Vladimír Mačejovský, Jarmila Schmidtová, Matúš Hrivnák,
Diana Krajmerová, Ivana Sarvašová, Dušan Gömöry*

Interspecific differentiation and gene exchange among the Slovak *Quercus* sect. *Quercus* populations

Abstract: European white oak species (*Quercus* sect. *Quercus*) are known to share a substantial part of their nuclear and cytoplasmic genomes as a result of extensive interspecific hybridization and introgression. We studied natural populations of three species distributed in Slovakia, namely *Q. robur*, *Q. petraea* and *Q. pubescens*, which are largely sympatric but have contrasting ecological requirements, using a combination of leaf morphometry, neutral nuclear microsatellite markers (nSSR) and potentially adaptive amplified fragment length polymorphisms (AFLP). Bayesian analysis of population structure relying on nSSR revealed that the degree of admixture was relatively low in *Q. robur* but higher in *Q. petraea* and *Q. pubescens*. The inferred *Q. robur* ancestry closely correlated with morphometric scores of the canonical discriminant analysis based on 13 leaf traits, while for the other two species the correlation was much weaker. We identified two AFLP fragments associated with climatic variables, mainly with air vapour pressure and characteristics of temperature regime at the sites of origin. These associations indicate that climatic adaptation is the mechanism driving interspecific divergence and maintaining integrity of white oak species.

Addresses: V. Mačejovský, M. Hrivnák, D. Krajmerová, D. Gömöry, Faculty of Forestry,
Technical University in Zvolen, T.G. Masaryka 24, SK-960 01 Zvolen, Slovakia, e-mail: gomory@tuzvo.sk
J. Schmidtová, Faculty of Wood Technology, Technical University in Zvolen, T.G. Masaryka 24,
SK-960 01 Zvolen, Slovakia
I. Sarvašová, Borová hora Arboretum, Technical University in Zvolen, Borovianska cesta 2171/66,
SK-960 01 Zvolen, Slovakia
*corresponding author

Introduction

In spite of a wide variety of species concepts, defining species by various criteria such as phenotypic distinction, cohesion or monophyly, the biological species concept emphasizing reproductive isolation (Mayr, 1963) has become the predominating view of species among biologists, especially zoologists.

Under this viewpoint, species are seen as the fundamental level in the hierarchy of biological diversity, as discrete units with boundaries, which are well defined by the ability/disability of producing viable and fertile offspring with members/non-members of the same species, respectively. In contrast to the other taxonomical ranks, the species level was considered 'real', i.e., defined by truly biological, non-arbitrary

criteria. Consequently, supporters of the biological species concept consider hybridization as a breakdown of isolating mechanisms, and generally as a lapsus of nature, occurring incidentally, rarely, and having generally detrimental effects on fitness of hybrid offspring. Nevertheless, evidence has gradually accumulated in the course of the last decades that interspecific hybridization is a rather common phenomenon (Arnold, 2006; Rieseberg & Carney, 1998). The view of hybridization has progressively shifted towards recognizing this process as a creative force in plant evolution, at least among botanists.

Hybridization is not always accompanied by introgression, because it requires that F_1 hybrids are viable and fertile, while the fitness of backcrossed individuals is not too inferior compared to parental species. This is by far not a general rule; the crossability of species naturally depends on the degree of divergence and similarity of genomic structures between parental species. However, equivalence in fitness of hybrids or even hybrid vigor are not exceptional (Schwarzbach et al., 2001; Schweitzer et al., 2002); alternatively, the variation range in fitness for hybrids and backcrosses may be broad, and at least some offspring genotypes may outperform their parents (Arnold et al., 2001; Arnold & Hodges, 1995; Rieseberg & Carney, 1998). Evolutionary consequences of hybridization and introgression may vary from hybrid speciation, when the hybrid lineage evolves into a separate taxon (Gross & Rieseberg, 2005; Mallet, 2007), to genetic swamping, when a rare species is merged with a more common one after a contact is established (Rhymer & Simberloff, 1996).

In the case of genetically proximate species, sympatry and parapatry are unavoidably associated with the "risk" of genetic exchange and, ultimately, merging into a single taxon. Factors warranting species integrity may range from purely extrinsic (e.g., flowering asynchrony) to purely intrinsic (chromosomal rearrangements, different ploidy). In many cases, genetic cohesion is maintained by divergent selection, when species are adapted to different environments (Gross & Rieseberg, 2005). In that case, species may share a substantial part of their genomes (either shared ancestral polymorphisms or gained de novo through interspecific gene flow) and differentiation is retained only in the segments of the genome under selection.

The genus *Quercus* is a good example of interspecific hybridization and introgression. Interspecific gene exchange is quite common within most subgenera and sections of this genus (Rushton, 1993; Cannon & Petit, 2019), and even intersectional hybrids were reported (Vazquez et al., 1993). Chloroplast and nuclear genes are known to be largely shared among European white oak species (sect. *Quercus*), which may have resulted from interspecific hybridization (Petit et al., 2003; Gömöry & Schmidtová, 2007; Lexer et al., 2006). The local patterns of gene admixture are, of course, not identical across the whole range. They depend primarily on the species-history factors (especially the pathways of Holocene spread), the resulting horizontal and vertical overlaps of distribution ranges, and environmental pressures. On the other hand, hybridization rates between white oak species are known to be non-identical in both directions (Bacilieri et al., 1996), and there is a number of both prezygotic and postzygotic reproductive barriers, ranging from decelerated heterospecific pollen tube growth to lower seed germination rates or poorer height growth of hybrid progenies, which contribute to reproductive isolation (Abadie et al., 2012; Lepais et al., 2013).

