Genetic differentiation of *Quercus robur* in the South-Ural

Bernd Degen[1]*, Ruslan Yanbaev[2], Yulai Yanbaev[2, 3]

[1] Thünen Institute of Forest Genetics, Sieker Landstrasse 2, 22927 Grosshansdorf, Germany
[2] Bashkirian Agrarian University, Ufa, Russia
[3] Ufa Institute of Biology of UFIC RAS, Ufa, Russia

* Corresponding author: Dr. Bernd Degen, Email: bernd.degen@thuenen.de

Abstract

We studied the genetic composition of 200 pedunculate oak (*Quercus robur*) trees at nine nuclear microsatellite gene loci. We sampled nine locations in an area of 1100 km by 400 km in the South-Ural. The question was to analyse the genetic differentiation of the oaks at the south-east edge of the species distribution area. We observed relatively high values of genetic differentiation and fixation (*delta*=0.387, F_{ST}=0.0652, $F_{ST(Hedrick)}$=0.407) compared to values from the centre of the species distribution range. Bayesian clustering analysis revealed three genetic groups. Presence of all genetic groups was detected at all locations, but oak trees in the extreme east of the Ural Mountains were genetically most different. We hypothesise that genetic drift influenced the observed pattern.

Keywords: : *Marginal populations, microsatellites, pedunculate oak, genetic structure, Russia*

Introduction

The genetic differentiation of *Quercus robur* is well studied in most parts of its natural distribution area. Especially the application of chloroplast markers has shown the genetic footprints of the re-immigration of pedunculate oak in Western, Central and Northern Europe after the last glacial period (Neophytou and Michiels, 2013; Petit et al., 2002). But also the influence of seed transfer by humans has been demonstrated for central Europe (Konig et al., 2002). Another frequent topic of genetic studies was the level of hybridisation of pedunculate oak with other European white oak species (Curtu et al., 2015; Rellstab et al., 2016). The genetic differentiation in Eastern Europe is much less known (Ballian et al., 2010; Chmielewski et al., 2015) and only very few studies have been published until now on the genetic structure of *Quercus robur* in Russia (Chokheli et al., 2018; Gomory et al., 2001).

The aim of this study was to get novel insights into the genetic differentiation at highly variable nuclear SSR (nSSR) gene markers of pedunculate oak population at its south-eastern distribution edge in the South-Ural Mountains. Of particular interest were two questions: a) do we have evidence of different gene pools (refugia lineages) present in the Ural oak populations and b) did genetic drift combined with limited gene flow cause a higher genetic differentiation in the South-Ural region compared to other parts of the natural distribution range?

Materials and Methods

Sampling

We collected a total of 200 oak samples at nine locations in the South-Ural (table 1, figure 1). The sample size ranged from 11 to 40 individuals per sample point. In all locations only reproductive trees have been collected.

Table 1
Sampled material at the nine locations in the South-Ural (N total = 200)

Aber.	Name	Longitude	Latitude	N adults
AR	Archangelsk	56.963	54.530	40
DU	Dubovaya gora	55.680	56.557	14
KD	Kuvandik	57.566	51.303	38
MI	Michailovka	55.900	54.800	11
SA	Sarashi	55.760	56.780	15
SH	Shulgan	57.047	53.050	15
ZI	Zilair	57.320	52.230	40
KU	Kuseevo	58.335	52.976	20
SI	Sibai	58.515	52.726	7

DOI:10.2478/sg-2019-0019
edited by the Thünen Institute of Forest Genetics

Laboratory Procedures

DNA was extracted from leaves of adult trees following a protocol by (Dumolin et al., 1995). Extracted DNA from each sample was quantified and diluted to 10 ng/µL. Two sets of multiplexed microsatellite primers were used to genotype individuals at nine nSSRs: QrZAG112, QrZAG96, QpZAG110, QrZAG11, QrZAG87, QrZAG7, QrZAG20, QrZAG5b, QrZAG65 (Lepais et al., 2006). We used the same PCR and genotyping protocol as described in Buschbom et al. (2011).

