 m (Gura Humorului) and 1000–1200 m (Latorita) (Table 1). Distal parts of side branches having young light green leaves were picked for DNA extraction from three adult trees (estimated age between 80 and 125 years) situated in different parts of the tree stand of each population.

2.2. DNA extraction

Total genomic DNA was isolated from young needle-like leaves according to the CTAB protocol (Doyle and Doyle, 1987) with slight modifications. The extracted DNA was dissolved in Milli-Q water and its concentration and quality (230/260 and 260/280 nm ratios) were initially checked with a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). For samples not having high-quality parameters, the extraction was repeated. Quality of these samples was confirmed in agarose gel at 0.8% and with a Qubit 2.0 Fluorometer (Thermo Fisher Scientific). DNA was diluted with Milli-Q water to a concentration of 30 ng/ μ L for PCR amplification of SSRs.

2.3. SSR genotyping

Twelve genomic SSR markers (Table 2) developed by Isoda and Watanabe (2006) and Wagner (2012) were used to

screen the *L. decidua* individuals. SSRs were organized in two multiplexes according to the expected allele range size (Wagner et al., 2012) (Table 2). The SSR amplification was performed by PCR in a total volume of 12 μ L including 7.21 μ L of water, 1.2 μ L of 1X PCR buffer, 0.6 μ L of $MgCl_2$ (50 mM), 0.24 μ L of dNTPs (10 mM), 0.3 μ L of reverse primer (10 μ M), 0.06 μ L of forward primer with M13 tail (10 μ M), 0.24 μ L of fluorochrome (FAM, VIC, NED, and PET; 10 μ M), 0.15 μ L of Taq DNA polymerase (5 U/ μ L), and 2 μ L of DNA template (30 ng/ μ L). The PCR program used was the following: 95 °C for 3 min for denaturation, 30 cycles of 30 s at 95 °C followed by 30 s at 60 °C and of 30 s at 72 °C, and finally 72 °C for 5 min for the last step of extension. The PCR products were diluted in formamide and sequenced by capillary electrophoresis through an ABI PRISM 3100-Avant sequencer (Thermo Fisher Scientific) using a 600 LIZ GeneScan size standard (Thermo Fisher Scientific). PCR fragments were analyzed using GeneScan software (Thermo Fisher Scientific) to obtain the electropherograms. Polymorphisms were analyzed with Genotyper DNA Fragment Analysis software (Thermo Fisher Scientific).

2.4. Data analysis

Software packages PowerMarker (Liu and Muse, 2005) and GenAEx 6.503 (Peakall and Smouse, 2012) were used for the molecular marker analysis of SSRs. For each SSR locus, the following parameters were calculated using the PowerMarker package: number of alleles (N_a), major allele

Table 2. SSR markers used in the present study, including the multiplex in which they were included, their repeat motif, dye, size range obtained in this study, and the original source of the SSR.

SSR locus	Motif	Dye	Size range (bp)	Source
Multiplex 1				
bcLK189	(AG) ₁₇ AT(AG) ₆	NED	150–172	Wagner et al. (2012)
bcLK211	(CT) ₁₆	FAM	188–222	Wagner et al. (2012)
bcLK228	(AG) ₁₈	VIC	177–215	Wagner et al. (2012)
bcLK253	(AG) ₁₇	NED	195–227	Wagner et al. (2012)
lardec001529 (Ld30)	(AC) ₁₈	VIC	114–132	Isoda and Watanabe (2006)
lardec022835 (Ld50)	(CA) ₁₈	PET	157–185	Isoda and Watanabe (2006)
Multiplex 2				
bcLK263	(TC) ₂₀	PET	195–227	Wagner et al. (2012)
lardec023929 (Ld42)	(TG) ₁₄	VIC	179–185	Isoda and Watanabe (2006)
lardec024823 (Ld45)	(CA) ₁₃	FAM	202–212	Isoda and Watanabe (2006)
lardec023228 (Ld56)	(AC) ₁₆	NED	231–243	Isoda and Watanabe (2006)
lardec022359 (Ld58)	(AC) ₁₅	FAM	139–171	Isoda and Watanabe (2006)
lardec025807 (Ld101)	(AC) ₁₂	NED	189–197	Isoda and Watanabe (2006)

frequency (f), number of effective alleles (N_e), number of genotypes (N_g), polymorphic information content (PIC) values calculated as

