



HAL
open science

Using joint multivariate analyses of leaf morphology and molecular-genetic markers for taxon identification in three hybridizing European white oak species (*Quercus* spp.)

Christian Rellstab, Andreas Bühler, René Graf, Catherine Folly, Felix Gugerli

► To cite this version:

Christian Rellstab, Andreas Bühler, René Graf, Catherine Folly, Felix Gugerli. 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*, 2016, 73 (3), pp.669-679. <10.1007/s13595-016-0552-7>. <hal-01532466>

HAL Id: hal-01532466

<https://hal.science/hal-01532466v1>

Submitted on 2 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Using joint multivariate analyses of leaf morphology and molecular-genetic markers for taxon identification in three hybridizing European white oak species (*Quercus* spp.)

Christian Rellstab¹ · Andreas Bühler¹ · René Graf¹ · Catherine Folly¹ · Felix Gugerli¹

Received: 16 November 2015 / Accepted: 7 April 2016 / Published online: 2 May 2016
© INRA and Springer-Verlag France 2016

Abstract

• **Key message** We show that joint multivariate analyses of leaf morphological characters and molecular-genetic markers improve the taxonomic assignment in hybridizing European white oaks. However, model-based approaches using genetic data alone represent straightforward alternatives to laborious, detailed morphological assessments.

Handling Editor: Ricardo Alia

Contribution of the co-authors Christian Rellstab: was involved in project planning and field work, performed all analyses, and wrote the manuscript.

Andreas Bühler: performed the morphological measurements and the multivariate analyses.

René Graf: established the molecular methods and genotyped the samples.

Catherine Folly: planned and performed the field work.

Felix Gugerli: led the project, was involved in planning the field and lab work, supported the statistical analyses, and reviewed and commented on successive drafts of the paper.

Electronic supplementary material The online version of this article (doi:10.1007/s13595-016-0552-7) contains supplementary material, which is available to authorized users.

✉ Christian Rellstab
christian.rellstab@wsl.ch

Andreas Bühler
andreasbuehler@gmail.com

René Graf
rene.graf@wsl.ch

Catherine Folly
catherinefolly@hotmail.com

Felix Gugerli
felix.gugerli@wsl.ch

- **Context** In European white oaks, species delimitation is debated because of large overlap of morphological characteristics likely due to hybridization.
- **Aims** We tested whether joint multivariate analyses of leaf morphology and molecular markers improve the identification of three oak species (*Quercus petraea*, *Quercus pubescens*, *Quercus robur*) compared to approaches using morphological or genetic variables only.
- **Methods** We assessed 13 leaf morphological characters and applied eight nuclear microsatellite markers in almost 1400 trees of 71 oak populations across Switzerland. We performed two multivariate approaches with three variable sets (morphology, genetics, combined) and assessed their performance in separating the taxa. We also compared the taxon assignment to a model-based clustering approach (STRUCTURE) based on genetic data alone.
- **Results** A joint use of morphological and genetic variables led to an improved taxon assignment. Whereas *Q. robur* could clearly be separated from the two other taxa, there was a certain overlap between *Q. petraea* and *Q. pubescens*. The STRUCTURE clustering led to the same taxon assignment in 85 % of the individuals.
- **Conclusion** It is important to consider both morphological and genetic properties in morphologically similar and hybridizing species. However, it might be more efficient to concentrate only on genetic markers than on time-consuming morphological assessments.

Keywords Factor analysis of mixed data · Introgression · Linear discriminant analysis · Nuclear microsatellite markers · Species assignment

1 Introduction

The identification of a species is of crucial importance not only for scientists that study, for example, biodiversity,

¹ WSL Swiss Federal Research Institute, Birmensdorf, Switzerland

hybridization, or paleoecology, but also for practitioners dealing with, e.g., species conservation or forest management. Ideally, species assignment is done directly in the field with diagnostic morphological characters. This is, however, not always possible, since in many genera (see below for examples) several morphologically similar taxa exist that are difficult to separate on the basis of a single or few morphological parameters. Moreover, especially in plants, some taxa hybridize and backcross (Mallet 2005; Whitney et al. 2010), leading to a morphological gradient along the character expression of the parental species involved.

In cases as described above, using molecular-genetic information can help in the delimitation of species. For example, the threatened butternut from eastern North America (*Juglans cinerea*) is hardly distinguishable morphologically from its hybrid with the exotic Japanese walnut (*Juglans ailantifolia*), but two randomly amplified polymorphic DNA (RAPD) markers can clearly differentiate between the two taxa (Ross-Davis et al. 2008). In the polyploid *Cardamine digitata* species aggregate, Jørgensen et al. (2008) used six microsatellite markers (hereafter referred to as nuclear simple sequence repeats, nSSRs) to separate four taxa of this controversial species complex. Other well-known examples come from the animal kingdom like the water flea species complex of *Daphnia longispina* (Petrušek et al. 2008; Rellstab et al. 2011) or different species of fruit fly (*Drosophila* spp., Harr et al. 1998; Routtu et al. 2007), whose members are difficult to distinguish morphologically and for which genetic markers helped for taxon assignment. In recent times, DNA barcoding—taxon identification using standardized DNA regions from organellar genomes (Valentini et al. 2009)—has become a common approach for identifying species that are morphologically difficult to assign or whose samples are hard to obtain. DNA barcoding, however, often uses rather conserved regions of the genome (like chloroplast or mitochondrial DNA) with low taxonomic resolution and uniparental inheritance.