Genetic distance, genetic differentiation, population fixation

Wright's F_{ST} (Wright, 1978), the standardized $G_{ST(Hedrick)}$ (Hedrick, 2005) and delta (Gregorius, 1987) were calculated using the program GDA_NT (Degen unpublished) as measures of fixation and genetic differentiation among populations. Genetic distance (D_0) was computed to measure the genetic differentiation between pairs of populations (Gregorius, 1984). Numerical tests using Monte Carlo methods with 10,000 permutations shifting individual genotypes among the sampled locations were applied to estimate the significance of delta, F_{ST}, $F_{ST(Hedrick)}$ and D_0. We calculated the Spearman's rank coefficient between the geographic and spatial distances among the populations.

Cluster analysis

To visualize differences between populations, cluster analysis based on the pairwise gene pool distances D_0 between locations was performed using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) as implemented in the software PAST (Hammer et al., 2001).

Bayesian clustering analysis

We used a Bayesian clustering method implemented in the software STRUCTURE v.2.3.4 (Pritchard et al., 2000) to check the number of genetic groups in the nine sampled locations. For the analysis with STRUCTURE, the model assumption is that the involved loci are not linked and that they are in Hardy-Weinberg-Equilibrium. We set the length of burn-in and Markov chain Monte Carlo simulations to 15,000 and 100,000, respectively, and tested K values from 1 to 7 each 10 times. We used the admixture model with correlated allele frequencies. The optimal number of genetic clusters was estimated with the ΔK (Evanno et al., 2005). For each tested K-value, data from the ten STRUCTURE runs was permuted with CLUMPP v.1.1.2 (Jakobsson and Rosenberg, 2007) to obtain the final Q values for each individual. Results were finally analysed and graphically represented using the software CLUMPAK (Kopelman et al., 2015).

