Genetic comparison of planted and natural *Quercus robur* stands in Russia

Bernd Degen[1]*, Yulai Yanbaev[2], Celine Blanc-Jolivet[1], Ruslan Ianbaev[2], Svetlana Bakhtina[2], Malte Mader[1]

[1] Thünen Institute of Forest Genetics, Sieker Landstrasse 2, D-22927, Grosshansdorf, Germany
[2] Bashkir State Agrarian University, 50-letiya Oktyabrya str.-34, 450001, Ufa, Russia

* Corresponding author: Bernd Degen, E-mail: bernd.degen@thuenen.de

Abstract

Genetic diversity and the optimal genetic composition are essential for the adaptability and adaptation of tree populations. Artificial regeneration of stands might reduce the genetic diversity and increase family structures if the seeds were collected from a limited number of mother trees. We did a genetic inventory in 12 pedunculate oak stands in Russia using a set of 366 nuclear gene markers (361 SNPs, 5 Indels) in order to look for differences in the genetic composition among natural and artificial stands. Our results did not reveal any systematic differences among both types of stands. However, we found two extreme cases of limited genetic diversity and increased proportion of full-sibs and half-sibs in urban man-made stands. The implications for the forestry and gene conservation programs were discussed.

Keywords: : *Family structure*, *genetic diversity*, *genotyping by sequencing*, *pedunculate oak*, *regeneration*, *seeds*, *silviculture*, *SNP*

Introduction

Genetic diversity and the optimal genetic composition are elementary for the adaptability and adaptation of tree populations (Prunier et al. 2016, Kremer and Hipp 2020). Here adaptability is the potential to adapt and adaptation is a process. Several studies have shown that the way of stand creation has a very important impact on the genetic composition. The type of seed collection influences the genetic composition of future artificially regenerated stands (Hosius et al. 2006). This creates a genetic bottleneck if the seed source is composed of only a few parent trees (Blanc-Jolivet and Degen 2014). But there are also examples of artificially regenerated stands with high genetic diversity (Gauli et al. 2009). Moreover, naturally regenerated stands might suffer as well a lack of genetic diversity. This can be the case if the regeneration is based on a few individuals or if over time due to a lack of gene flow, increasingly stronger family structures get established (Jolivet et al. 2012, Paluch et al. 2019).

In this study we focus on *Quercus robur* in Russia. This is a very important tree species in the Western part of the country. Developments in agriculture and overexploitation caused a high decrease of the species' distribution area in Russia in the last centuries (Kozharinov and Borisov 2013). This decline was reinforced by extremely cold winters and insect attacks. Recent surveys show that the area of pedunculate oak in Russia was reduced by 23 % over the past 50 years (Tsaralunga 2015). Nevertheless, *Quercus robur* occupies still about 3.5 million hectares and is a key species of broadleaf forests ecosystems in Western-Russia (Anonymous 1999).

Since a few decades the main strategies of restoration of this species in Russia rely on artificial reforestation (Shutyaev 2000). There were three types of areas that have been used for artificial regeneration of the species in this region: a) agricultural land, b) forest area, and c) urban territories. There were different actors involved in the reforestation and there were different guidelines applied for plantations in these three types of areas. Based on the applied methods (Koldanov 1967) we assumed the strongest genetic shift by the reforestations in the urban and agricultural areas (unknown and foreign seed sources, very limited number of seed trees), whereas we expected due to traditional good practise in forest areas (use of local seed sources, seed collection from many stands and mother trees) only little impact on the reforestation (Anonymous 1982).

DOI:10.2478/sg-2021-0001
edited by the Thünen Institute of Forest Genetics

We intended to compare the genetic composition and genetic parameters of artificially regenerated stands. Our hypothesis was that artificial stands have lower genetic diversity, larger genetic distances to stands in the same region and more relatives because the collection of the seed material was done on a narrow basis (only seeds of a few seed trees). Further, we analysed if there are any genetic differences for the three types of area used for artificial regeneration.