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i p_j^2$$

(where n is the total number of alleles detected, p_i is the frequency of the i th allele, and p_j is the frequency of the j th allele) (Botstein et al., 1980), observed heterozygosity (H_o) calculated as the number of heterozygous alleles/number of alleles, expected heterozygosity (H_e) calculated as $1 - \sum_{i=1}^n p_i^2$ (where p_i is the frequency of the i th allele) (Nei, 1972), and fixation index (F_{is}) calculated as $F_{is} = 1 - (H_o/H_e)$ (Wright, 1965). A consensus tree was obtained in PowerMarker with genetic similarity data, using Nei's genetic distance (Nei, 1972), to evaluate the relatedness among populations using the neighbor-joining method (Saitou and Nei, 1987). Branch support on the phenogram was tested by bootstrap analysis with 1000 replications using PowerMarker and visualized by PHYLIP version 3.67 software (Felsenstein, 2007). Analysis of molecular variance (AMOVA) was performed using GenAlEx 6.503 software (Peakall and Smouse, 2012) using the data of the 12 SSR loci. Total molecular variation was partitioned in the following sources of variation: among populations, among individuals, and within individuals. The genetic distance matrix (Nei, 1972) among the different accessions was calculated with the GenAlEx 6.5 software package. The relationship between pair-wise genetic distances and geographical distances among individuals was calculated using linear regression analysis. Population

structure was estimated using a model-based Bayesian structure implemented in the software STRUCTURE 2.3.4 (Pritchard, 2000). Twenty runs of STRUCTURE were performed by setting the number of clusters (K) from 1 to 7 (number of locations of the sampled accessions). Each run consisted of a length of burn-in period of 5000 steps followed by 50,000 Monte Carlo Markov chain (MCMC) replicates, assuming an admixture model and uncorrelated allele frequencies. No prior knowledge of the population of origin was introduced. The ΔK method (Evanno et al., 2005) was used to identify the most likely number of clusters (K) using STRUCTURE HARVESTER 0.6.94 software (Earl, 2012). Based on the estimation of the best K of the latter, an additional run of STRUCTURE was performed using the following parameters: K3 to K5, length of burn-in period of 500,000, number of MCMC reps after burn-in of 750,000, number of iterations of 20. The ultimate most likely K was found running STRUCTURE HARVESTER software again. Each individual was assigned to its corresponding group based on maximum membership probability, as indicated by Remington et al. (2001). Principal coordinates analysis (PCoA) to graphically represent genetic relationships among individuals was performed using GenAlEx 6.5 software.

3. Results

3.1. SSR characterization and diversity

No overlap in the ranges of marker sizes was detected within the two multiplexes for the SSR loci tagged with the same dye (Table 2), and therefore the 12 SSR markers

could be unambiguously scored. For the 12 SSR markers, the size ranges we obtained were included within the size ranges described by Wagner et al. (2012). These 12 SSR markers were amplified in all individuals, except in one individual for each of the markers bcLK26, Ld30, Ld50, and Ld101, and in two individuals for Ld58, resulting in a missing data rate of 2.38%. In this way, only six of the 21 individuals genotyped had one missing SSR marker, while 15 had no missing data.

All the SSR loci scored were polymorphic (Table 3), and all the individuals genotyped had a different genetic profile. The number of SSR alleles (N_a) detected ranged between four (for Ld42) and 11 (for bcLK253 and bcLK263). The major allele frequency (f) ranged between 0.214 (bcLK228) and 0.750 (Ld101), while the number of effective alleles (N_e) varied between 1.73 (Ld101) and 7.74 (bcLK253) and the number of genotypes (N_g) varied between six (Ld101) and 14 (bcLK253) (Table 3). The PIC for SSR loci ranged between 0.397 (Ld101) and 0.858 (bcLK253), with an average value of 0.713, and all loci but one (Ld101) had PIC values above 0.5. The observed and expected heterozygosities (H_o and H_e , respectively) had average values of $H_o = 0.542$ and $H_e = 0.738$, resulting in an average fixation index (F_{is}) of $F_{is} = 0.256$ (Table 3). Although wide ranges of variation were found among SSR loci for H_o (0.300 to 0.857) and H_e (0.421 to 0.871), the correlation between these parameters was reduced ($r = 0.327$) and nonsignificant ($P = 0.2989$). This resulted in a wide variation among SSR loci for F_{is} , with values from -0.194 (Ld56) to 0.621 (bcLK263) (Table 3).

3.2. Multivariate cluster analysis of populations

The multivariate neighbor-joining cluster analysis of the seven populations studied shows that population 3 (Gura Humorului) from the north of Romania and from the lowest altitude (520–570 m) is basal to the other populations from central Romania (Figure 2). Among these, two clusters are clearly differentiated, with a bootstrap value of 100%, one containing population 6 (Brasov Valea Popii) which is the one ranking second in a low altitudinal range (670–700 m), and the other one consisting of the remaining populations. Among these, bootstrap values above 50% are only detected for the node separating population 7 (Sinaia) from populations 1 (Anina) and 4 (Brasov Valea Cetatii), which in turn are also differentiated by a bootstrap value of 96% (Figure 2). These three latter populations are from altitudes of 800 m or lower, while the two other populations, i.e. populations 2 (Latorita) and 5 (Sacele), are from altitudes above 850 m. Interestingly, with the exception of population 3, no clear geographical pattern is observed in the clustering of populations (Figure 2).