In forest trees, the most prominent example comes from European white oaks, including the hybridizing *Quercus petraea*, *Quercus pubescens*, and *Quercus robur*. There has been a long debate of how to differentiate among these three species and which parameters are best suited to do so. Hybridization is known to occur between all three species pairs (Curtu et al. 2007a; Lepais and Gerber 2011) but is often asymmetric and depends on the local relative abundance and the mixture of the parental species (Bacilieri et al. 1996; Curtu et al. 2007a; Gerber et al. 2014; Lepais et al. 2009). *Q. robur* and *Q. petraea* show distinct differences in leaf morphology (Kremer et al. 2002), fruit structure (Rushton 1983), and neutral nSSR allele frequencies (Curtu et al. 2007b; Gugerli et al. 2007; Guichoux et al. 2011; Muir et al. 2000). In contrast, *Q. petraea* and *Q. pubescens* rather represent a morphological continuum than two morphologically distinct species (Dupouey and Badeau 1993; Viscosi et al. 2009). Although

some species-specific micromorphological characteristics exist (Fortini et al. 2015), they often greatly overlap (Bruschi et al. 2000). Moreover, genetic differentiation was found to be very low between *Q. petraea* and *Q. pubescens* (Bruschi et al. 2000; Curtu et al. 2007b; Neophytou et al. 2015; Viscosi et al. 2009), in contrast to the two other species pairs.

In all the examples described above, a joint, simultaneous analysis of morphology and genetic information was not applied. Instead, the two approaches were often performed in a stepwise procedure, by using one approach to confirm or support the other (e.g., Curtu et al. 2007a; Gugerli et al. 2007) or by using morphologically identified reference individuals to help in resolving the clustering or delimitation obtained by the genetic analysis (like in the case of *Daphnia* in Rellstab et al. 2011). A joint analysis, however, will most likely improve the classification of individuals and result in a higher resolution of species delimitation. If no single morphological or genetic diagnostic trait exists, multivariate analyses are ideal tools to identify parameters that differentiate among groups of individuals, which in turn can then be assigned to taxa.

In the present study, we aimed at thoroughly assessing both leaf morphology and nuclear genetic markers in all three white oak species occurring in Switzerland to clarify how the two types of parameters and their joint application support taxon identification. We sampled leaves of 71 Swiss oak populations and used 13 morphological parameters, eight nSSRs, and multivariate statistics to answer the following questions: Does the joint application of morphological parameters and neutral genetic markers in multivariate analyses improve the species assignment compared to using only morphological or genetic parameters? Which are the main parameters that differentiate among the species? We also performed a model-based clustering approach (STRUCTURE) using genetic data only and checked if it leads to congruent results compared to the multivariate analyses. Our study shows that the combination of morphological and genetic parameters improves species assignment in a hybridizing species complex including morphologically similar taxa. However, at least in the case of the oak species investigated here, model-based clustering using only genetic information already leads to a relatively accurate species assignment and may be considered as an alternative approach to the labor-intensive morphological measurements, especially since it allows more precise taxonomic assignment in the case of hybrids that do not always exhibit intermediate morphology.

2 Material and methods

2.1 Study species

European white oaks are represented in Switzerland by the three widely distributed *Q. petraea* (Matt.) Liebl. (hereafter

sometimes abbreviated as Pet), *Q. pubescens* Willd. (Pub), and *Q. robur* L. (Rob). *Q. petraea* and *Q. robur* are largely sympatric, though they are often locally separated according to their ecological niche. While *Q. petraea* is more drought-tolerant, *Q. robur* is often associated to deep, moist soil that expands into riparian hardwood forests. In turn, *Q. pubescens* as a sub-Mediterranean species grows mainly in the low-elevation inner alpine area (Valais) and along the calcareous, south-exposed dry Jura slopes within its Swiss range and can often be found in mixed stands with *Q. petraea*.

2.2 Sampling and genotyping

From May to July 2013, we sampled 71 oak tree populations (max. 20 individuals per population) from their entire distribution range within Switzerland (Table S1). We aimed at collecting all three species in all biogeographic regions of Switzerland if present. Information about potential populations were retrieved from previously published studies (e.g., Mátyás et al. 2002; Mátyás and Sperisen 2001), the Swiss National Forest Inventory (Brändli 2010), and the forest soil database of the Swiss Federal Institute WSL (described in Walther et al. 2013). The minimum distance between sampled trees was 20 m. For morphological analyses, we collected, if possible, one sunlit twig with extendable loppers and used a herbarium press to dry the leaves. For the genetic analyses, we dried four to five young leaves per tree on silica gel.

Per tree, DNA from 25 mg dried leaf tissue was extracted by LGC Genomics (Berlin, Germany) with a KingFisher 96 (Thermo Scientific, Waltham, USA) using the beadex maxi plant kit optimized for oak tree leaves (LGC Genomics). We genotyped all individuals with the eight nSSR markers of multiplex kit 2 from Guichoux et al. (2011) as described in the Appendix. Null allele frequencies were calculated with GENEPOP 4.2.2 (Rousset 2008). No marker exhibited substantially increased null allele frequencies across all populations; only in 23 of 568 cases, estimated null allele frequencies were above 10 %. To convert the genetic data (alleles) for use in multivariate analyses, we used GENALEX 6.5 (Peakall and Smouse 2012) to perform a principal coordinates analysis (PCoA) on pairwise co-dominant genotypic distances (Smouse and Peakall 1999). The resulting six principal components (accounting for 100 % of the variation) were later used as genetic parameters (PCo1 to PCo6) in the multivariate analyses.

2.3 Morphological parameters

Per tree, we analyzed the morphology of three intact leaves. In total, we measured nine parameters as described and denoted in Kremer et al. (2002): number of lobes (NL), number of intercalary veins (NV), basal shape of the lamina (BS), lamina length (LL), petiole length (PL), lobe width (LW), sinus width at the first lobe (SW), and length

of the lamina at the widest lobe of the leaf (widest point, WP). Additionally, similar to SW, we measured the width at the sinus above the widest lobe (SWW).