Materials and Methods

Sampling

In a broader study we conducted a genetic survey in 90 locations well distributed over the natural distribution range of *Quercus robur* in Russia. Most sample plots were located in natural forests but also a few sample plots were located in the three area types of artificial stands (agricultural land, forest area, urban area). We selected 12 stands out of the 90 for the present work in a way that they form three geographic groups (figure 1, table 1). Each group included natural and artificial stands in a radius of up to 400 km. The artificial stands were indicated in the names by the extension "_A", for example "Tvr2_A". We expected on a large geographic scale a strong genetic differentiation among the stands. The grouping of our study stands in three regional geographic groups should help to see genetic differences that are caused by a different stand establishment and not by large-scale isolation by distance effects. At each location, ten trees with a minimum diameter at breast height (DBH) of 20 cm and distant by at least 50 m among each other, were sampled. Cambium or leaves were collected and the geographic position was recorded. The material was stored in dry paper until DNA-extraction in the lab.

DNA extraction and genotyping

The DNA was extracted according to Dumolin et al. (1995). For all samples, 366 polymorphic nuclear loci (361 SNPs, 5 INDELs) were analysed based on targeted genotyping by sequencing (Degen et al. 2020a). Sequencing was done by LGC Genomics GmbH with Illumina NextSeq using single primer enrichment technology.

Data analysis

For the trees of each location we computed the effective number of alleles (A_e), the observed and expected hetereozygosity (H_o, H_e) and the fixation index (F_{IS}). We tested the statistical significance of the F_{IS}-values, and thus the significant departure from Hardy-Weinberg-proportions, by 1000 permutations shifting the alleles among individuals at the same sample location. The genetic differences between pairs of locations were quantified using the genetic distance d_0 (Gregorius 1984). As a measure of genetic differentiation *delta* (Gregorius 1987) was calculated. This parameter represents a genetic distance of each stand to the genetic composition of all other stands, thus the complementary group. We applied numerical tests using 1000 permutations shifting individual genotypes among samples to estimate the statistical significance of d_o. With the software PAST 4.3 (Hammer et al. 2001), we did an UPGMA -cluster analysis based on the pairwise gene pool distances d_0. The cluster analysis should visualize genetic differences of different locations. As part of the cluster analysis, bootstrapping was done 1000 times. Here the allele frequencies of individual loci were subjected to resampling. The percentage of replicates where each node is supported is given on the dendrogram.

Further, we used the full-likelihood approach implemented in the software COLONY to identify the number of full- and half-sibs at each location (Jones and Wang 2010). Only sibships with a probability of 0.99 and higher were included into the analysis. We applied the program CERVUS to compute the combined non-exclusion probability for sib identity (Marshall et al. 1998).

Results

Genetic differences

The genetic distances d_0 among pairs of stands for the gene pools at all 366 loci varied between 0.086 for the pair Penza / ArcA_A and 0.15 for the pair Tvr1 / Tvr2_A (table 2). All genetic distances $d_0 > 0.09$ were statistically significant. Within each geographic group, genetic distances among artificial and natural stands were not systematically higher compared to pairs of natural stands. The UPGMA-dendrogram did not show any clustering according to the three geographic groups (figure 2). Only the two neighboured stands Arc1k and Arc2 were grouped together. The artificial stands Tvr2_A and Msk1_A as well as the small village stand Tvr1 were significant outliers in the dendrogram. The artificially regenerated stand Lipetsk_A was separated in the dendrogram from all the others, but the node showed only 88 % support. A similar picture is given by the values of the genetic differentiation of the gene pools *delta* (figure 3, table 3). *Delta* values above the average of 0.08 were found for Tvr2_A, Tvr1 and Msk1_A. The artificial stands ArcA_A, and Lipetsk_A had less than average genetic differentiation from the complementary genetic composition of all other stands.

Genetic variation

The average effective number of alleles (A_e) over all 366 loci was 1.431 (table 3). Two out of four artificial stands had lower values (Msk1_A, Tvr2_A). Remarkable was the lowest value of A_e found in the stand Tvr2_A (1.389). Except for the stand Msk1_A, all artificial stands had higher than average observed heterozygosities. None of the fixation-indices gave evidence for an excess of homozygotes. They were all slightly negative and all values smaller than $F_{IS} < -0.04$ were statistically significant. The stand Tvr2_A had a pronounced excess of heterozygotes compared to Hardy-Weinberg-proportions ($F_{IS} = -0.20$).