In the case of white oaks, hybridization and introgression have of course relevance for forestry practice as well. *Q. robur* and *Q. petraea* oak belong to commercially important tree species, and oak acorns have been extensively traded and transferred across Europe since the advent of modern forestry (Tulstrup, 1959). Not long ago oak seeds and seedlings were identified only on the genus level in Slovakia when marketed. Consequently, oak stands (especially *Q. robur*) were planted often outside their natural habitats, which established secondary contacts between species and potentially promoted hybridization. Only the recently adopted legislation on forest reproductive materials (law no. 217/2004) restricted (but not completely terminated) this practice. Knowledge of the extent of past hybridization (both natural and human-induced) and mechanisms maintaining species integrity is thus indispensable for the conservation of genetic resources of white oaks.

Therefore, the primary motivation for this study was assessing the genetic consequences of these past practices, to allow judging whether the current practices are *lege artis*. The aim of this study is to estimate the extent of allele sharing among three white oak species in Slovakia, where their distribution ranges largely overlap, and oaks have frequently been planted outside their natural ranges. Moreover, we used potentially adaptive AFLP markers to identify environmental drivers of species divergence within this taxonomical complex.

Material and Methods

Experimental material

Nineteen populations of pedunculate (*Quercus robur* L.), sessile (*Q. petraea* (Matt.) Liebl.) and pubescent (*Q. pubescens* Willd.) oaks were selected to cover the distribution ranges of these species in Slovakia

(Table 1). As we also attempted to identify potential adaptive patterns, we restricted the sampled area to the Slovak Western Carpathians and adjacent parts of the Pannonian lowland to avoid variation associated with the Holocene recolonization history. Although this area was colonized by three different refugial lineages, their descendants are spatially intermixed (Petit et al., 2003); therefore, no geographical trends or patterns resulting from different origins are expected. The populations are indigenous (mostly natural or near-natural forests). Species identification relied on leaves, bark, tree form and (whenever available) cupules and fruits. The choice of populations relied on the data of the state nature conservation agency of Slovakia and focused on pure populations. However, morphological inspection of the sampled trees during the material collection revealed species admixture in some cases (Table 1); in those cases, all species present in the forest were sampled.

Twigs with at least six fully developed leaves were collected from 30 adult trees per population on average (1 twig per tree) from the insolated part of the crown. After bringing the samples to laboratory, the best-preserved leaf was archived in a herbarium and used for morphometry, another leaf was silica-dried and used for DNA isolation. Thirteen leaf traits were scored (electronic supplementary materials, Table S1), which are traditionally used in the taxonomical evaluation of the species concerned (cf Kremer et al., 2002).

Climatic variables of the population sites were taken from two sources: the WorldClim high-resolution interpolated climate database (Hijmans et al., 2005), and the ClimateEU database (Hamann et al., 2013). The climatic data represent the mean values of each climatic variable within the period 1960–1990 at the 30 arc-second resolution for both databases (the used climatic characteristics are listed in Table S2).

DNA extraction, microsatellite and AFLP analysis

Total genomic DNA was extracted from the silica-dried leaves using a modified CTAB protocol following Doyle & Doyle (1987). Twenty nuclear microsatellite markers were studied. The PCR were done in two multiplex reactions: a 12-plex of EST-SSR (Durand et al., 2010) was composed of 0.8 μl of extracted DNA, 0.16 μl of 12 primers 10μM each (except PIE 242 and PIE152 80μM), 1 μl of Q-solution, 2.5 μl Qiagen multiplex PCR Kit and 0.54 μl water, while the 8-plex of genomic SSR (Guichoux et al., 2011) was composed of 0.8 μl of extracted DNA, 0.15 μl of 8 primers 10μM each, 1 μl of Q-solution, 2.5 μl Qiagen multiplex PCR Kit and 0.55 μl water (total volume of 5 μl). Primer sequences are listed in Table S1. The PCR profile was identical for both multiplexes: initial denaturation at 95 °C for 15 minutes followed by 35 cycles of denaturation at 95 °C for 30 seconds; annealing at 56 °C for 1 minute and extension at 72 °C for 45 seconds, and final extension at 60 °C for 20 minutes.

For the ALFP analyses, seven randomly chosen samples per population were used. DNA cleaving was made in 10 μl mixtures containing 5 U EcoRI (NEB R3101S), 2.5 U MseI (NEB R0525S), 1 μl Cut-Smart buffer and 2 μl extracted DNA. This mixture was incubated at 37 °C for 20 minutes and then at

Table 1. Geographical positions of the studied populations

No.	Location	Latitude	Longitude	Altitude [m a.s.l.]	Dominant species	Admixture
1	Bogdalický vrch	48°24'19"	16°54'17"	150	*Q. robur*	
2	Číčov	47°46'12"	17°43'49"	115	*Q. robur*	
3	Chynoriansky luh	48°37'16"	18°15'59"	190	*Q. robur*	
4	Turík	49°06'48"	19°22'52"	640	*Q. robur*	
5	Moldava nad Bodvou	48°37'16"	21°00'06"	210	*Q. robur*	
6	Latorický luh	48°27'40"	22°05'59"	105	*Q. robur*	
7	Kurinec	48°20'58"	20°00'58"	240	*Q. robur*	
8	Dolné Orešany	48°26'47"	17°24'45"	300	*Q. petraea*	
9	Bajtava	47°51'21"	18°45'50"	220	*Q. petraea*	
10	Kašivárová	48°28'16"	18°46'23"	580	*Q. petraea*	
11	Príboj	48°44'47"	19°13'48"	400	*Q. petraea*	
12	Kvetnica	49°00'27"	20°17'12"	740	*Q. petraea*	
13	Sivec	48°51'04"	21°05'31"	590	*Q. petraea*	
14	Šarišský hradný vrch	49°02'58"	21°10'45"	460	*Q. petraea*	
15	Kamenica nad Hronom	47°49'33"	18°45'24"	170	*Q. pubescens*	
16	Soroška	48°36'13"	20°38'22"	490	*Q. pubescens*	
17	Humenec	48°51'26"	21°09'58"	470	*Q. pubescens*	
18	Kokošovská dubina	48°57'05"	21°20'30"	540	*Q. petraea*	*Q. robur*
19	Horša	48°14'26"	18°42'01"	190	*Q. pubescens*	*Q. petraea*