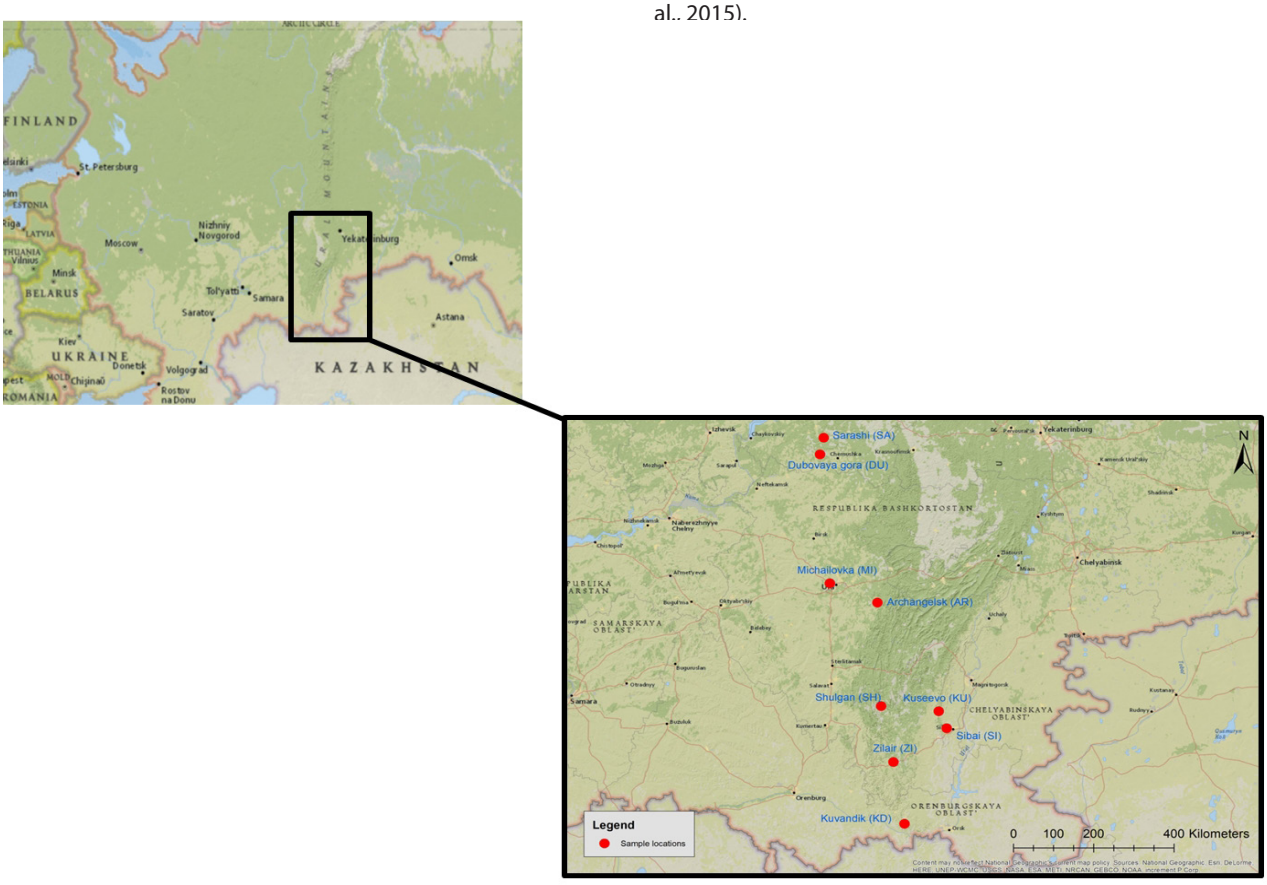

Figure 1
Position of the nine sampled populations in the South-Ural

Results

Genetic differentiation

We observed highly significant values (P=1) for the genetic differentiation (*delta*=0.396) of the gene pools among all nine populations as well as for the population fixation F_{ST}=0.067 and $F_{ST(Hedrick)}$=0.417. Also all pairwise genetic distances D_0 among populations were significant to highly significant with values ranking from 0.331 to 0.601 (table 2). There was only a very weak and non-significant correlation among genetic and spatial distances (r=0.063). Also, the cluster analysis did not indicate any spatial pattern (figure 2).

Table 2
Genepool distances D_G among the nine populations (below the diagonal) and proportion within 10000 permutations of alleles among populations with smaller genepool distances D_G than the observed ones (above the diagonal)

	AR	DU	KD	MI	SA	SH	ZI	KU	SI
AR		0.896	1.000	0.919	0.490	0.963	0.918	0.999	0.781
DU	0.436		1.000	0.999	0.992	0.999	0.996	1.000	1.000
KD	0.421	0.515		0.998	1.000	1.000	1.000	1.000	1.000
MI	0.479	0.548	0.500		1.000	1.000	0.946	0.999	0.999
SA	0.379	0.461	0.516	0.531		1.000	0.995	0.998	0.993
SH	0.438	0.503	0.547	0.533	0.481		1.000	1.000	1.000
ZI	0.331	0.441	0.409	0.451	0.428	0.454		0.976	0.725
KU	0.410	0.536	0.538	0.485	0.421	0.498	0.379		1.000
SI	0.524	0.601	0.592	0.572	0.540	0.582	0.522	0.566	

As indicated by the dendrogram of the cluster analysis the population of Sibai (SI) followed by Michailovka (MI) and Kuvandik (KD) are genetically the most differentiated ones (figure 2). On the other side there was a high genetic similarity between the populations of Archangelsk (AR) and Zilair (ZI).

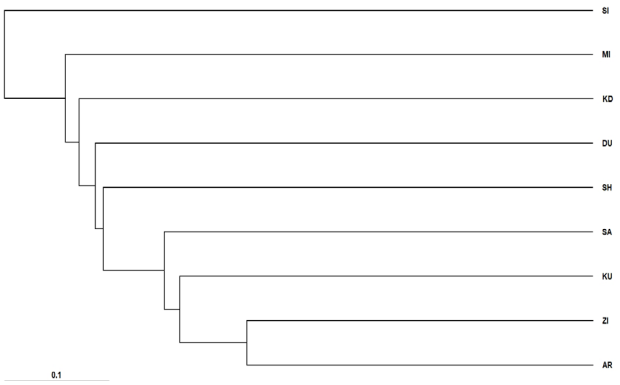

Figure 2
Dendrogram based on a cluster analysis (UPGMA: Unweighted Pair Group Method with Arithmetic Mean) using the genetic distance matrix among the allele frequencies of the nine oak populations.

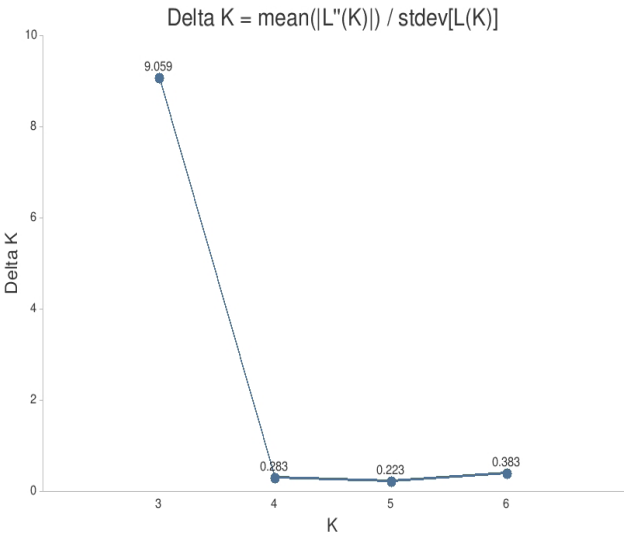

Figure 3
Delta k values for the tested genetic groups K=2 to 6 in program STRUCTURE