Full-sibs and half-sibs

The analysis with the software COLONY was repeated three times. In all runs, we got the same result. The combined

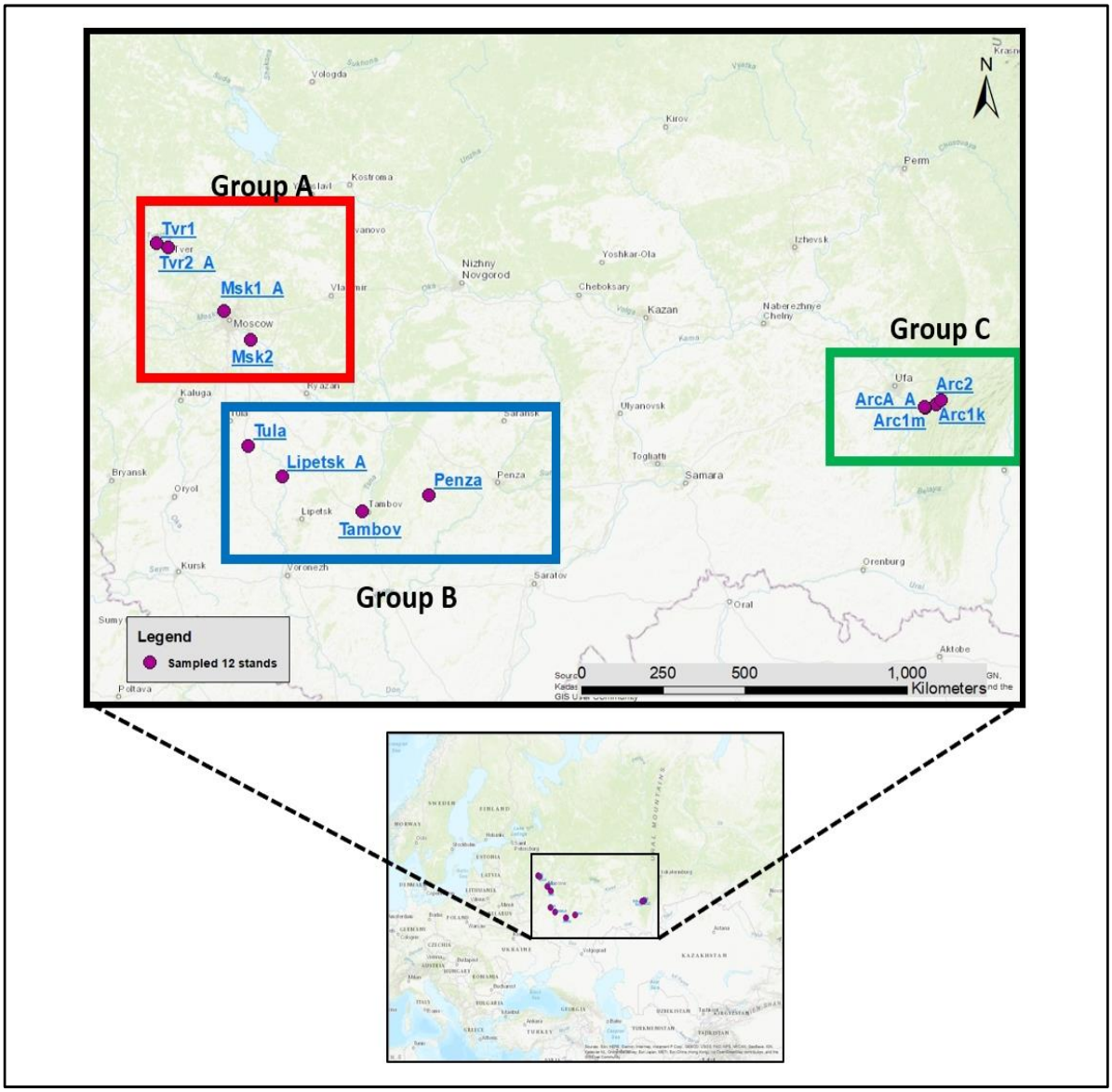

Figure 1
Distribution of the sampled stands

non-exclusion probability for sib identity was estimated by the software CERVUS at 1.316e^{-44}. Therefore, only full-sibs and half-sibs with a probability of at least 0.99 were counted. The strongest family structure with 6 full-sibs and 12 half-sibs was found in the samples of stand Tvr2_A (table 3). Note that the number of half-sibs can be larger than the sample size of 10 when more than two sampled individuals share the same parent. Also the samples from the small village population Tvr1 had a pronounced family structure (1 full-sib + 12 half-sibs). The two artificial stands Msk1_A and ArcA_A had no full- and half-sibs. The samples from the stand Tambov had 5 half-sibs.