65 °C for another 20 minutes. Subsequently, 5 μl of ligation mixture was added so final mixture contained 1 x T4 DNA ligase buffer, 1 U T4 DNA ligase (NEB M0202S), 0.11 M NaCl, 1.5 μg BSA, 1 mM ATP (NEB P07567), 1.5 μl MseI adapter (50 μM) and 1.5 μl EcoRI adapter (5 μM). The total mixture was incubated at 16 °C for 12 hours and then at 65 °C for 10 minutes. For preselective PCR, the product of ligation was diluted by 100 μl ultrapure water. 2 μl of the diluted product was added to 8 μl mixture containing 0.4 U FIREPol DNA polymerase (Solis BioDyne), 1× FirePol buffer B, 0.2 μM dNTP, 1.5 mM MgCl$_2$, 0.25 μM primer EcoRI+A, 0.25 μM primer MseI+C. Thermal profile for preselective PCR was as follows: initial denaturation at 72 °C for 2 minutes followed by 25 cycles of denaturation at 94 °C for 30 seconds, annealing at 56 °C for 30 seconds, and elongation at 72 °C for 2 minutes, and final extension at 60 °C for 10 minutes. The product of this preselective PCR was diluted with 90 μl of ultrapure water. Selective PCR was accomplished in 10 μl of mixture containing 0.4 U HOT FIREPol DNA polymerase (Solis Bio-Dyne), 1 x HOT FIREPol DNA polymerase buffer B2, 0.2 mM dNTPs, 0.2 mM MgCl$_2$, combination of 0.54 μM selective EcoRI (+AGG; +AGC; +ACT; +AAG) primers and 0.54 μM of selective primer MseI+C-CA and 2 μl of diluted product of preselective PCR. The program for selective PCR was as follows: initial denaturation at 95 °C for 15 minutes, 33 cycles of denaturation at 95 °C for 30 seconds; annealing starting at 61 °C for 30 seconds during the first 8 cycles when temperature was reduced by 0.5 °C per cycle and stayed constant at 57 °C for 30 seconds during the remaining 25 cycles, and extension at 72 °C for 1 minute, followed by final extension at 72 °C for 10 minutes. The mixture of 0.5 μl of each selective combination (total 2.0 μl) was added to 7.5 μl of formamide and 0.5 μl of a size standard and denatured for 3 min at 94 °C.

Fragment analyses were performed on ABI PRISM 3130 genetic analyzer. Genotypes were obtained from raw data using GeneMapper 4.0.

Data analysis

Prior to data treatment, specimens with more than 30% missing data at nSSR loci were excluded from all analyses.

Canonical discriminant analysis (CDA) was applied to morphometric data (procedure CANDISC; SAS, 1988) as a way to classify the species on an objective basis. Prior species classification of specimens relied on field identification. A separate CDA was run to distinguish *Q. petraea* from *Q. pubescens*.

To assess the pattern of allele sharing among species, a modelbased clustering procedure following Pritchard et al. (2000) was used. The clustering procedure assigns multilocus genotypes of the sampled trees to a predefined number K of genetic groups ('gene pools', clusters) in such a way that Hardy-Weinberg and linkage disequilibria within clusters are maximally explained, based on Bayesian inference. The program STRUCTURE version 2.3.5 (Pritchard et al., 2000) was used for the analyses. The procedure was run 10 times for each $K = 1$–10, with a burn-in period of 100 000 and subsequent 1 000 000 iterations without prior information on the population of origin to determine the number of clusters. Since we supposed the existence of gene exchange among populations of different origins as well as among species, we used the admixture model estimating the fraction of the genome that each individual draws from each cluster. The optimum number of clusters was determined using the ΔK criterion of Evanno et al. (2005). The run with the highest posterior probability at each K was selected for interpretation.

To identify signals of divergent selection in potentially adaptive AFLP markers, we used two approaches. First, frequency-based detection of outlier loci was performed using BayeScan (Foll & Gaggiotti, 2008) to identify loci exhibiting excessive interspecific differentiation. BayeScan v2.1 was used with default settings (20 pilot runs of 5000 iterations each and a burn-in of 50,000 iterations followed by 50,000 iterations for the estimation of the posterior distributions with a thinning interval of 10; prior odds for the neutral model were set to 10). Second, AFLP frequencies were tested for correlations with climatic variables using multiple univariate logistic regression models with the spatial analysis method (SAM), as implemented in the Sambada v0.5.1 software (Joost et al., 2007). The significance of correlations was assessed using the likelihood-ratio test (*G*-test) and the Wald test implemented in SAM, applying an initial 95% confidence interval. The results of Wald and G tests from Samβada were examined for the significance and then corrected for multiple testing by the Benjamini-Hochberg procedure.

Results

Morphometry

Surprisingly, the canonical discriminant analysis revealed the presence of four quite separated clusters (Fig. 1). The cluster of *Q. robur* seems to be the most consistent, not containing any misplaced individual of the other two species. However, 6 specimens of *Q. robur* are placed into the cluster of *Q. petraea*, which also contains 7 misplaced specimens of *Q. pubescens*. The cluster of *Q. pubescens* also comprises two misplaced *Q. petraea* individuals. Moreover, in the center
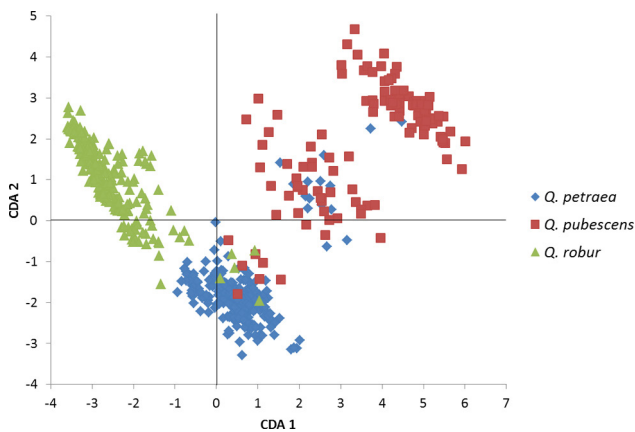
Fig. 1 Canonical discriminant analysis based on morphological traits: a plot of first two CDA axes

between the clusters of *Q. petraea* and *Q. pubescens*, there is a fourth cluster, well separated from the former two, containing a mixture of *Q. petraea* and *Q. pubescens*.