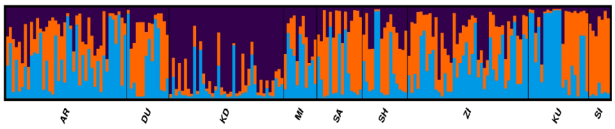

Figure 4
Mean membership coefficients of all 200 oak individuals for the majority mode of three (K = 3) among the 10 repeated STRUCTURE runs as computed with CLUMPAK (Kopelman et al. 2015) for all nine populations.

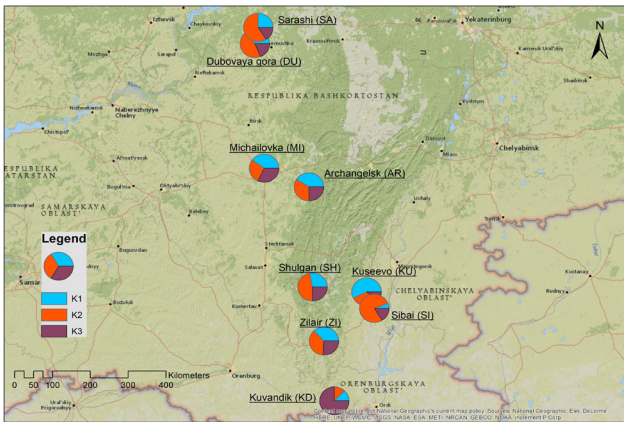

Figure 5
Spatial distribution of the mean membership coefficients for the majority mode of three (K = 3) among the 10 repeated STRUCTURE runs as computed with CLUMPAK (Kopelman et al. 2015) for all nine populations.

Bayesian cluster analysis

The results of the STRUCTURE analysis showed that based on delta K (Evanno et al., 2005), the best representation of the data is gained if three groups are considered (figure 3). These three genetic groups are present in all nine locations (figure 4, figure 5), but the frequency of the groups varied especially for the populations in Sibai (dominance of group K2), Kuseevo (dominance of group K1) and Kuvandik (dominance of group K3). At the six other locations the three genetic clusters were present at more or less equal frequencies.

Discussion

We observed relatively high values of population differentiation (*delta*=0.396) and fixation (F_{ST}=0.0676 and $F_{ST(Hedrick)}$=0.417) for the nine studied oak populations at nSSR-loci in the South-Ural. In most studies of *Quercus robur* with nuclear microsatellites in the centre of the species distribution range much lower values have been observed. For example, Gregorius et al. (2007) found for six stands in North Germany a *delta* of 0.13, F_{ST}=0.016 and $F_{ST(Hedrick)}$=0.053. Neophytou et al. (2010) observed an F_{ST} of 0.039 in three populations from Germany, Bulgaria and Greece. Nevertheless, similar to our results a high F_{ST}-value of 0.12 was observed at the northern edge of the species area in Finland (Pohjanmies et al., 2016). Also for other broadleaved tree species such as *Tilia cordata* a stronger genetic differentiation at the edges of the species distribution range was observed (Logan et al., 2019). In none of our sampled location we found an excess of homozygotes compared to the Hardy-Weinberg-expectations (data not shown). Thus, differences in the level of inbreeding cannot explain the observed genetic differentiation.

All pairwise genetic distances D_0 were statistically significant and relatively high (0.331 to 0.601). The oak trees of three very isolated locations in the south-east edge (Kuseevo, Sibai and Kuvandik) were the genetically most differentiated ones. This was also indicated by cluster analysis using the D_0 values. The high differentiation could be explained by genetic drift and insufficient gene flow. However, we have analysed the amount of external gene flow for the Sibai population before, and found it to be as high as 35 % (Buschbom et al., 2011).