3.3. Genetic structure of populations

AMOVA (Table 4) revealed that around one-sixth of the genetic variance observed is caused by the differences among populations, while another one-sixth is accounted for by genetic differences among individuals. Two-thirds of the genetic variation observed is accounted for by within-individual genetic variation (Table 4) due to heterozygosity. The lack of population structure is confirmed when the relationship between genetic

Table 3. Genetic diversity statistics of SSR markers in *Larix decidua* populations, including size range, number of alleles (N_a), major allele frequency (f), number of effective alleles (N_e), number of genotypes (N_g), polymorphic information content (PIC), observed heterozygosity (H_o), expected heterozygosity (H_e), and fixation index (F_{is}).

SSR locus	N_a	f	N_e	N_g	PIC	H_o	H_e	F_{is}
bcLK189	9	0.333	4.79	7	0.764	0.429	0.791	0.458
bcLK211	10	0.500	3.42	10	0.685	0.714	0.707	-0.010
bcLK228	10	0.214	7.06	12	0.843	0.667	0.858	0.223
bcLK253	11	0.238	7.74	14	0.858	0.857	0.871	0.016
bcLK263	11	0.375	4.79	11	0.769	0.300	0.791	0.621
Ld30	8	0.325	5.19	10	0.784	0.550	0.808	0.319
Ld42	4	0.500	2.67	7	0.558	0.381	0.626	0.391
Ld45	5	0.333	4.26	10	0.727	0.571	0.765	0.253
Ld50	6	0.425	3.54	9	0.675	0.300	0.718	0.582
Ld56	6	0.500	3.11	8	0.684	0.810	0.678	-0.194
Ld58	9	0.289	5.78	12	0.806	0.526	0.827	0.363
Ld101	6	0.750	1.73	6	0.397	0.400	0.421	0.050
Mean \pm SE	7.92 \pm 0.70	0.399 \pm 0.043	4.51 \pm 0.51	9.67 \pm 0.68	0.713 \pm 0.037	0.542 \pm 0.055	0.738 \pm 0.036	0.256 \pm 0.072

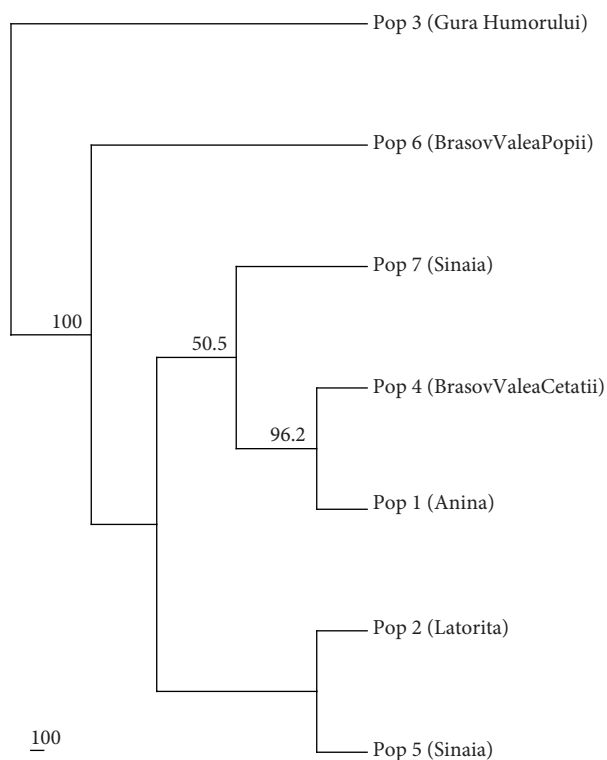


Figure 2. Neighbor-joining clustering phenogram based on Nei's (1972) genetic distances for seven populations of *Larix decidua*. Bootstrap values (in percentage; based on 1000 replications) greater than 50% are indicated at the corresponding nodes.

and geographical distance is studied (Figure 3). In this respect, no significant correlation exists between these distances ($r = -0.017$; $P = 0.7984$).