Since many of the above-described characteristics are dependent on the size of the leaves, we calculated seven relative parameters (of which two were transformed to meet the assumption of the methods applied) from those measured, similar as described and denoted in Kremer et al. (2002): lamina shape or obversity ($OB = 100 * WP/LL$), petiole ratio ($PR = 100 * PL/(LL + PL)$), lobe depth ratio at first lobe ($LDR = 100 * (LW - SW)/LW$), lobe depth ratio at widest lobe ($LDRW = 100 * (LW - SWW)/LW$), square root-transformed percentage of venation ($RPV = \text{SQRT}(100 * NV/NL)$), lobe width ratio ($LWR = 100 * LW/LL$), and natural log of lobe number ratio ($LLNR = \ln(NL/LL)$), the number of lobes relative to the length of the lamina.

Moreover, all leaves were inspected for the presence or absence of the following five hair types (for examples, see Fortini et al. 2015; Kissling 1977) using a stereo lens: stellate hair on the lamina (LS), clustered (fasciculate) hair on the lamina (LC), intermediate (between stellate and clustered) hair on the lamina (LI), stellate hair on the vein (VS), and clustered hair on the vein (VC). Single hairs were ignored.

For the multivariate analyses, we finally used (a) the seven relative parameters (OB, PR, LDR, LDRW, RPV, LWR, LLNR) and basal shape (BS), all averaged per tree; (b) the five hair parameters (LS, LC, LI, VS, VC) as presence/absence data per tree (presence = present on at least one leaf of the tree); and (c) the scores of the six principal components (PCo1 to PCo6) of the PCoA. For all analyses, we used only the samples with complete morphological measurement and more than five successfully scored nSSR loci (only ten samples yielded less than five loci), resulting in a total of 1369 samples (Table S1).

2.4 Multivariate analyses

All multivariate analyses were run in R 3.2.2 (R Development Core Team 2015). We performed two different approaches: a factorial analysis of mixed data (FAMD) followed by a hierarchical clustering on principal components (HCPC) using the FACTOMINER package (Lê et al. 2008) and a linear discriminant analysis (LDA) using MASS (Venables and Ripley 2002). Both multivariate analyses were run for three different datasets: using only morphological variables, using only genetic variables, and using all variables.

FAMD is a principal component method to explore data comprising both continuous and categorical variables (Pagès and Camiz 2008). FAMD does not require a prior division into groups. In our case, the continuous variables were the eight morphological variables and the scores of the six PCs on the basis of the genetic data as described above, and the categorical variables were represented by the five hair variables. After analysis, HCPC (Husson et al. 2010) was used to divide the

data into clusters keeping the first ten dimensions of the FAMD. The actual number of clusters is suggested by HCPC based on inertia gain. HCPC was performed with a minimum number of clusters of three, as suggested by the authors. Note that the FAMD with only genetic variables does not contain categorical variables and thus represents a principal component analysis (PCA).

LDA is a multivariate analysis that creates discriminant functions that best describe a priori defined groups using a training dataset. Subsequently, it uses these discriminant functions to assign individuals of a separate or larger dataset. In our analysis, we used hair types (that were shown to be important in the FAMD and HCPC, see Section 3) to define the groups (species) in the training dataset. We ignored the presence of intermediate hairs and hairs on the vein and used the samples having only one lamina hair type (stellate lamina hair for Pet, clustered lamina hair for Pub, no hairs on the lamina for Rob, as described in Kissling 1977) as training data. After performing the LDA with standardized and centered values for all parameters (except hair types used for defining the groups), the linear discriminant function was used to predict the entire dataset. In contrast to the FAMD, LDA calculates posterior probabilities of each tree to belong to each of the groups. We used two different probability thresholds to assign trees to a group. First, we applied the “majority rule” where each tree is assigned to the taxon with the highest assignment probability. Second, we used a probability threshold of 0.8. Trees above this threshold were assigned to the respective taxon. All others were tagged as “unclassified”; they represent intermediate, possibly hybrid trees.

2.5 Model-based clustering using genetic data

We used STRUCTURE 2.3.3 (Pritchard et al. 2000) and STRUCTURE HARVESTER 0.6.93 (Earl and vonHoldt 2012) to group individuals into genetic clusters based on genetic data only. We ran ten simulations with different seeds for K (number of clusters) from 1 to 20, using 1,000,000 repetitions after a burn-in period of 100,000 runs, admixture model, correlated allele frequencies, and no prior location information. Assignment probabilities to the clusters were calculated with CLUMPP 1.1.1 (Jakobsson and Rosenberg 2007) using the Greedy algorithm. We then had a closer look at the assignment probabilities for $K=3$ (i.e., the number of species involved). Every tree was assigned to a cluster according to the majority rule and the clusters were assigned to taxa based on hair types.

2.6 Comparison among different variable sets and approaches

We tested—for both the FAMD/HCPC and LDA (majority rule)—which variable set (morphological variables only, genetic variables only, all variables) best separates among the

different clusters. We used the separation/cohesion ratio (described in, e.g., Janert 2010) as quality criterion. A high separation/cohesion ratio indicates that the distance between clusters is large and/or the variation within clusters is small, leading to a good separation of the clusters. Separation (among clusters) and cohesion (within clusters) were calculated after Davies and Bouldin (1979) using distances calculated from the first two dimensions of the multivariate analyses. Separation was calculated as the average distance between the centroids of each cluster. Cohesion was assessed by using the average distance (weighted by cluster size) of each point to the centroid of its cluster. Finally, we calculated how well all six multivariate and the STRUCTURE approach corresponded. To do so, we compared the clustering obtained from the software STRUCTURE ($K=3$) with the grouping/clustering of the FAMD/HCPC and LDA. For both STRUCTURE and LDA, cluster assignment was done according to the majority rule.