The STRUCTURE analysis identified three different gene pools present in all nine locations. Again the three most differentiated locations had the most different composition of STRUCTURE membership coefficients. This could be simply an effect of genetic drift or an indication of a stronger footprint of glacial refugia. It is also possible that the south-east Ural served as a refugium itself. To disentangle these possibilities we would need data from more neighbouring populations. In order to resolve the large scale genetic structure of *Quercus robur* in the Russian distribution area a new project started in 2019. In this project, samples were collected in 95 locations (figure 6). They will be genotyped for a large set of SNP markers.

We can exclude for *Quercus robur* in the studied area – in contrast to other tree species in Russia such as Pine or Larch – different levels of hybridisation as a source of genetic differentiation (Petrova et al., 2018). The distribution range of *Q. petraea* as a potential hybridisation partner ends more than one thousand km to the west of the studied area. The presented work highlights several interesting research questions and it will be exciting to elucidate some of them with the help of the future large scale genetic data.

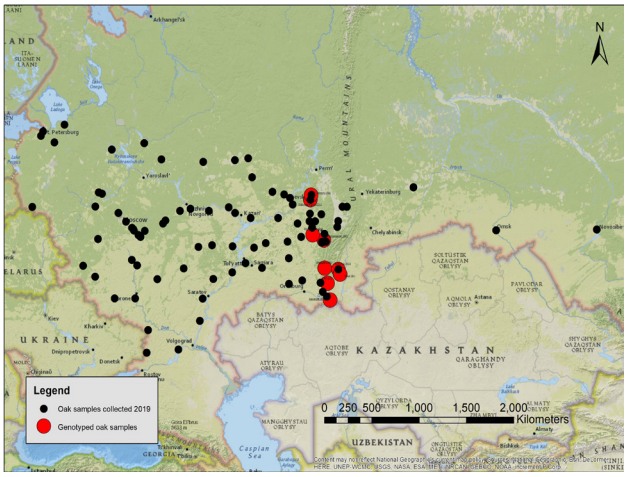

Figure 6
Spatial distribution of sampled locations of genotyped oak samples (red cycles) and newly collected material (black cycles).

Acknowledgements

We would like to thank J. Buschbom and I. Schulze for supervising and conducting the molecular lab work for this study and the reviewers of the manuscript for their helpful comments and suggestions that improved the study greatly. The new collection of oak samples was done in the frame of the grant No 19-16-00084 from the Russian Science Foundation.

References

Ballian D, Belletti P, Ferrazzini D, Bogunic F, Kajba D (2010) Genetic variability of Pedunculate Oak (Quercus robur L.) in Bosnia and Herzegovina. Period. Biol. 112(3):353-362

Buschbom J, Yanbaev Y, Degen B (2011) Efficient Long-Distance Gene Flow into an Isolated Relict Oak Stand. Journal of Heredity 102(4):464-472 https://doi.org/10.1093/jhered/esr023

Chmielewski M, Meyza K, Chybicki IJ, Dzialuk A, Litkowiec M, Burczyk J (2015) Chloroplast microsatellites as a tool for phylogeographic studies: the case of white oaks in Poland. iForest 8:765-771. https://doi.org/10.3832/ifor1597-008

Chokheli V, Kagan D, Rajput V, Kozlovsky B, Sereda M, Shmaraeva A, Khibuhina T, Fedyaeva V, Shishlova Z, Dmitriev P, Varduny T, Kapralova O, Usatov A (2018) Genetic Variability in Cenopopulations of Pedunculate Oak (Quercus robur) in Rostov Region, Russia, with the Use of ISSR-Markers. Int. J. Agric. Biol. 20(11):2544-2548

Curtu AL, Craciunesc I, Enescu CM, Vidalis A, Sofletea N (2015) Fine-scale spatial genetic structure in a multi-oak-species (Quercus spp.) forest. iForest 8:324-332. https://doi.org/10.3832/ifor1150-007

Dumolin S, Demesure B, Petit RJ (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. Theor. Appl. Genet. 91(8):1253-1256 https://doi.org/10.1007/bf00220937

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14(8):2611-2620. https://doi.org/10.1111/j.1365-294x.2005.02553.x

Gomory D, Yakovlev I, Zhelev P, Jedinakova J, Paule L (2001) Genetic differentiation of oak populations within the Quercus robur/Quercus petraea complex in Central and Eastern Europe. Heredity 86:557-563 https://doi.org/10.1046/j.1365-2540.2001.00874.x