Discussion

Genetic diversity

We did not systematically observe lower genetic diversity in artificially regenerated stands compared to the natural stands of the same region. In two out of the three groups an artificially regenerated stand had the lowest genetic diversity (A_e): Tvr2_A in group A and Lipetsk_A in group B. Also, no general trend was observed for the heterozygosity and the fixation-indices. Thus, artificial regeneration can be done in a way that has no negative impact. Or in a way that is at least not worse than natural regeneration. All depends on the applied practice. Harvesting of seeds from many spatially well distributed seed trees (> 30) in large populations catches the genetic

Table 1
Information on the sampled stands

Group	Name	latitude	longitude	type	Comment
A	Msk1_A	55.897778	37.457778	artificial	urban area, park in a city
A	Msk2	55.461389	38.193333	natural	
A	Tvr1	56.913611	35.592222	natural	urban area, small population in village
A	Tvr2_A	56.847222	35.916111	artificial	urban area, park in city, 45 years old, trees grow in lines
B	Lipetsk_A	53.301111	39.060833	artificial	agricultural area, plantation along the road
B	Penza	53.006389	43.101944	natural	
B	Tambov	52.734722	41.283611	natural	
B	Tula	53.799444	38.127778	natural	
C	Arc1k	54.517015	57.255834	natural	
C	Arc1m	54.449049	57.11635	natural	
C	Arc2	54.420473	56.798293	natural	large population
C	ArcA_A	54.3973	56.802758	artificial	forest area, stand in forest conditions

composition of the adult trees very well, as could also be shown by simulations (Blanc-Jolivet and Degen 2014) .

Naturally regenerated stands might also be subject to genetic drift and an increase of family structure (Degen et al. 2019). This is the case especially when the stands are isolated and only a limited number of seed trees reproduce. The negative effect can even be enforced by vegetative propagation. This has been observed for several tree species: e.g. *Abies alba* (Paluch et al. 2019), *Prunus avium* (Jolivet et al. 2012) or *Quercus robur* in Finland (Vakkari et al. 2020). Other studies support our observation that artificial regeneration does not necessarily mean a reduction of genetic diversity. For example, this has been observed for *Pinus sylvestris* in Sweden (Gil et al. 2015). And Aravanopoulos (2018) came in his review to the same conclusion and called for more sensitive genetic surveys to see an impact.

Genetic differentiation and large-scale genetic structure

We observed a mean genetic differentiation of *delta* = 0.08 among the samples from different locations. This value was smaller than results of another study (*delta*=0.13) on pedunculate oak in the same region (Degen et al. 2020b) but larger than the average value of *delta* = 0.06 for a study on nuclear SNPs of Mongolian oak stands in the Russian Far East (Schroeder et al. 2019). The dendrogram based on the genetic distances of our 12 studied locations did not show the expected substructure according to the three selected regions. We explain this with a rather small isolation by distance effect and thus with a relative effective gene flow among the oak populations (Buschbom et al. 2011).

Family structure

We would expect a higher proportion of half-sibs and full-sibs if the seeds for an artificial regeneration were collected from a limited number of mother trees. In three out of four artificial stands we observed no full-sibs in our samples. Remarkable is the stand Tvr2_A with 6 full-sibs and 12 pairs of half-sibs. This stand is a park in the city Tver. Obviously, the material has been collected from a very limited number of trees. Also, the small stand Tvr1 in a village had 12 pairs of half-sibs. Also, there was no evidence for planting this small stand seems to be genetically isolated. In most cases the presence of half-sibs and full-sibs at a location was linked with statistically significant negative F_{IS}-values. This underlined the differences to Hardy-Weinberg-proportions.