3 Results

3.1 Comparison among different variable sets and approaches

HCPC suggested three clusters for all three variable sets. Species assignment of these clusters was done according to hair types (see below). For the FAMD, the joint application of both morphological and genetic parameters (Fig. 1) led to the best separation of the clusters (Table 1). The average distance among clusters (separation) was almost four times

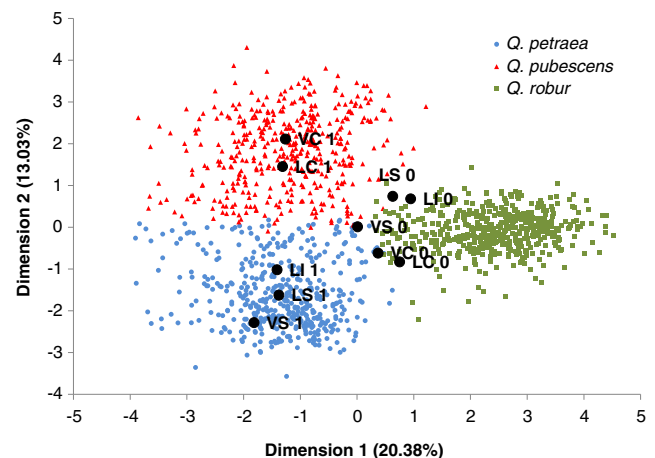


Fig. 1 Taxon assignment of 1369 *Quercus* individuals using factor analysis of mixed data (FAMD) and subsequent hierarchical clustering of principal components (HCPC). This multivariate procedure combines 13 morphological and six synthetic genetic variables. Shown is the position of each tree along the first two dimensions of FAMD, with the clustering of HCPC in different colors and symbols. Species assignment of the three proposed clusters was done according to hair types. Factor coordinates of the presence (1) or absence (0) of hair types (V = on the vein, L = on the lamina, C = clustered, S = stellate, I = intermediate, see also Table 3) are also shown

Table 1 Separation of the clusters (representing oak species) using two multivariate approaches with three variable sets (all, only morphological, only genetic)

Approach	Variable set	<i>n</i> (Pet)	<i>n</i> (Pub)	<i>n</i> (Rob)	Cohesion within clusters	Separation among clusters	Separation/cohesion ratio
FAMD	All variables	466	428	475	1.01	3.87	3.84
	Morphological variables only	431	437	501	1.01	3.47	3.43
	Genetic variables only	510	439	420	1.10	1.19	1.09
LDA	All variables	430	471	468	0.90	2.99	3.33
	Morphological variables only	388	450	531	0.97	2.38	2.45
	Genetic variables only	419	476	474	0.90	2.80	3.13

n = number of trees assigned to the different species (*Pet* = *Quercus petraea*, *Pub* = *Q. pubescens*, *Rob* = *Q. robur*) based on HCPC (hierarchical clustering of principal components) for FAMD or highest probability (majority rule) for LDA. Cohesion represents the weighted average distance within clusters. Separation is the average distance among clusters. A large separation/cohesion ratio indicates good separation among species. Distances are based on the first two axes of the multivariate analyses

FAMD factor analysis of mixed data, LDA linear discriminant analysis

higher than the average distance among points within the clusters (cohesion). FAMD/HCPC based on morphological data alone (Fig. S1a) also showed a high separation/cohesion ratio, while the analysis based on genetic data alone (Fig. S1b) showed the lowest values. The separation using genetic data only is, however, markedly improved when looking at the first two principal coordinates of the initial PCoA (Fig. S2). Separation/cohesion ratios of the LDA were mostly lower than for the FAMD/HCPC (Table 1). Also here, the joint use of both leaf morphology and genetic data led to the highest separation of the clusters (Fig. 2a), followed by the LDA using genetic data only (Fig. S3b) and morphological variables only (Fig. S3a).

The overlap in species assignment using different variable sets and statistical approaches varied from 66 to 95 % (Table 2). For all comparisons, *Q. robur* reached the highest congruence in all three approaches (results not shown). For example, FAMD/HCPC or LDA using all variables led to the same species assignment in 85 % of the cases. However, 96 % of all *Q. robur* trees (assigned by FAMD/HCPC) were also identified as the same species by LDA (compared to 83 % for *Q. pubescens* and 74 % for *Q. petraea*). The differences in species assignment therefore derive mainly from the classification of *Q. petraea* and *Q. pubescens*. STRUCTURE had its best congruency with the multivariate approaches that used all variables (overall 87 and 85 % for FAMD/HCPC and LDA, respectively). The following two sections will concentrate on this combined variable set because for both multivariate approaches, a joint application of morphological and genetic variables proved to deliver the best results in terms of cluster separation.

3.2 Factorial analysis of mixed data and hierarchical clustering on principal components

After performing the FAMD with all variables, we used the first ten dimensions, which accounted for 78.6 % of the total

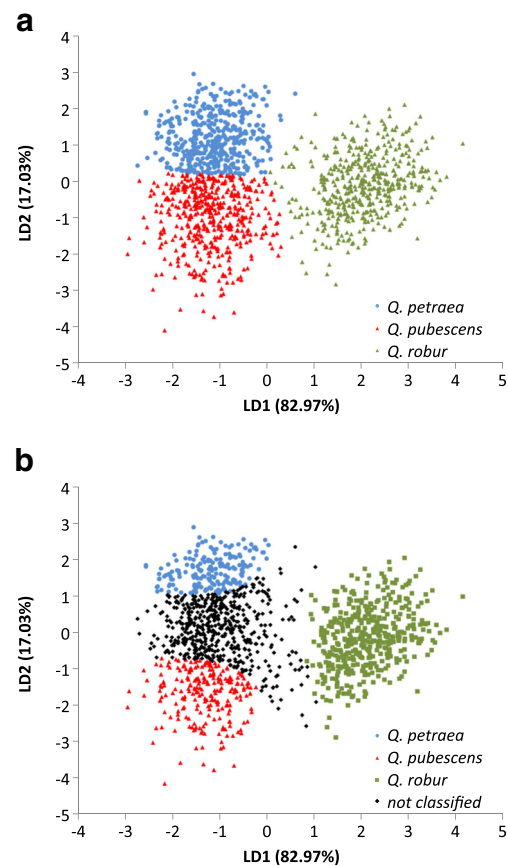


Fig. 2 Taxon assignment of 1369 *Quercus* individuals using linear discriminant analysis (LDA) with different thresholds of species assignment probability. This multivariate approach combines eight morphological and six synthetic genetic variables. Unambiguous species-specific hair types were used to predefine the groups of the training dataset. Shown are the position and classification of each tree along the two discriminant functions, using **a** highest probability (majority rule) and **b** 80 % as probability threshold for species assignment.