The ΔK statistic had a maximum peak at $K = 4$ (Figure 4), indicating the presence of four genetic clusters in the plants of the seven populations of *L. decidua* sampled. With the exception of individuals from populations 2 (Latorita) and 4 (Brasov Valea Cetatii), in which all the individuals of each of these populations clearly belonged to the same genetic cluster (Figure 5), individuals of

each population were assigned to more than one genetic cluster.

In three individual plants, corresponding to two individuals from population 3 (Gura Humorului) and one from population 6 (Brasov Valea Popii), no clear ascription could be made to any of the genetic clusters and they were considered as admixed between two or three genetic clusters (Figure 5). A first genetic cluster (I) is composed of one individual of population 1, the three individuals of population 4, and one of population 7; a second cluster (II) is the largest and is made up of two individuals of population 1, the three individuals of population 2, one from population 5, and two from population 6; the third genetic cluster (III) is made up of just two individuals of population 5; and the fourth genetic cluster (IV) by two individuals of population 7 (Figure 5). Two individuals of population 3 correspond to an admixture of clusters II, III, and IV, while one individual of population 6 is an admixture of genetic clusters II and IV (Figure 5).

3.4. Multivariate PCoA analysis of individuals

The four genetic clusters are confirmed in the multivariate PCoA (Figure 6). In this analysis, the first and second principal coordinates account, respectively, for 20.6% and 12.3% of the SSR genetic variation. The first coordinate basically separates genetic cluster I, with positive values; genetic cluster II, which plots in the middle part of the graph; and genetic clusters III and IV, which have negative values for this first coordinate (Figure 6). These two latter clusters are separated in the third component, with cluster III having positive values and cluster IV negative ones for the second principal component. The admixed individuals plot intermediately between the clusters of which they are an admixture (Figure 6).

4. Discussion

The multiplexing strategy developed by Wagner et al. (2012) proved effective in genotyping Romanian *L. decidua* populations. In this way, as occurred in other works using these same markers in *L. decidua* (Pluess, 2011; Nardin et al., 2015), no overlap in the range of

Table 4. Molecular analysis of variance (AMOVA) among populations, among individuals within populations, and within individuals (due to heterozygosity) for seven *Larix decidua* populations evaluated. Results are based on 12 SSR markers.

Source of variation	d.f.	S.S.	M.S.	Variance components	Percentage of variation
Total	41	188.83			
Among populations	6	56.33	9.389	0.785	16.63
Among individuals	14	65.50	4.679	0.744	15.77
Within individuals	21	67.00	3.190	3.190	67.60

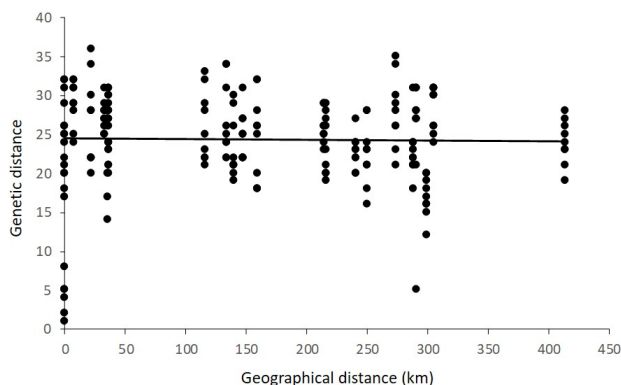


Figure 3. Relationships between geographical distance and Nei's (1972) genetic distances based on 12 SSR markers of 21 *Larix decidua* individuals belonging to seven populations (three individuals per population).

marker sizes was observed for SSR markers labeled with the same dye, allowing an efficient use of multiplexing. The size ranges obtained for individual SSRs matched the ones obtained by Wagner et al. (2012), although the range of variation was narrower in our case, probably due to the fact that these authors sampled a broader geographical area (several countries vs. Romania in our case) of *L. decidua* populations. The interpretation of SSR data is particularly sensitive to missing data (Reeves et al., 2016);

however, in our case the missing data rate for individual SSR markers was zero or very low and therefore did not justify removing any of the SSR markers from the analyses (Nardin et al., 2015; Reeves et al., 2016).

The diversity statistics obtained for the genomic SSR markers reveals that a large genetic variation exists in the materials scored. The number of alleles (N_a) per marker is somewhat lower than those obtained by Wagner et al. (2012) and Nardin et al. (2015), although those authors sampled more individuals. However, the number of effective alleles (N_e) does not differ significantly from the values obtained by Nardin et al. (2015). The fact that the number of genotypes (N_g) is generally different and larger than the number of alleles (N_a) is an indication of the presence of heterozygous loci (Guo and Thompson, 1982). PIC values above 0.5 are considered as high (Pandey et al., 2012), and in our case all SSR loci except one had values above this threshold, which indicates that a reduced number of highly informative genomic SSR loci, like the ones we used, provides relevant information for the reliable evaluation of genetic diversity and structure in *L. decidua* (Pluess, 2011; Wagner et al., 2012; Nardin et al., 2015). In this respect the PIC values obtained by us are higher than those obtained for genic SSRs in *L. kaempferi* (Chen et al., 2015) and similar to highly informative SSR markers in other conifer species, like those from the genus *Picea* (A'Hara and Cottrell, 2009; Shi et al., 2014).