Plantations in urban, agricultural and forest areas

The sampled location Lipetsk_A represented a plantation in agricultural land. In the first half of the 20th century, 50 million hectares (= 60 %) of the arable agricultural area in the European part of the former Soviet Union were damaged by water and wind erosion. To counteract this erosion, almost 5 million hectares of forest were created in the steppe, forest-steppe zones, and along the main roads and railways by the middle of the last century (Koldanov 1967). Pedunculate oak was with 31.7 % one of the most important tree species for these new plantations. There are records that the acorns for these plantations were at least partly from Belarus and Western Ukraine (Koldanov 1967).The *delta* value for the samples from Lipetsk_A is not above average. This did not support the assumption of the use of a foreign seed source.

The sample of the location ArcA_A were typical for an artificial regeneration in forest land. Usually, seeds of local populations from many stands and trees were used for such plantations (Anonymous 1982). Our data supported the assumption that the practice very well transferred the local genetic diversity from one generation to the next one (no reduction A_e, below average *delta*, no full- and half-sibs).

Our samples from urban areas Msk1_A and Tvr2_A showed different genetic parameters. Both were in the dendrogram located in branches significantly different from all the others.

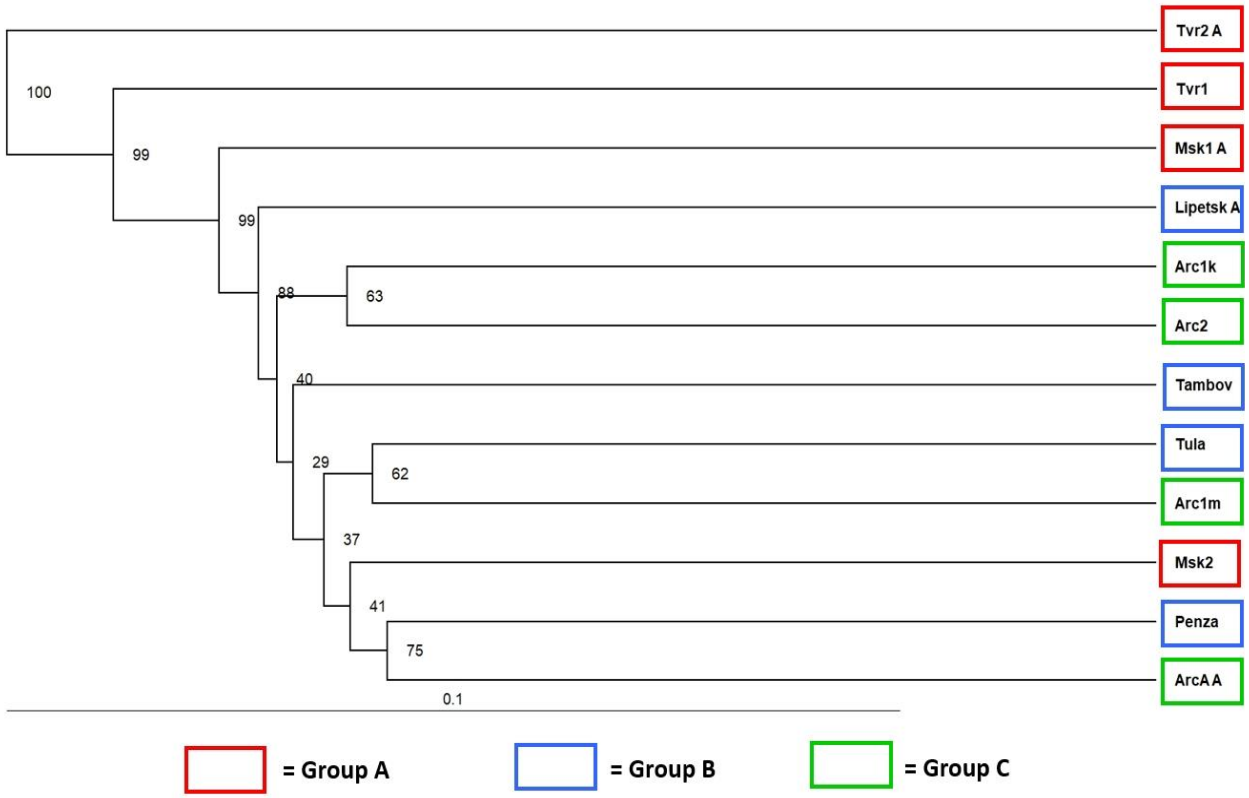

Figure 2
Dendrogram of a cluster analysis (UPGMA) based on the matrix of genetic distances of the gene pools for the 366 gene markers; the numbers represent the percentage of permutations that supported the node