Neutral variation patterns

In the Bayesian analysis (STRUCTURE; Pritchard et al., 2002) run for the nSSR dataset, the ΔK measure (Evanno et al., 2005) indicated $K = 2$ as the most likely number of groups (Fig. S1a). Nevertheless, genetic structure may be hierarchical, and consequently several numbers of groups can be adequate to describe the data. Therefore, results of the STRUCTURE analyses for both *K* ranging from 2 to 4 are presented (Fig. 2). As expected, the results of all analyses reflect species identity well. Under $K = 2$, *Q. robur* trees have generally a high proportion of one cluster while the second predominates in the genomes of *Q. petraea* and *Q. pubescens*. Under $K = 3$, the clusters correspond almost completely to one species each. The situation is very similar for $K = 4$, just *Q. robur* is split into two clusters. However, their proportions do not show substantial differences between populations, and neither a geographical trend. Nevertheless, in most populations, there is a small part of individuals whose inferred ancestry contradicts their species identity based on field classification relying on phenotypic traits.

As the CDA showed quite a good separation of *Q. robur* from the remaining two species along the first CDA axis, we used the CDA 1 scores as a proxy for species identity and compared it to the outcomes of the Structure analysis for $K = 3$ (Fig. 3). The correlation between canonical scores and the proportion of *Q. robur* (3rd cluster) genes was highly significant ($r = -0.768^{***}$). In spite of this, there were five *Q. robur* trees having less than 50% of *Q. robur* genes, in two cases even the proportion was close to zero. Nevertheless, all these trees had relatively high CDA 1 scores, which is an indication of intermediate leaf morphology and thus potential introgression in previous
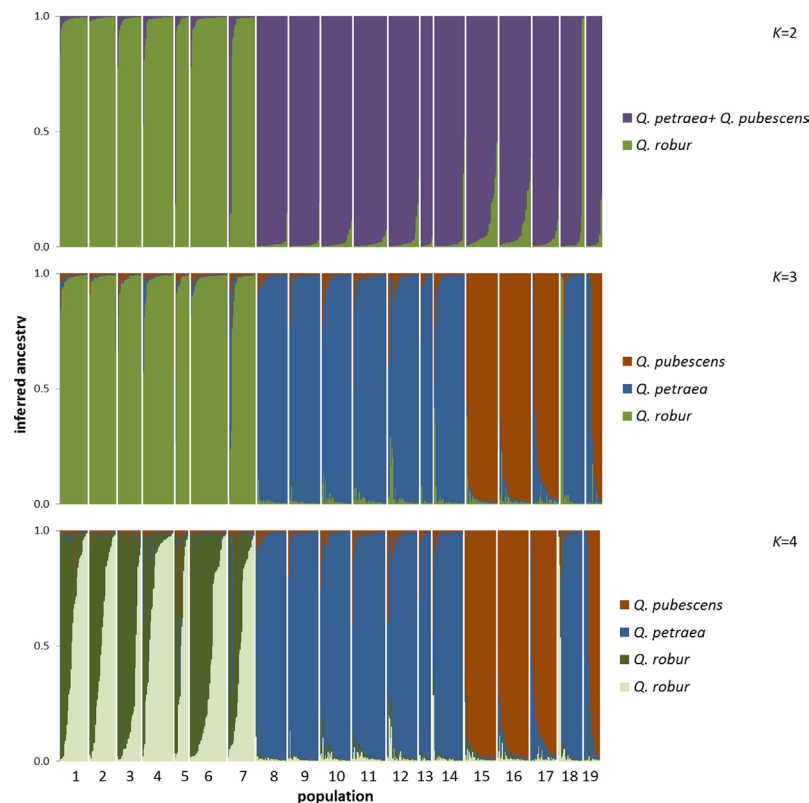

Fig. 2 Ancestry of individuals inferred by the Structure analysis based on nSSR for numbers of clusters $K = 2$ to 4. Species identity derived from Structure
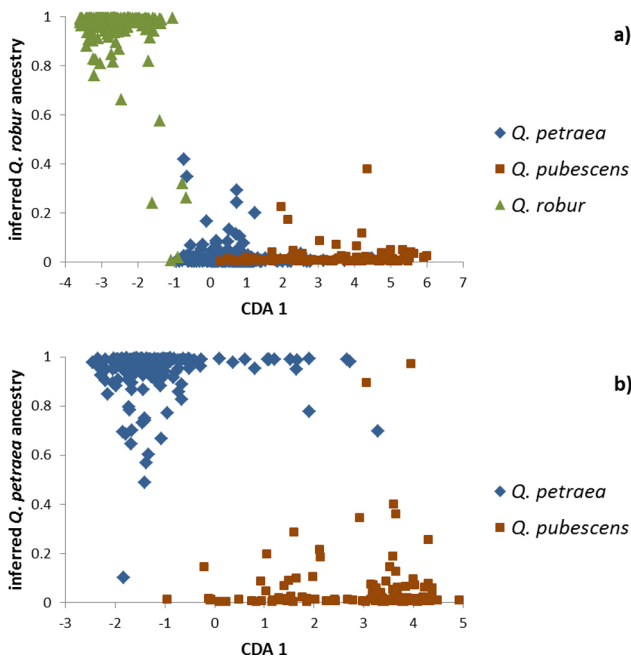
Fig. 3 The relationship between classification based on CDA and the ancestry of individuals inferred by the Structure analysis based on nSSR at $K = 3$: a) CDA for all three species, b) separate CDA for *Q. petraea* and *Q. pubescens*. Species identity relies on field classification

generations. The CDA run for all three species did not properly discriminate *Q. petraea* from Q. *pubescens*; a separate analysis was thus done for these two species, and the scores on the 1st DCA axis were again plotted against the inferred *Q. petraea* ancestry (based again on $K = 3$). The overlap along both axes is substantially larger in this case. However, the proportion of trees with inferred ancestry of a particular species around 50 % (which may represent F1 hybrids) is generally quite low. Although low numbers of phenotypically or genetically intermediate individuals did not allow statistical testing of differences in their frequencies, the two mixed populations (18 Kokošovská dubina and 19 Horša) do not seem to have higher proportion of intermediates than pure populations.