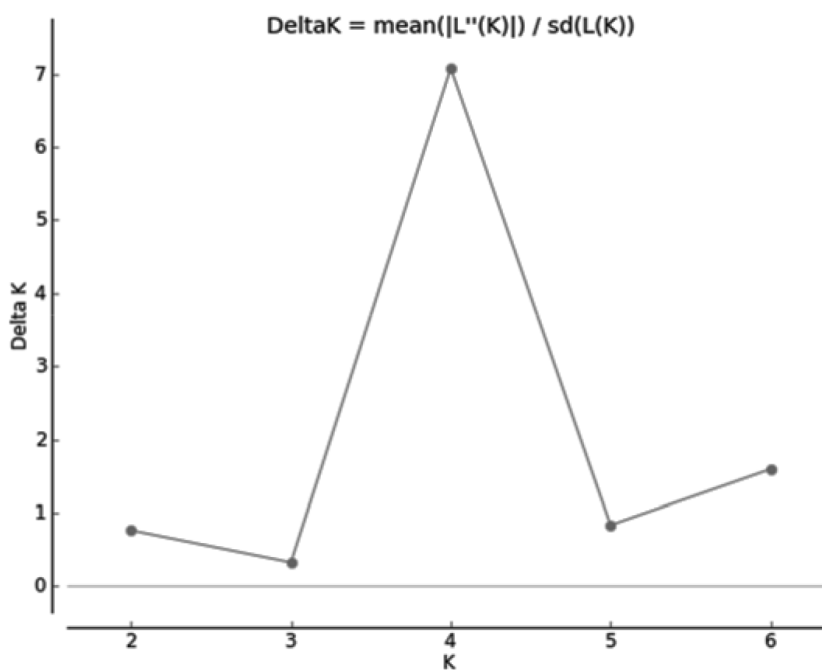


Figure 4. Delta K values for 2 to 6 genetic clusters for 21 individuals of *Larix decidua* collected from seven populations. Delta K was calculated according to Evanno et al. (2005).

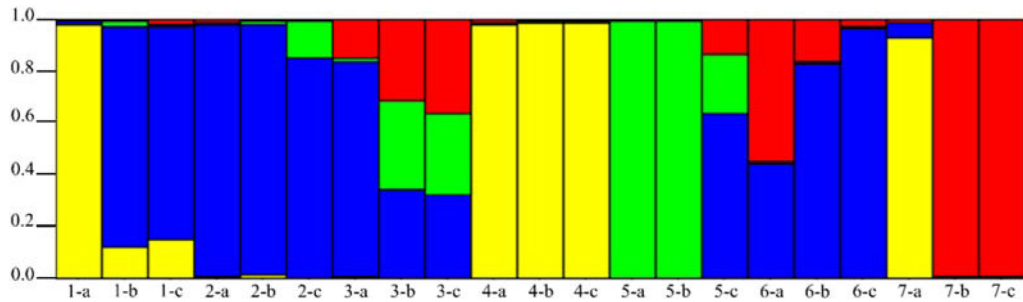


Figure 5. Estimated population structure for 21 plants of seven *Larix decidua* populations based on a number of clusters (K) $K = 4$. Each individual plant is represented by a vertical bar, which is partitioned into colored segments that provide an estimate of the membership fraction in each of the 4 genetic clusters. Each bar is marked with the population number (1–7) and individual (a–c).