Table 2 Proportion of agreement in species assignment of 1369 *Quercus* individuals using seven different analytical approaches

No.	Approach	2	3	4	5	6	7
1	FAMD/HCPC (all variables)	0.95	0.78	0.85	0.75	0.81	0.87
2	FAMD/HCPC (morphological variables only)		0.74	0.80	0.73	0.76	0.82
3	FAMD/HCPC (genetic variables only)			0.85	0.66	0.88	0.79
4	LDA (all variables)				0.78	0.90	0.85
5	LDA (morphological variables only)					0.69	0.71
6	LDA (genetic variables only)						0.84
7	STRUCTURE (genetic data only)						

Shown is the congruence among two multivariate analyses (FAMD/HCPC = factor analysis of mixed data with subsequent hierarchical clustering of principal components; LDA = linear discriminant analysis) using three different sets of variables (all, only morphological, only genetic) and a model-based approach (STRUCTURE with $K=3$) using only genetic data

variation, for the clustering (HCPC). The three clusters suggested by HCPC contained relatively equal numbers of trees (475 for cluster 1, 428 for cluster 2, and 466 for cluster 3). All five hair parameters were significantly different among the clusters, with presence/absence of clustered and stellate hair on the lamina having the lowest p values (Table S2). Therefore, clusters were assigned to species according to hair type (Table 3). Intermediate hairs were frequent in *Q. petraea*, but also occurred in 30 % of *Q. pubescens*. Additionally, all 14 continuous variables significantly differed among clusters defined by the HCPC (Table S2).

We then used the clustering of the HCPC described above to interpret the FAMD on the total dataset. The first two dimensions of the FAMD accounted for 33.4 % of the variation (Fig. 1). Dimension 1 of the FAMD clearly separated *Q. robur* from the two other species (Figs. 1 and S4) and was mainly influenced by the nSSR data (PCo1), petiole ratio (PR), and venation (RPV, Table 3). In general, *Q. robur* had shorter petioles and a higher proportion of intercalary veins (Fig. S5). Also, other parameters clearly distinguished between *Q. robur* and the rest, like hair parameters, lobe depth ratio (LDR), basal shape (BS), and lobe number ratio (LLNR). Dimension 2 separated all three species (Figs. 1 and S4), with *Q. pubescens* and *Q. petraea* on the extremes of the axis and *Q. robur* in the middle. Hair parameters had the highest contribution (Table 3), followed by PCo2 (and also PCo3 and PCo4), lobe depth ratio at the widest lobe (LDRW), and lobe number ratio (LLNR). *Q. pubescens* exhibited higher sinus depth at the widest width and a higher number of lobes per leaf compared to *Q. petraea* (Fig. S5), but variation within taxa was large. Overall, 35 of 71 populations represented pure populations, and in only 11 populations, the frequency of the most abundant species was below 80 % (Fig. 3).

3.3 Linear discriminant analysis

For the LDA with all variables, we used a training dataset of 1286 a priori grouped trees (346 *Q. petraea*, 414

Q. pubescens, and 526 *Q. robur*) with unambiguous hair types and produced two linear discriminant functions that accounted for 83.0 and 17.0 % of the variation. LD1 split the dataset into *Q. robur* and the rest (Fig. S6) and was mainly influenced by (in this order) PCo1, RPV, PR, PCo4, and PCo5 (Table 3). LD2 represented a gradient between *Q. petraea* and *Q. pubescens* (while *Q. robur* had intermediate values, Fig. S6); the most important variables were PCo2, PCo3, PCo4, LDRW, and LLNR (in decreasing order). Thus, the importance, average, and range of morphological and genetic variables characterizing the three groups (Table 3) were very similar to the ones described for the FAMD. Applying the LD functions on the training dataset led to grouping confirmation of 79.3 % of the trees using the majority rule. We then used the two LD functions to predict the whole dataset, resulting in 430 *Q. petraea*, 471 *Q. pubescens*, and 468 *Q. robur* trees with the majority rule (Figs. 2a and 4). Only 1 of the 83 trees with mixed hair types was assigned to *Q. robur*, and the rest was equally assigned to the other two species.

In general, the maximum probability to belong to a certain species was much higher in *Q. robur* than in the other two species (Fig. S7). Consequently, with the higher probability threshold of 0.8 for species assignment, the number of *Q. petraea* and *Q. pubescens* trees was substantially smaller, while the number of *Q. robur* trees remained more or less the same. The proportion of unclassified trees was 40.0 %; these intermediate forms mostly and equally derived from the original *Q. petraea* and *Q. pubescens* clouds, dissolving this taxon cluster (Fig. 2b). With a probability threshold of 80 %, only 11 populations consisted of only one species (all *Q. robur* populations), and the maximum proportion of intermediates in a population was as high as 73.7 % (Fig. S8).