Search for selection signals

For potentially adaptive AFLP markers, the ΔK measure applied to the outcomes of the Bayesian analysis of population structure indicated $K = 4$ as the most likely number of clusters, which seems also to be the only structure interpretable in terms of species identity (Fig. S1b). One of the clusters predominates in *Q. robur*, another one mainly in *Q. petraea*, and the third one in *Q. pubescens*. The fourth one is present at higher frequency in just one population of *Q. pubescens*. The structure at $K = 2$ is completely uninterpretable; one of the gene pools occur at high frequencies in all three species, the other appears just in few individuals. For $K = 3$, one group is attributable to *Q. robur*, another one to *Q. petraea* and *Q. pubescens*, while the third one could not be assigned to any species (Fig. S2).

Although the correspondence between the patterns of nSSR-based and AFLP-based genetic structures is not poor (correlation coefficients between the inferred ancestry based on nSSR and AFLP were 0.756[***] and 0.709[***] for the *Q. robur* and *Q. petraea* gene pool, respectively), there is still a substantial portion of trees, for which nSSRs predict different ancestry than AFLPs. *Q. robur* seems to be the most consistent in this respect – only three individuals were found, for which the phenotypic classification in the field contradicts ancestry inferred based on AFLP genotypes (Fig. 4a). On the other hand, AFLPs predict different ancestry in quite many *Q. petraea* and *Q. pubescens* individuals (Fig. 4a, b).

As direct approaches to search for selection signals, BayeScan and spatial analysis method were used. BayeScan failed to find any polymorphism showing unusual level of differentiation; for all AFLP loci, posterior odds for model including diversifying selection (Bayes factors) lied far below the limit of sufficient evidence (Table S3). On the other hand, SAM revealed two AFLP markers responsive to climatic variables: the G_73.8309 locus was correlated
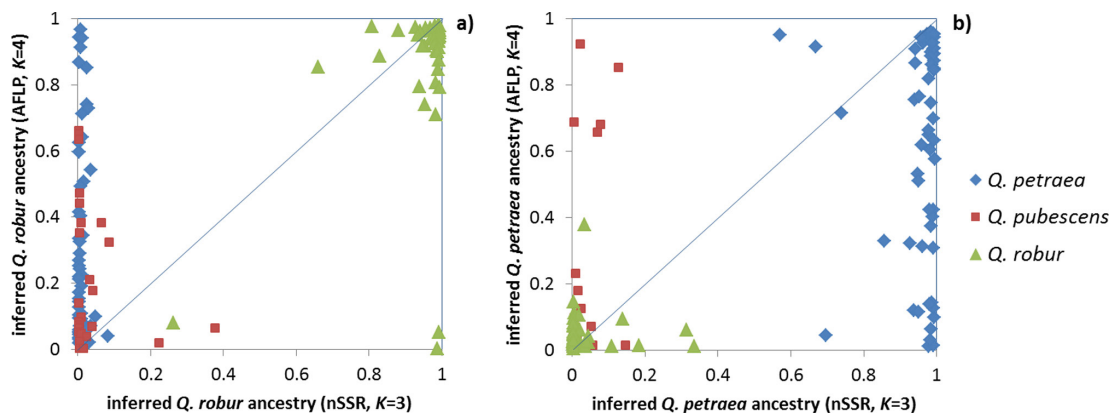

Fig. 4 Relationship between the ancestries of individuals inferred by the Structure analysis based on nSSR and AFLP: a) *Q. robur* ancestry, b) *Q. petraea* ancestry. Species identity relies on field classification

Table 2 Summary of AFLP loci exhibiting environmental associations, remaining significant at least at $P<0.01$ after Benjamini-Hochberg correction

Marker	Environmental variable	G-score	P^1	BHc[2]	Wald-score	P^1	BHc[2]	McFadden adjusted R^2	β_0	β_1
G_73.8309	VAPRVS	29.58	5.37E-08	***	22.81	1.79E-06	**	0.1336	−21.474	15.885
	t_{max03}	23.17	1.48E-06	***	18.19	2.00E-05	**	0.1001	−7.167	0.6703
	t_{max04}	22.59	2.01E-06	***	17.82	2.43E-05	**	0.0971	−9.505	0.6027
	altitude	22.58	2.02E-06	***	17.66	2.64E-05	**	0.0970	0.870	−0.0047
	BIO7	20.16	7.11E-06	**	17.03	3.68E-05	**	0.0844	−18.445	0.5823
Y_124.3018	bFFP	20.07	7.45E-06	**	17.72	2.55E-05	**	0.0790	10.526	−0.1015
	BIO4	19.39	1.07E-05	**	17.27	3.25E-05	**	0.0756	−21.952	0.0278

[1]test probability without correction for multiple testing; [2]significance of the test after Benjamini-Hochberg correction; β_0 and β_1 – intercept and slope of the logistic regression model, respectively; VAPRVS – average air vapour pressure during the vegetation season (May to October); t_{max03} and t_{max04} – maximum daily temperature in March and April, respectively; BIO7 – temperature annual range; bFFP – the Julian date on which frost-free period begins; BIO4 – temperature seasonality; **$P<0.01$, ***$P<0.001$.

primarily with vapor pressure during the vegetation season and with temperature-related variables (maximum temperatures in March and April, temperature annual range, and altitude, which is also correlated with temperature regime of the site), which the Y_124.3018 locus was correlated with the date of the beginning of frost-free period and temperature seasonality (Table 2). Nevertheless, both markers showed correlations (even though not as close) with a plenty of environmental indicators; in general, more with temperature-related than with precipitation-related variables (Table S4).