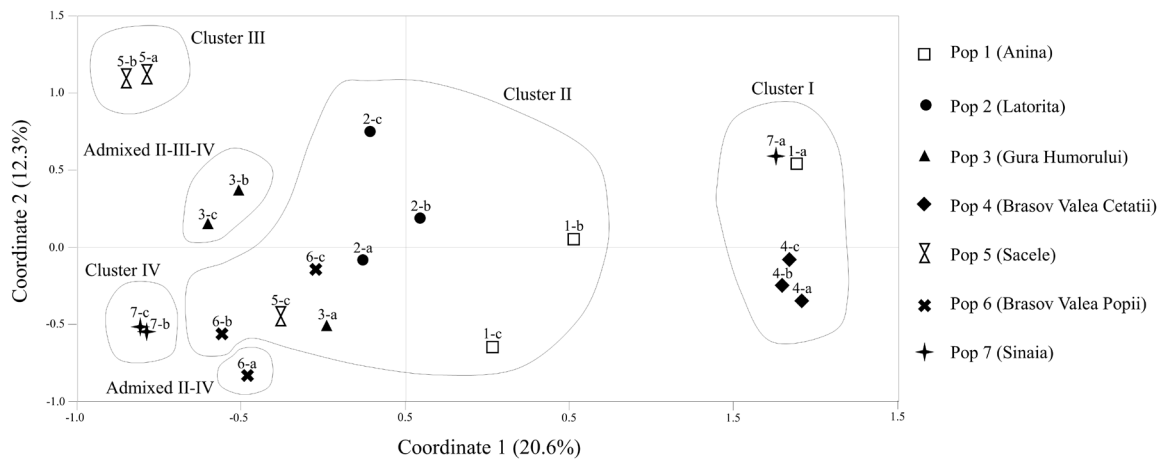


Figure 6. Principal coordinates analysis (PCoA) similarities based on 12 polymorphic SSR markers for *Larix decidua* individuals corresponding to seven populations based on the first (PCo1) and second (PCo2) principal coordinates. PCo1 and PCo2 together and PCo3 account for 20.6% and 12.3% of the total variation, respectively. Each point is marked with the population number (1–7) and individual (a–c). The four genetic clusters (I to IV) identified with the population structure analysis as well as the admixed individuals are indicated.

The high values for observed heterozygosity (H_o) were expected, given the allogamous reproduction of *L. decidua* (Lewandowski and Burczyk, 2000) and large sizes of the populations. Expected heterozygosity (H_e) values were also high, with an average value above 0.7, and similar to the values observed for this species (Pluess, 2011) and for the related *L. kaempferi* (Isoda and Watanabe, 2006), providing further evidence of the utility and highly informative nature of the SSRs used. As observed in *L. decidua* (Pluess, 2011; Wagner et al., 2012; Nardin et al., 2015) and in other conifer species (Mosca et al., 2012; Leonarduzzi et al., 2016), H_o and H_e values were quite similar, revealing a high degree of allogamy and low genetic isolation in the materials evaluated, as observed with RAPD markers in these same populations (Vilcan et al., 2017). As a result, the fixation index (F_{is}) values for the SSR markers were low and, as occurs in both *Larix* (Nardin et al., 2015; Vilcan et

al., 2017) and other conifers (Leonarduzzi et al., 2016), in some cases negative, revealing an excess of heterozygotes over those predicted by the Hardy–Weinberg equilibrium for these markers (Wright, 1965). Overall, because of their codominant nature, high degree of amplification, efficient multiplexing, robustness, reduced rate of missing data, high polymorphism, and high informativeness, the markers used here, which were developed by Isoda and Watanabe (2006) and Wagner et al. (2012), are of great utility for evaluating the diversity and genetic structure of *L. decidua* populations.

The cluster analysis of the seven populations basically grouped them according to altitudinal ranges rather than by geographical proximity. In a study of the genetic differentiation of *L. decidua* along four different altitudinal ranges in the French Alps, Nardin et al. (2015) found a low but significant genetic differentiation signal among them.

A similar result, this time with two altitudinal ranges, was obtained in *L. decidua* from the Swiss Alps (Zhao et al., 2000) and in *L. kaempferi* from Mount Fuji in Japan (Nishimura and Setoguchi, 2011).

The AMOVA results reveal a low level of genetic structure among the seven populations, even though they are situated at distances of up to 358 km. In fact, genetic variation among individuals within a population were similar to the differences among populations. These results are in agreement with those obtained for *L. decidua* populations from the Alps, Poland, the Czech Republic, and Slovakia (Maier, 1992) and a low level of genetic differentiation among plots and among individuals within plots in a population of European larch from the Alps (Nardin et al., 2015). Most of the genetic variation corresponded to within-individual variation, which is due to the presence of different alleles at individual loci (i.e. heterozygosity). This situation is common in forest trees, where high levels of heterozygosity are commonly found (González-Martínez et al., 2006; Carabeo et al., 2016; Tóth et al., 2017). In our case, the low levels of genetic structure have important implications for selection and breeding purposes, indicating that wide genetic diversity can be found in the offspring of a single individual, which would result from the high level of heterozygosity combined with the allogamous nature of *L. decidua* (Lewandowski and Burczyk, 2000). The low genetic structuration is confirmed by the lack of correlation between genetic and geographical distances, which was also observed with RAPD markers in these same populations (Vilcan et al., 2017). This provides an indication that, despite fragmented habitats, as occurs in other forest trees (Kremer et al., 2012), long distance extensive genetic flow takes place in *L. decidua*.