Gregorius H-R (1987) The relationship between the concepts of genetic diversity and differentiation. Theor. Appl. Genet. 74(3):397-401 https://doi.org/10.1007/bf00274724

Gregorius H-R, Degen B, Konig A (2007) Problems in the analysis of genetic differentiation among populations--A case study in Quercus robur. Silvae Genet. 56(3-4):190-199. https://doi.org/10.1515/sg-2007-0029

Gregorius HR (1984) A unique genetic distance. Biometrical Journal 26(1):13-18 https://doi.org/10.1002/bimj.4710260103

Hammer Ø, Harper D, Ryan P (2001) PAST: Paleontological Statistics Software Package for education and data analysis. Palaeontolia Electronica 4

Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59(8):1633-1638. https://doi.org/10.1111/j.0014-3820.2005.tb01814.x

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801-1806 https://doi.org/10.1093/bioinformatics/btm233

Konig AO, Ziegenhagen B, van Dam BC, Csaikl UM, Coart E, Degen B, Burg K, de Vries SMG, Petit RJ (2002) Chloroplast DNA variation of oaks in western Central Europe and genetic consequences of human influences. Forest Ecology and Management 156(1-3):147-166 https://doi.org/10.1016/s0378-1127(01)00640-5

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. 15(5):1179-1191 https://doi.org/10.1111/1755-0998.12387

Lepais O, Leger V, Gerber S (2006) Short note: High throughput microsatellite genotyping in oak species. Silvae Genet. 55(4-5):238-240 https://doi.org/10.1515/sg-2006-0031

Logan SA, Phuekvilai P, Sanderson R, Wolff K (2019) Reproductive and population genetic characteristics of leading-edge and central populations of two temperate forest tree species and implications for range expansion. Forest Ecology and Management 433:475-486 https://doi.org/10.1016/j.foreco.2018.11.024

Neophytou C, Aravanopoulos FA, Fink S, Dounavi A (2010) Detecting interspecific and geographic differentiation patterns in two interfertile oak species (Quercus petraea (Matt.) Liebl. and Q. robur L.) using small sets of microsatellite markers. Forest Ecology and Management 259(10):2026-2035 https://doi.org/10.1016/j.foreco.2010.02.013

Neophytou C, Michiels HG (2013) Upper Rhine Valley: A migration crossroads of middle European oaks. Forest Ecology and Management 304:89-98 https://doi.org/10.1016/j.foreco.2013.04.020

Petit RJ, Csaikl UM, Bordacs S, Burg K, Coart E, Cottrell J, van Dam B, Deans JD, Dumolin-Lapegue S, Fineschi S, Finkeldey R, Gillies A, Glaz I, Goicoechea PG, Jensen JS, Konig AO, Lowe AJ, Madsen SF, Matyas G, Munro RC, Olalde M, Pemonge MH, Popescu F, Slade D, Tabbener H, Taurchini D, de Vries SGM, Ziegenhagen B, Kremer A (2002) Chloroplast DNA variation in European white oaks - Phylogeography and patterns of diversity based on data from over 2600 populations. Forest Ecology and Management 156(1-3):5-26 https://doi.org/10.1016/s0378-1127(01)00645-4

Petrova EA, Zhuk EA, Popov AG, Bondar AA, Belokon MM, Goroshkevich SN, Vasilyeva GV (2018) Asymmetric introgression between Pinus sibirica and Pinus pumila in the Aldan plateau (Eastern Siberia). Silvae Genet. 67(1):66-71 https://doi.org/10.2478/sg-2018-0009

Pohjanmies T, Elshibli S, Pulkkinen P, Rusanen M, Vakkari P, Korpelainen H, Roslin T (2016) Fragmentation-related patterns of genetic differentiation in pedunculate oak (Quercus robur) at two hierarchical scales. Silva. Fenn. 50(2):15. https://doi.org/10.14214/sf.1510

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945-959

Rellstab C, Buhler A, Graf R, Folly C, Gugerli F (2016) Using joint multivariate analyses of leaf morphology and molecular-genetic markers for taxon identification in three hybridizing European white oak species (Quercus spp.). Annals of Forest Science 73(3):669-679 https://doi.org/10.1007/s13595-016-0552-7

Wright S (1978) Evolution and the genetics of populations: a treatise in four volumes: Vol. 4: variability within and among natural populations. Chicago: University of Chicago Press