Discussion

The genus *Quercus* is known to be a complex of species, which retain cohesion but are interconnected by limited gene exchange (Cronk & Suarez-Gonzalez, 2018; Cannon & Petit, 2019). Hybridization and introgression are among closely related oak species are common and widespread, as evidenced by the presence of morphologically intermediate individuals (Viscosi et al., 2012), various types of gene markers (Petit et al., 2003; Gömöry & Schmidtová, 2007), as well as genome-wide investigations (McVay et al., 2017; Leroy et al., 2019a). A natural consequence of such a common gene exchange is a complicated and controversial systematics, mainly on the species level. While Linnaeus (1753) distinguished just 12 oak species, the number of species within the genus has increased to several hundred. Their arrangement in subgenera and sections is not unambiguously resolved either. In Slovakia, three white oak species have traditionally been accepted as native, namely *Quercus robur* L., *Q. petraea* (Matt.) Liebl. and *Q. pubescens* Willd. However, five other taxa were identified in natural or seminatural forests of Slovakia: *Q. pedunculiflora* C. Koch., *Q. banatus* P. Kučera (formerly *Q. dalechampii* Ten.), *Q. polycarpa* Schur., *Q. virgiliana* Ten. and *Q. frainetto* Ten. (Magic, 1975). In Flora Europaea, they are listed as separate species (except *Q.*

virgiliana, which is treated as synonym of *Q. pubescens*) having the centers of their distribution ranges mainly in southern and eastern Europe (https://eunis.eea.europa.eu/references/1780/species). Exact delimitation of the boundaries of their distribution and even their taxonomic status is a matter of controversies among dendrologists, however. European white oaks are generally capable of interspecific hybridization (although the hybridization success varies and depends on a particular pair of parental species); consequently, oaks are characterized by a great intraspecific morphological variation making taxonomical classification eventually difficult. Therefore, we treated *Q. robur*, *Q. petraea* and *Q. pubescens* as species *sensu lato*, comprising the related minor species. Our *Q. robur* samples may thus include some *Q. pedunculiflora* while *Q. petraea* sample set may contain *Q. banatus*.

In contrast with our earlier allozyme-based study (Gömöry & Schmidtová, 2007), *Q. robur* seems to be the most consistent species both morphologically and genetically. On the other hand, there is much overlap between *Q. petraea* and *Q. pubescens*; intermediate individuals in terms of leaf morphology even form a separate group. Naturally, opportunities for hybridization between these two species are much more frequent than with *Q. robur*. As shown by Lepais et al. (2009), direction and rate of introgression depends on relative abundance of hybridizing species. In Slovakia, *Q. pubescens* is predominantly the main component of xerothermophilous communities on very specific sites (steep south-oriented slopes on mostly volcanic or carbonate rocks). Even though the admixture of other white oak species directly in the stands is typically negligible, *Q. petraea* stands are frequently located in close vicinity; this was also the case of all sampled *Q. pubescens* populations. On the other hand, most of our sampled *Q. robur* stands were riparian forests located in lowlands and geographically separated from populations of the other two species. There were two exceptions. The population Turík is situated on a site very untypical for *Q.*

robur, at the northern limit of *Q. robur* distribution in Slovakia at quite a high elevation. It may represent a colonization relict, which has persisted until the present times surrounded by beech and fir forests, without any neighboring or closely located *Q. petraea*. The population Moldava nad Bodvou is a riparian forest, but located on the foothills of a dry calcareous plateau with a rich occurrence of *Q. pubescens*. Genetic admixture in both populations is negligible, only one hybrid tree was found in Moldava nad Bodvou. *Q. robur* can much more easily be fertilized by *Q. petraea* or *Q. pubescens* pollen than vice versa (Bacilieri et al., 1996; Lepais et al., 2013); consequently, higher genomic contributions of these species in the *Q. robur* gene pool are expected. Nevertheless, Moldava na Bodvou is an exception, and here a shift of flowering time caused by altitudinal difference may play a role in gene flow limitation. The other *Q. robur* populations within our population set are geographically isolated, which may be the reason why this expectation was not fulfilled. In the light of this fact, our sampling strategy may be questioned. However, our study focused on indigenous natural or close-to-natural forests, which in case of *Q. robur* can only be found in riparian forests. *Q. robur* can of course be found also elsewhere but the autochthony of such stands is not warranted, as oaks have been extensively planted in the past. We thus avoided sampling of *Q. robur* at higher elevations, where the stands are in a closer contact with *Q. petraea* but the autochthony is questionable.

Taking into consideration geographical settings, hybridization and introgression levels derived from the inferred ancestry (both nSSR- and AFLP-based) are consistent with the findings of Curtu et al. (2007) in a mixed community of four oak species, where the introgression rates ranged from 1.7 % to 16.2 % (the latter for a pair of closely related species *Q. pubescens* and *Q. frainetto*, however). On the other hand, much higher introgression rates were observed in other studies (e.g., Beatty et al., 2016, found only 45 % of trees in mixed *Q. robur–Q. petraea* stands to be pure species, which, however, can partly be attributed to a different methodology).