The number of genetic clusters is lower than that of populations and, except for two populations, individuals from the same population belong to different genetic clusters. This again provides support for extensive migration and gene flow in *L. decidua* (Lewandowski and Burczyk, 2000; González-Martínez et al., 2006).

References

- A'Hara SW, Cottrell JE (2009). Development of a set of highly polymorphic genomic microsatellites (gSSRs) in Sitka spruce (*Picea sitchensis* (Bong.) Carr.). *Mol Breeding* 23: 349-355.
- Botstein D, White RL, Skolnick M, Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32: 314-331.
- Budeanu M, Petritan AM, Popescu F, Vasile D, Tudose NC (2016). The resistance of European beech (*Fagus sylvatica*) from the eastern natural limit of species to climate change. *Not Bot Horti Agrobi* 44: 625-633.
- Carabeo M, Simeone MC, Cherubini M, Mattia C, Chiocchini F, Bertini L, Caruso C, La Mantia T, Villani F, Mattionia C (2016). Estimating the genetic diversity and structure of *Quercus trojana* Webb populations in Italy by SSRs: implications for management and conservation. *Can J Forest Res* 47: 331-339.
- Chen XB, Xie YH, Sun XM (2015). Development and characterization of polymorphic genic-SSR markers in *Larix kaempferi*. *Molecules* 20: 6060-6067.
- Doyle JJ, Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19: 11-15.
- One major genetic cluster is distributed across five out of the seven populations, suggesting that this genetic background may confer broad adaptation. On the other hand, the fact that two small genetic clusters (III and IV) are confined to some individuals of a single population may be the consequence of an adaptive advantage to specific environmental conditions (Hamrick, 2004). This would allow the maintenance of these genetic clusters despite extensive gene flow, migration, and hybridization.
- In conclusion, despite the limited sample size per population used, highly informative genomic SSR markers (Isoda and Watanabe, 2006; Wagner et al., 2012) have proved useful for evaluating the diversity and assessing the genetic structure of *L. decidua* populations from Romania. The results reveal that a large genetic diversity of highly heterozygous individuals exists in the European larch populations from Romania and that, despite significant geographic distances and altitudinal ranges in the population's samples, there is limited genetic differentiation. This has important implications for the conservation of European larch genetic diversity in a climate change scenario, as well as for selection and breeding of this important timber species (Matras and Pâques, 2008; Marchal et al., 2017; Vilcan et al., 2017). The high genetic diversity also may facilitate adaptation, through selection, of European larch to climate change.

Acknowledgments

Pietro Gramazio is grateful to Universitat Politècnica de València for a predoctoral contract (Programa FPI de la UPV-Subprograma 1/2013 call). Ioana M Plesa is grateful to the Doctoral School of the University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca and ICHAT - Institute of Advanced Horticulture Research of Transylvania, and Universitat Politècnica de València, for their support. Mariola Plazas is grateful to Spanish Ministerio de Economía, Industria y Competitividad for a post doctoral grant within the Juan de la Cierva programme (FCJI-2015-24835).