3.4 STRUCTURE

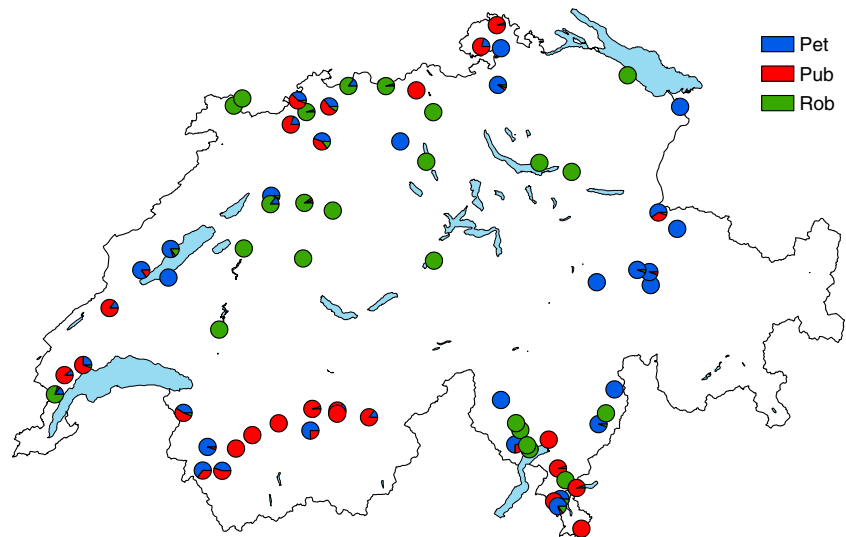
STRUCTURE returned the highest $L(K)$, which is the model choice criterion proposed by Pritchard et al. (2000), with

Table 3 Results of the multivariate analyses (using both morphological and genetic variables) describing the three *Quercus* species studied

Variables	Abbr.	FAMD/HPCP						LDA									
		Contribution to FAMD		HPCP cluster 1		HPCP cluster 2		HPCP cluster 3		Standardized coefficient		LDA group 1		LDA group 2		LDA group 3	
		Dim. 1	Dim. 2	<i>Q. robur</i>	<i>Q. pubescens</i>	<i>Q. petraea</i>	LD1	LD2	<i>Q. robur</i>	<i>Q. pubescens</i>	<i>Q. petraea</i>	LDA group 1	LDA group 2	LDA group 3			
Morphological																	
Basal shape	BS	7.27	0.00	5.32 ± 1.9	3.77 ± 1.53	3.7 ± 1.44	0.11	-0.05	5.31 ± 1.43	3.73 ± 1.54	3.76 ± 1.91						
Lamina shape/obversity	OB	3.57	0.22	62.98 ± 7.41	57.18 ± 7.57	58.23 ± 7.03	0.09	-0.03	63.15 ± 6.95	57.05 ± 7.49	58.37 ± 7.4						
Petiole ratio	PR	16.81	0.00	6.2 ± 2.33	13.18 ± 3.64	13.84 ± 3.89	-0.27	0.31	6.12 ± 3.79	13.4 ± 3.75	13.62 ± 2.27						
Lobe depth ratio	LDR	7.35	0.05	73.35 ± 10.16	65.67 ± 9.62	65.72 ± 9.89	0.05	0.26	73.13 ± 9.8	65.18 ± 9.79	66.63 ± 10.26						
Lobe depth ratio at widest sinus	LDRW	0.63	6.52	37.96 ± 8.55	39.51 ± 9.95	33.61 ± 7.53	-0.08	-0.37	37.91 ± 7.06	39.83 ± 9.67	32.8 ± 8.62						
Lobe number ratio	LLNR	4.93	2.17	0.45 ± 0.34	0.76 ± 0.39	0.59 ± 0.42	0.15	-0.35	0.47 ± 0.34	0.8 ± 0.43	0.5 ± 0.36						
Percentage venation	RPV	16.12	0.00	6.54 ± 1.63	3.33 ± 1.97	3.25 ± 1.64	0.41	-0.02	6.57 ± 1.55	3.22 ± 1.98	3.39 ± 1.63						
Lobe width ratio	LWR	0.08	0.31	30.02 ± 4.38	31.57 ± 4.84	30.9 ± 4.41	-0.05	-0.14	29.93 ± 4.35	31.4 ± 4.8	31.1 ± 4.42						
Presence of stellate hair on lamina	LS	5.78	19.65	0.01	0.07	0.84	Used for grouping	0.03	0.27	0.67							
Presence of stellate hair on vein	VS	0.11	0.44	0.00	0.00	0.02	Not used	0.00	0.00	0.01							
Presence of clustered hair on lamina	LC	6.56	19.65	0.01	0.96	0.18	Used for grouping	0.01	0.77	0.30							
Presence of clustered hair on vein	VC	3.12	21.30	0.00	0.70	0.03	Not used	0.01	0.53	0.13							
Presence of intermediate hair on lamina	LI	8.89	11.30	0.03	0.30	0.88	Not used	0.04	0.47	0.72							
Genetic																	
PCoA component 1	PCo1	17.58	0.66	0.17 ± 0.09	-0.06 ± 0.06	-0.11 ± 0.07	1.04	-0.13	0.18 ± 0.06	-0.06 ± 0.06	-0.12 ± 0.08						
PCoA component 2	PCo2	0.10	9.89	0.01 ± 0.08	-0.05 ± 0.09	0.04 ± 0.09	0.18	0.70	0.01 ± 0.07	-0.06 ± 0.08	0.06 ± 0.08						
PCoA component 3	PCo3	0.02	3.25	-0.01 ± 0.07	0.03 ± 0.09	-0.02 ± 0.09	-0.04	-0.41	0 ± 0.09	0.03 ± 0.09	-0.03 ± 0.07						
PCoA component 4	PCo4	0.45	3.10	-0.01 ± 0.07	0.03 ± 0.09	-0.01 ± 0.09	-0.24	-0.39	-0.01 ± 0.09	0.03 ± 0.09	-0.02 ± 0.07						
PCoA component 5	PCo5	0.21	0.58	0.01 ± 0.08	-0.01 ± 0.08	0.01 ± 0.08	0.16	0.21	0.01 ± 0.08	-0.01 ± 0.08	0 ± 0.08						
PCoA component 6	PCo6	0.42	0.91	-0.01 ± 0.09	0.02 ± 0.07	0.00 ± 0.07	-0.13	-0.23	-0.01 ± 0.07	0.02 ± 0.07	-0.01 ± 0.09						