The patterns of ancestry derived from AFLP are substantially different, especially for the pair of species *Q. petraea – Q. pubescens*, while *Q. robur* appeared again to be quite consistent. However, the presumption of selective neutrality required for the Bayesian clustering procedure Structure is not fully valid for AFLPs. As anonymous markers distributed randomly across the genome, polymorphisms underlying the presence/absence of AFLP fragments may be located in exons of genes or other sequences with adaptive significance. Diversifying selection has always been proposed as the mechanism maintaining species integrity of white oaks (Petit et al., 2003), and the genes

or sequences involved in this process were studied. Gömöry (2000) found an allozyme locus (*Gludh-A*) showing unusually high interspecific differentiation (F_{ST} = 0.664). Similarly, a beta-tubulin gene was found to be a differentiation outlier in the study of Porth et al. (2016). Another type of evidence comes from gene-expression studies: a set of differentially expressed genes in response to waterlogging was identified by Le Provost et al. (2012). Recently, and extensive study based on 31 million SNPs revealed genes contributing to reproductive isolation between white oak species (Leroy et al., 2019b).

We identified two AFLP fragments showing significant associations with several climatic variables. Of course, climatic predictors were partly collinear (which is a problem of almost all genotype-environment association studies). It is therefore difficult to assess, for which variable the association with genotype is causal, especially when physiological mechanisms underlying interspecific differentiation are largely unknown. Nevertheless, for all these climatic proxies, their inclusion among drivers of interspecific divergence is plausible.

In the case of sympatric complexes of closely related species, the taxa frequently share large parts of their gene pools. The reason for this may be both shared ancestral polymorphisms and interspecific gene flow. Genetic drift is expected to eliminate shared polymorphisms unless other microevolutionary mechanisms interfere (Wakeley & Hey, 1997); however, allele loss requires a certain time. There is experimental evidence that genes may be shared among well-established species even under the absence of gene flow (Whittemore, 2005). Nevertheless, there is enough empirical data about interspecific gene exchange in white oaks to expect it also in Slovak populations. Even though none of AFLP fragments showed excessive differentiation, we consider the observed associations with climatic variables linked to temperature and water regimes to be sufficient proof of selective forces driving interspecific divergence and maintaining integrity of white oak species.

In terms of forestry practice, our study confirmed that species boundaries among white oak species are partly permeable but exist. The proportion of hybrids or introgressed individuals is lower than reported from western Europe (Rushton, 1993; Petit et al., 2003). This may, however, be related to the fact that the sites suitable for either species are geographically more separated in the Carpathians and adjacent areas than in the western regions, and in spite of a mosaic-like sympatry the interspecific gene flow under natural circumstances is less effective. If the practice of planting of oak species on inappropriate sites or mixing of reproductive materials continued, it could break this isolation and increase hybridization rates. This would not be a huge problem as such, as natural

selection seems to eliminate improper genotypes and maintain species integrity, but it may decrease average fitness of newly established stands and enhance the stress imposed by the ongoing climate change, which already can be observed in oak stands in several regions of eastern Europe. Inspection of oak basic materials (especially seed stands) and removing trees not belonging to the major species is advisable. Careful species identification is also necessary in the case of materials for *ex situ* conservation (indispensable, considering a large-scale decline of oak forests in the climatically most stressed areas), even when all collection is made in semi-natural stands.

The study also underlines the importance of oak reserves (both those established by nature conservation agencies and gene reserves) as natural laboratories, where the extent of interspecific hybridization and introgression can be monitored and quantified, and which can serve as reference standards to evaluate these processes in commercial stands.

Acknowledgements

References

Abadie P, Roussel G, Dencausse B, Bonnet C, Bertocchi E, Louvet JM, Kremer A & Garnier-Géré P (2012) Strength, diversity and plasticity of postmating reproductive barriers between two hybridizing oak species (*Quercus robur* L. and *Quercus petraea* (Matt) Liebl.). Journal of Evolutionary Biology 25: 157–173. doi:10.1111/j.1420-9101.2011.02414.x.

Arnold ML (2006) Evolution through genetic exchange. Oxford University Press, Oxford, UK.

Arnold ML & Hodges SA (1995) Are natural hybrids fit or unfit relative to their parents? Trends in Ecology and Evolution 10: 67–71.

Arnold ML, Kentner EK, Johnston JA, Cornman S & Bouck AC (2001) Natural hybridisation and fitness. Taxon 50: 93–104.

Bacilieri R, Ducousso A, Petit RJ & Kremer A (1996) Mating system and asymmetric hybridization in a mixed stand of European oaks. Evolution 50: 900–908.

Beatty GE, Montgomery WI, Spaans F, Tosh DG & Provan J (2016) Pure species in a continuum of genetic and morphological variation: sympatric oaks at the edge of their range. Annals of Botany 117: 541–549. doi:10.1093/aob/mcw002.

Cannon CH & Petit RJ (2019) The oak syngameon: more than the sum of its parts. New Phytologist. doi:10.1111/nph.16091.

Cronk QC & Suarez-Gonzalez A (2018) The role of interspecific hybridization in adaptive potential at range margins. Molecular Ecology 27: 4653–4656. doi:10.1111/mec.14927.

Curtu AL, Gailing O & Finkeldey R (2007) Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. BMC Evolutionary Biology 7: 218.

Doyle JJ & Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11–15.

Durand J, Bodénès C, Chancerel E, Frigerio JM, Vendramin G, Sebastiani F, Buonamici A, Gailing O, Koelewijn HP, Villani F, Mattioni C, Cherubini M, Goicoechea PG, Herrán A, Ikaran Z, Cabane C, Ueno S, Alberto F, Dumoulin PY, Guichoux E, de Daruvar A, Kremer A & Plomion C (2010) A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. BMC Genomics 11: 570. doi:10.1186/1471-2164-11-570.

Evanno G, Regnaut S & Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611–2620.

Foll M & Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180: 977–993.

Gömöry D (2000) A gene coding for a non-specific NAD-dependent dehydrogenase shows a strong differentiation between *Quercus robur* and *Quercus petraea*. Forest Genetics 7: 167–170.

Gömöry D & Schmidtová J (2007) Extent of nuclear genome sharing among white oak species (*Quercus* L. subgen. *Lepidobalanus* (Endl.) Oerst.) in Slovakia estimated by allozymes. Plant Systematics and Evolution 266: 253–264.

Gross BL & Rieseberg LH (2005) The ecological genetics of homoploid hybrid speciation. Journal of Heredity 96: 241–252.