Table 2
Genetic distances (d_0) among the gene pools at 366 loci; results of permutation tests: ***, **, * = statistically significant with p < 0.999, 0.99, 0.95; ns = not statistically significant, the colours mark genetic distance of pairs within each group (group A = red, group B= blue, group C= green)

	Msk2	Tvr1	Tvr2_A	Lipetsk_A	Penza	Tambov	Tula	Arc1k	Arc1m	Arc2	ArcA_A
Msk1_A	0.107***	0.121***	0.134***	0.110***	0.104***	0.106***	0.101***	0.100***	0.107***	0.108***	0.100**
Msk2		0.113***	0.130***	0.102***	0.089ns	0.096***	0.089ns	0.092***	0.098***	0.095***	0.091*
Tvr1			0.150***	0.116***	0.114***	0.123***	0.107***	0.117***	0.115***	0.121***	0.119***
Tvr2_A				0.125***	0.128***	0.130***	0.121***	0.130***	0.127***	0.123***	0.118***
Lipetsk_A					0.097**	0.102***	0.092*	0.100***	0.102***	0.104***	0.105***
Penza						0.098***	0.094**	0.093*	0.091*	0.102***	0.086ns
Tambov							0.095**	0.101***	0.098***	0.107***	0.095**
Tula								0.093***	0.088ns	0.103***	0.092*
Arc1k									0.100***	0.091*	0.092*
Arc1m										0.108***	0.094**
Arc2											0.095***

Table 3
Mean values per stand for the effective number of alleles (A_e), observed and expected heterorygosity (H_o, H_e), Fixation index (F_{IS}), sub-population differentiation *delta*, number of pairs of full-sibs (N FS) and number of pairs of half-sibs (N HS); results of permutation tests for F_{IS}: ***, **, * = statistically significant with p < 0.999, 0.99, 0.95, ns = not statistically significant

Group	Name	*Ae*	H_o	H_e	F_{is}	*delta*	N FS	N HS
A	Msk1_A	1.4210	0.2505	0.2491	-0.0148ns	0.0830	0	0
A	Msk2	1.4320	0.2597	0.2519	-0.0369ns	0.0717	1	3
A	Tvr1	1.4020	0.2548	0.2383	-0.0683***	0.1002	1	12
A	Tvr2_A	1.3890	0.2783	0.2261	-0.2006***	0.1139	6	12
B	Lipetsk_A	1.4440	0.2771	0.2595	-0.0578**	0.0780	0	2
B	Penza	1.4560	0.2704	0.2642	-0.0257ns	0.0703	0	0
B	Tambov	1.4510	0.2755	0.2604	-0.0538**	0.0789	0	5
B	Tula	1.4490	0.2652	0.2608	-0.0201ns	0.0674	0	0
C	Arc1k	1.4420	0.2617	0.2578	-0.0199ns	0.0732	0	1
C	Arc1m	1.4280	0.2544	0.2503	-0.0238ns	0.0749	0	1
C	Arc2	1.4200	0.2634	0.2453	-0.0627**	0.0786	0	2
C	ArcA_A	1.4380	0.2665	0.2554	-0.0389ns	0.0701	0	0
Mean		1.4310	0.2648	0.2516	-0.0519	0.0800		