Listed are the variables and their abbreviations; the contribution of each variable to the first two dimensions of the factor analysis of mixed data (FAMD); averages (± standard deviation) of the variables in the three clusters defined by hierarchical clustering of principal components (HPCP; in case of hair types, the proportion is given); contribution of each variable (standardized coefficient) to the first two linear discriminants of the linear discriminant analysis (LDA); averages (± standard deviation) of the variables in the three groups of the LDA (in case of hair types, the proportion is given). Values of LLNR are log-transformed; RPV is square root-transformed

Fig. 3 Species composition of the 71 investigated populations of *Quercus* in Switzerland using factor analysis of mixed data (FAMD) and subsequent hierarchical clustering of principal components (HCPC). Species assignment of the clusters was done according to hair types (Fig. 1). *Pet* = *Q. petraea*, *Pub* = *Q. pubescens*, *Rob* = *Q. robur*



$K=8$. However, the increase in likelihood for $K>3$ was only minimal (Fig. S9). In this study, we were mainly interested in $K=3$ to compare the population genetic clustering to the groups and clusters in the multivariate analyses. To assign the three clusters from STRUCTURE to species, we looked at the lamina hair types occurring in the different clusters; 85.9 % of the trees in cluster 1 had clustered hair on the lamina and were therefore assigned to *Q. pubescens*. Most of the trees in cluster 2 (90.3 %) were hairless and therefore assigned to *Q. robur*. In cluster 3, 68.4 % of the trees had stellate hair on the lamina, whereas 27.5 % had clustered hair. This cluster was assigned to *Q. petraea*. Across the study range, 15 of the 71 populations were pure populations (3 Pet, 2 Pub, and 10 Rob), and in 52 populations, the frequency of the most abundant taxon was at least 80 % (Fig. 5). In general, posterior probabilities to belong to a specific cluster were high, especially in *Q. robur* (Fig. S10).

4 Discussion

Scientists and practitioners often rely on rapid and accurate species identification. In some genera, however, species assignment is not so straightforward, because there might be a lack of reliable and unambiguous morphological characters and because hybridization among species might lead to intermediate forms or a combination of parental characters. In these cases, species identification cannot be reliably done in the field, and additional effort is needed. Popular approaches to resolve such complex patterns are the use of multivariate statistics based on many morphological characters or the use of genetic markers. However, the joint application of these two approaches, which might result in the highest resolution in taxon identification, is rarely found in the literature. In the present study, we tested whether joint multivariate analyses (FAMD/HCPC and LDA) of leaf morphological characters

Fig. 4 Species composition of the 71 investigated populations of *Quercus* in Switzerland using linear discriminant analysis (LDA). Species assignment was done according to the highest assignment probability (majority rule). *Pet* = *Q. petraea*, *Pub* = *Q. pubescens*, *Rob* = *Q. robur*

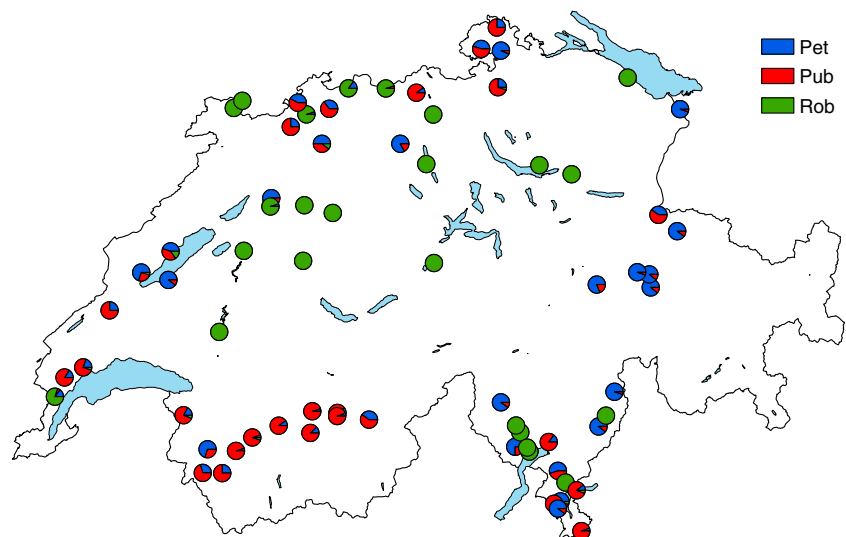
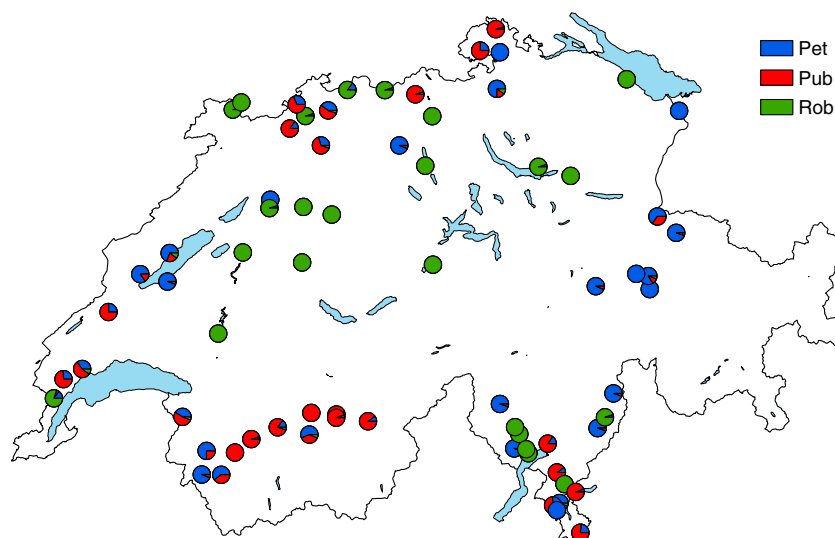