Guichoux E, Lagache L, Wagner S, Léger P & Petit RJ (2011) Two highly validated multiplexes (12-plex and 8-plex) for species delimitation and parentage analysis in oaks (*Quercus* spp.). Molecular Ecology Resources 11: 578–585. doi:10.1111/j.1755-0998.2011.02983.x.

Hamann A, Wang T, Spittlehouse DL & Murdock TQ (2013) A comprehensive, high-resolution database of historical and projected climate surfaces for western North America. Bulletin of the American Meteorological Society 94: 1307–1309.

Hijmans RJ, Cameron SE, Parra JL, Jones PG & Jarvis A (2005) Very high resolution interpolated surfaces for global land areas. International Journal of Climatology 25: 1965–1978.

Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G & Taberlet P (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Molecular Ecology 16: 3955–3969.

Kremer A, Dupouey JL, Deans JD, Cottrell J, Csaikl U, Finkeldey R, Espinel S, Jensen J, Kleinschmit J, Van Dam B, Ducousso A, Forrest I, Lopez de Heredia U, Lowe AJ, Tutková M, Munro RC, Steinhoff S & Badeau V (2002) Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. Annals of Forest Science 59: 777–787.

Le Provost, G, Sulmon C, Frigerio JM, Bodénès C, Kremer A & Plomion C (2012) Role of waterlogging-responsive genes in shaping interspecific differentiation between two sympatric oak species. Tree Physiology 32: 119–134. doi:10.1093/treephys/tpr123.

Lepais O, Petit RJ, Guichoux E, Lavabre JE, Alberto F, Kremer A & Gerber S (2009) Species relative abundance and direction of introgression in oaks. Molecular Ecology 18: 2228–2242. doi:10.1111/j.1365-294X.2009.04137.x.

Lepais O, Roussel G, Hubert F, Kremer A & Gerber S (2013) Strength and variability of postmating reproductive isolating barriers between four European white oak species. Tree Genetics & Genomes 9: 841–853. doi:10.1007/s11295-013-0602-3.

Leroy T, Louvet J-M, Lalanne C, Le Provost G, Labadie K, Aury J-M, Delzon S, Plomion C & Kremer A (2019a) Adaptive introgression as a driver of local adaptation to climate in European white oaks. New Phytologist doi:10.1111/nph.16095.

Leroy T, Rougemont Q, Dupouey J-L, Bodénès C, Lalanne C, Belser C, Labadie K, Le Provost G, Aury J-M, Kremer A & Plomion C (2019b) Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. New Phytologist doi: 10.1111/nph.16039.

Lexer C, Kremer A & Petit RJ (2006) Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. Molecular Ecology 15: 2007–2012.

Linnaeus C (1753) Species Plantarum, Tomus II. Laurentius Salvius, Stockholm, Sweden.

Magic D (1975) Taxonomické poznámky k doterajšiemu výskumu dubov v Západných Karpatoch. Biológia 30: 65–74.

Mallet J (2007) Hybrid speciation. Nature 446: 279–283.

Mayr E (1963) Animal species and evolution. Harvard University Press, Cambridge MA, USA.

McVay JD, Hauser D, Hipp AL & Manos PS (2017) Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in western North America. Genome 60: 733–742. doi:10.1139/gen-2016-0206.

Petit RJ, Bodénès C, Ducousso A, Roussel G & Kremer A (2003) Hybridization as a mechanism of invasion in oaks. New Phytologist 161: 151–164. doi:10.1046/j.1469-8137.2003.00944.x.

Porth I, Garnier-Géré P, Klápště J, Scotti-Saintagne C, El-Kassaby YA, Burg K & Kremer A (2016) Species-specific alleles at a beta-tubulin gene show significant associations with leaf morphological variation within *Quercus petraea* and *Q. robur* populations. Tree Genetics & Genomes 12 Article Number: 81. doi:10.1007/s11295-016-1041-8.

Pritchard JK, Stephens M & Donnelly P (2000) Inference of population structure from multilocus genotype data. Genetics 155: 945–959.

Rhymer JM & Simberloff D (1996) Extinction by hybridization and introgression. Annual Review of Ecology and Systematics 27: 83–109.

Rieseberg LH & Carney SE (1998) Plant hybridization. New Phytologist 140: 599–624.

Rushton BS (1993) Natural hybridization within the genus *Quercus* L. Annales des Sciences Forestières 50 (Supplement 1): S73–S90.

SAS (1988) SAS/STAT® User's guide: release 6.03 Ed. SAS Institute Inc., Cary NC, USA.

Schwarzbach AE, Donovan LA & Rieseberg LH (2001) Transgressive character expression in a hybrid sunflower species. American Journal of Botany 88: 270–277.

Schweitzer JA, Martinsen GD & Whitham TG (2002) Cottonwood hybrids gain fitness traits of both parents: A mechanism for their long-term persistence? American Journal of Botany 89: 981–990.

Tulstrup NP (1959) International trade in forest tree seed. Unasylva 13: 7.

Vazquez FM, Perez MC, Esparrago F & Burzaco A (1993) Hibridos del genero *Quercus* L. en Extremadura: Congreso Forestal Español, Lourizán 1993, Ponencias y comunicaciones, Tomo 1459 (ed. by FJ Silva-Pando) Xunta de Galicia, Vigo, Spain, pp. 459–465.

Viscosi V, Antonecchia G, Lepais O, Fortini P, Gerber S & Loy A (2012) Leaf shape and size differentiation in white oaks: assessment of allometric relationships among three sympatric species and their hybrids. International Journal of Plant Sciences 173: 875–884. doi:10.1086/667234.

Wakeley J & Hey J (1997) Estimating ancestral population parameters. Genetics 145: 847–855.

Whittemore AT (2005) Genetic structure, lack of introgression, and taxonomic status in the *Celtis laevigata – Celtis reticulata* complex (Cannabaceae). Systematic Botany 30: 809–817.