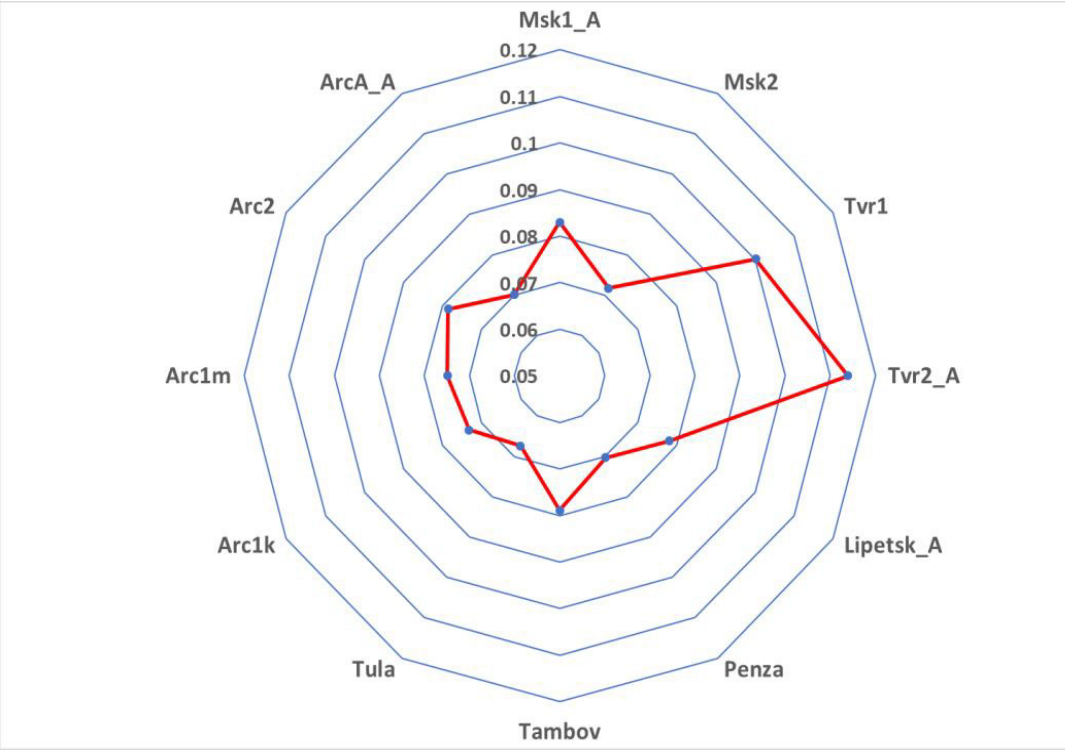

Figure 3
Genetic differentiation *delta* at the 12 locations

But whereas Tvr2_A had a high proportion of half- und full-sibs and lower genetic diversity (A_e), these negative effects could not be observed in the samples of Msk1_A. This stand is surrounded by many natural oak stands. Tvr2_A represented a park, and we got personal reports that the oak alley was created in 1975 using acorns of several old trees of the city of Tver, planted in the early 19th century.

Sample size

The small sample size of 10 individuals per stand could be criticised. But it should be noted that we got for each of these ten diploid individuals data for 366 loci. This provided per stand 10 x 366 x 2 = 7320 data points. A study with 100 individuals per stand and 10 microsatellite loci would have provided only 2000 data points. Of course, it is essential that the ten individuals were collected with a minimum distance of at least 50m in order to avoid an overrepresentation of individual families.

Conclusion

We see the urgent need for guidelines, information and training of involved persons to ensure genetic sustainable collection of seed material for plantation purpose in Russia. Most suitable would be if the lead on this would be taken by the Russian Federal Forestry Agency. Among its seven obligations of its subordinate organization "Roslesozashchita" there are three directly linked to our topic (monitoring of forest reproduction, monitoring of forest genetic resources, management of Federal fund for forest plant seeds). Further, a critical evaluation of future seed sources needs to be done.

Acknowledgement

The study was co-financed by grant WKF-WF04-22WC4111 01 (German Federal Ministry of Food and Agriculture & German Federal German Ministry for the Environment, Nature Conservation and Nuclear Safety) and for the study on the Russian pedunculate oak samples by the grant No 19-16-00084 from the Russian Science Foundation (project members: S. Bakhtina, R. Ianbaev, Y. Yanbaev and B. Degen). We are grateful to Stefan Jencsik, Petra Hoffmann and Susanne Hoppe for technical assistance with laboratory work.

References

Anonymous (1982) Forest seed zoning of the basic forest-forming species in the USSR (in Russian). Moscow: Forest Industry

Anonymous (1999) Forest Fund of Russia: guide (in Russian). Moscow Research and information center on forest resources

Aravanopoulos FA (2018) Do Silviculture and Forest Management Affect the Genetic Diversity and Structure of Long-Impacted Forest Tree Populations? Forests 9(6):14. https://dx.doi.org/10.3390/f9060355

Blanc-Jolivet C, Degen B (2014) Using simulations to optimize genetic diversity in Prunus avium seed harvests. Tree Genetics & Genomes 10(3):503-512. https://dx.doi.org/10.1007/s11295-014-0699-z

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://dx.doi.org/10.1093/jhered/esr023