- Earl DA (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4: 359-361.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14: 2611-2620.
- Felsenstein J (2007). PHYLIP (Phylogeny Inference Package) Version 3.67. Seattle, WA, USA: Department of Genome Sciences, University of Washington.
- González-Martínez SC, Krutovsky KV, Neale DB (2006). Forest-tree population genomics and adaptive evolution. *New Phytol* 170: 227-238.
- Guo SW, Thompson EA (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361-372.
- Hamrick JL (2004). Response of forest trees to global environmental changes. *Forest Ecol Manag* 197: 323-335.
- Isoda K, Watanabe A (2006). Isolation and characterization of microsatellite loci from *Larix kaempferi*. *Mol Ecol Resour* 6: 664-666.
- Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177: 309-334.
- Kharuk V, Ranson K, Dvinskaya M (2007). Evidence of evergreen conifer invasion into larch dominated forests during recent decades in central Siberia. *Euras J Forest Res* 10: 163-171.
- Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle, JR, Gomulkiewicz R, Klein EK, Ritland K et al. (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecol Lett* 15: 378-392.
- Leonarduzzi C, Spanu I, Labriola M, González-Martínez SC, Piotti A, Vendramin GG (2016). Development and characterization of three highly informative EST-SSR multiplexes for *Pinus halepensis* Mill. and their transferability to other Mediterranean pines. *Plant Mol Biol Rep* 34: 993-1002.
- Lewandowski A, Burczyk J (2000). Mating system and genetic diversity in natural populations of European larch (*Larix decidua*) and stone pine (*Pinus cembra*) located at higher elevations. *Silvae Genet* 49: 158-160.
- Liu K, Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128-2129.
- Maier J (1992). Genetic variation in European larch (*Larix decidua* Mill.). *Ann Sci Forest* 49: 39-47.
- Marchal A, Muñoz F, Millier F, Sánchez L, Pâques LE (2017). Hybrid larch heterosis: for which traits and under which genetic control? *Tree Genet Genomes* 13: 92.
- Martin MA, Mattioni C, Cherubini M, Turchini D, Villani F (2010). Genetic diversity in European chestnut populations by means of genomic and genic microsatellite markers. *Tree Genet Genomes* 6: 735-744.
- Matras J, Pâques L (2008). EUFORGEN Technical Guidelines for Genetic Conservation and Use for European Larch (*Larix decidua*). Rome, Italy: Bioversity International.
- Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez J, Buckler E, Doebley J (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *P Natl Acad Sci USA* 99: 6080-6084.
- Mihai G, Teodosiu M (2009). Genetic diversity and breeding of larch (*Larix decidua* Mill.) in Romania. *Ann For Res* 52: 97-108.
- Mosca E, Eckert AJ, Di Piero EA, Rocchini D, La Porta N, Belletti P, Neale DB (2012). The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Mol Ecol* 21: 5530-5545.
- Muñoz-Falcón JE, Vilanova S, Plazas M, Prohens J (2011). Diversity, relationships, and genetic fingerprinting of the Listada de Gandía eggplant landrace using genomic SSRs and EST-SSRs. *Sci Hortic* 129: 238-246.
- Nardin M, Musch B, Rousselle Y, Guérin V, Sanchez L, Rossi JP, Gerber S, Marin S, Pâques L, Rozenberg P (2015). Genetic differentiation of European larch along an altitudinal gradient in the French Alps. *Ann For Sci* 72: 517-527.
- Nei M (1972). Genetic distance between populations. *Am Nat* 106: 283-292.
- Nishimura M, Setoguchi H (2011). Homogeneous genetic structure and variation in tree architecture of *Larix kaempferi* along altitudinal gradients on Mt. Fuji. *J Plant Res* 124: 253-263.
- Pandey MK, Gautami B, Jayakumar T, Sriswathi M, Upadhyaya HD, Gowda MVC, Radhakrishnan T, Bertioli D, Knapp SJ, Cook DR et al. (2012). Highly informative genic and genomic SSR markers to facilitate molecular breeding in cultivated groundnut (*Arachis hypogaea*). *Plant Breeding* 131: 139-147.
- Peakall ROD, Smouse PE (2012). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research an update. *Bioinformatics* 28: 2537-2539.
- Pluess AR (2011). Pursuing glacier retreat: genetic structure of a rapidly expanding *Larix decidua* population. *Mol Ecol* 20: 473-485.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
- Reeves PA, Bowker CL, Fettig CE, Tembrock LR, Richards CM (2016). Effect of error and missing data on population structure inference using microsatellite data. Preprint. bioRxiv: 080630. Available online at <https://www.biorxiv.org/content/early/2016/10/13/080630>.
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Goodman MM, Buckler ES (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *P Natl Acad Sci USA* 98: 11479-11484.
- Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406-425.

- Shi YZ, Forneris N, Rajora OP (2014). Highly informative single-copy nuclear microsatellite DNA markers developed using an AFLP-SSR approach in black spruce (*Picea mariana*) and red spruce (*P. rubens*). PLoS One 9: e103789.
- Tóth EG, Vendramin GG, Bagnoli F, Cseke K, Höhn M (2017). High genetic diversity and distinct origin of recently fragmented Scots pine (*Pinus sylvestris* L.) populations along the Carpathians and the Pannonian Basin. Tree Genet Genomes 13: 47.
- Torchik V (2010). Evaluation of an assortment of ornamental forms of conifers in central botanical garden of Belarus National Academy of Science. Acta Hort 885: 375-381.
- Vilcan A, Mihalte L, Sestras AF, Holonec L, Sestras RE (2017). Genetic variation and potential genetic resources of several Romanian larch populations. Turk J Agric For 41: 82-91.
- Wagner S, Gerber S, Petit RJ (2012). Two highly informative dinucleotide SSR multiplexes for the conifer *Larix decidua* (European larch). Mol Ecol Resour 12: 717-725.
- Wright S (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19: 395-420.
- Young AG, Boyle TJB (2000). Forest Conservation Genetics: Principles and Practice. Canberra, Australia: CSIRO Publishing.
- Zhao G, Felber F, Kupfer P (2000). Genetic variation and differentiation of *Larix decidua* populations in Swiss Alps. Acta Bot Sin 43: 731-735.