Degen B, Blanc-Jolivet C, Bakhtina S, Ianbaev R, Yanbaev Y, Mader M, Nürnberg S, Schröder H (2020a) A new set of nuclear and plastid SNP and Indel loci for Quercus robur and Quercus petraea screened with targeted genotyping by sequencing Conservation Genetic Ressources:submitted

Degen B, Yanbaev R, Yanbaev Y (2019) Genetic differentiation of Quercus robur in the South-Ural. Silvae Genetica 68(1):111-115. https://dx.doi.org/10.2478/sg-2019-0019

Degen B, Yanbaev Y, Ianbaev R, Bakhtina S, Tagirova A (2020b) Genetic diversity and differentiation among populations of the pedunculate oak (Quercus robur) at the eastern margin of its range based on a new set of 95 SNP loci. Journal of Forestry Research:7. https://dx.doi.org/10.1007/s11676-020-01265-w

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

Gauli A, Gailing O, Stefenon VM, Finkeldey R (2009) Genetic similarity of natural populations and plantations of Pinus roxburghii Sarg. in Nepal. Annals of Forest Science 66(7):10. https://dx.doi.org/10.1051/forest/2009053

Gil MRG, Floran V, Ostlund L, Mullin TJ, Gull BA (2015) Genetic diversity and inbreeding in natural and managed populations of Scots pine. Tree Genetics & Genomes 11(2):12. https://dx.doi.org/10.1007/s11295-015-0850-5

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

Gregorius HR (1987) The relationship between the concepts of genetic diversity and differentiation. Theoretical and Applied Genetics 74(3):397-401. https://dx.doi.org/10.1007/bf00274724

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

Hosius B, Leinemann L, Konnert M, Bergmann F (2006) Genetic aspects of forestry in the central Europe. European Journal of Forest Research 125(4):407-417. https://dx.doi.org/10.1007/s10342-006-0136-4

Jolivet C, Holtken AM, Liesebach H, Steiner W, Degen B (2012) Mating patterns and pollen dispersal in four contrasting wild cherry populations (Prunus avium L.). European Journal of Forest Research 131(4):1055-1069. https://dx.doi.org/10.1007/s10342-011-0576-3

Jones OR, Wang J (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources 10(3):551-555. https://dx.doi.org/10.1111/j.1755-0998.2009.02787.x

Koldanov VJ (1967) Steppe afforestation (in Russian). Moscow: Forest Industry

Kozharinov A, Borisov P (2013) Distribution of oak forests in Eastern Europe over the last 13000 years. Contemporary Problems of Ecology 6(7):755-760 https://doi.org/10.1134/s199542551307007x

Kremer A, Hipp AL (2020) Oaks: an evolutionary success story. New Phytologist 226(4):987-1011. https://dx.doi.org/10.1111/nph.16274

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7(5):639-655. https://dx.doi.org/10.1046/j.1365-294x.1998.00374.x

Paluch J, Zarek M, Kempf M (2019) The effect of population density on gene flow between adult trees and the seedling bank in Abies alba Mill. European Journal of Forest Research 138(2):203-217. https://dx.doi.org/10.1007/s10342-019-01162-w

Prunier J, Verta JP, MacKay JJ (2016) Conifer genomics and adaptation: at the crossroads of genetic diversity and genome function. New Phytologist 209(1):44-62. https://dx.doi.org/10.1111/nph.13565

Schroeder H, Yanbaev Y, Kersten B, Degen B (2019) Short note: Development of a new set of SNP markers to measure genetic diversity and genetic differentiation of Mongolian oak (Quercus mongolica Fisch. ex Ledeb.) in the Far East of Russia. Silvae Genetica 68(1):85-91. https://dx.doi.org/10.2478/sg-2019-0016

Shutyaev AM (2000) Biodiversity of pedunculate oak and its using in breeding and afforestation Voronezh: Research Institute of forest genetics and breeding

Tsaralunga VV (2015) External signs of pathology of the pedunculate oak (in Russian). Voronezh: Voronezh State Technical University

Vakkari P, Rusanen M, Heikkinen J, Huotari T, Karkkainen K (2020) Patterns of genetic variation in leading-edge populations of Quercus robur: genetic patchiness due to family clusters. Tree Genetics & Genomes 16(5):12. https://dx.doi.org/10.1007/s11295-020-01465-9