Fig. 5 Species composition of the 71 investigated populations of *Quercus* in Switzerland using STRUCTURE (Pritchard et al. 2000). Species assignment of the $K=3$ clusters was done according to hair types. *Pet* = *Q. petraea*, *Pub* = *Q. pubescens*, *Rob* = *Q. robur*



and genetic markers improve the species assignment of the three morphologically similar and hybridizing oak species *Q. petraea*, *Q. pubescens*, and *Q. robur* compared to the same approaches using only morphological or genetic variables. We then compared the results to a model-based clustering method (STRUCTURE, Pritchard et al. 2000) that only takes genetic information into account.

For both the FAMD/HCPC and LDA, the combination of leaf morphological and genetic parameters led to the highest degree of separation among the three species (Table 1). This result suggests that joint multivariate analyses may improve taxon delimitation in closely related species such as the white oaks. Notably, *Q. robur* is more different from the two other species than *Q. petraea* from *Q. pubescens* (Figs. 1 and 2). It has been shown that *Q. robur* is morphologically more distinct from *Q. petraea* and *Q. pubescens* than are the latter two species (e.g., Curtu et al. 2007a; Dupouey and Badeau 1993). However, there might be more reasons for the challenging discrimination between *Q. petraea* and *Q. pubescens*. First, the genetic markers used in this study were designed to maximize differentiation between *Q. robur* and *Q. petraea* (Guichoux et al. 2011, see also Fig. S2). Nevertheless, a part of these markers has previously shown discrimination power for all three taxa (Gugerli et al. 2008). Second, hybrids between *Q. pubescens* and *Q. petraea* might be more common than hybrids between *Q. robur* and the other two species, at least within the study range. Several studies favor this hypothesis (e.g., Semerikov et al. 1988), but the degree and direction of hybridization in these three species, and in white oaks in general, is still under debate (e.g., Dupouey and Badeau 1993; Guichoux et al. 2013).

Our results show that several parameters can be used to reliably identify *Q. robur*. In general, *Q. robur* is hairless on the lamina and has short petioles, a high proportion of intercalary veins, a high lobe depth, an ear-like basal shape, and a

low number of lobes. Moreover, the nSSR markers clearly differentiate between *Q. robur* and the other two species. This result is in line with earlier studies (Curtu et al. 2007a; Gugerli et al. 2007; Kremer et al. 2002). The distinction between *Q. petraea* and *Q. pubescens* is more complex. Although the two species generally differ in hair type (*Q. pubescens* has mostly clustered hair on the lamina, whereas *Q. petraea* normally has stellate hair), hair types are not easy to classify due to intermediate forms. Therefore, additional parameters are needed to resolve the taxonomic assignment. Besides hair type and nSSR markers, the main difference between the two species are the lobes, with *Q. pubescens* having a higher lobe depth at the widest width and a larger number of lobes compared to *Q. petraea* (and *Q. robur*). However, overlap among species was still considerable. This means that the identification in the field on the basis of a few single leaf morphological characters, without the availability of acorns, is problematic, and genetic data as presented in this study improves taxonomic resolution.

The three analytical approaches we used differ in the way they can identify the taxa. FAMD can combine continuous and categorical variables and needs no a priori species information or training dataset. However, the subsequently applied hierarchical clustering (HCPC) does not directly allow for the recognition of intermediate forms or hybrids (this could be done a posteriori with additional methods). Species assignment of the resulting clusters has to be done with additional evidence, e.g., reference samples or previously known diagnostic parameters. Here, hair type proved to be the best predictor for the different groups. In contrast, LDA needs a training/reference dataset with individuals assigned to a specific species. Since hair presence/type was the most discriminating factor in FAMD and is known to be a good characteristic for species differentiation in general (Kissling 1977), we built the training dataset based on these characters. An

advantage of LDA is that it specifically aims at finding the best separation among the predefined groups, in contrast to FAMD that maximizes the variance without prior knowledge of groups of data. LDA yields an assignment probability for each individual, thus allowing the identification of the proportion of intermediate/hybrid forms. The threshold to belong to a certain taxon, however, is arbitrary and has substantial consequences on the outcome of the analysis (Fig. 2). Finally, STRUCTURE is a model-based approach that clusters individuals with no a priori knowledge based on genetic data and population genetic theory, also delivering assignment probabilities. In this approach, markers with alleles that are—at best—diagnostic for a taxon may reliably indicate hybrid origin of particular individuals, even when backcrossing is involved (Lepais et al. 2009). All three methods (using a majority rule for species assignment and all variables in the case of multivariate analyses) resulted in more or less congruent outcomes (overall average 85–87 %, Table 2), depending on the species (74–98 %).

The fact that 85–87 % of the sampled trees were assigned to the same species using STRUCTURE and the two multivariate analyses using both morphological and genetic variables show that a model-based test of species identification using solely genetic data may represent an alternative to morphological approaches. This is especially interesting if genetic analyses are anyway performed to answer specific study questions. Moreover, purely morphological approaches have several disadvantages, especially when intermediate forms and hybrids are present (Lopez-Caamal and Tovar-Sanchez 2014). First, morphological expression to a certain degree depends on the environment. Second, morphological intermediacy may originate from other processes than hybridization. Finally, hybrid individuals do not necessarily show intermediate phenotypes (Viscosi et al. 2009) because not all characteristics are under polygenic control with simple additive effects. Purely genetic approaches, on the other hand, can reliably identify hybrid taxa if highly differentiating markers and alleles can be established. Besides the nSSR markers applied in the present study (Guichoux et al. 2011), one possibility is to use species-diagnostic single-nucleotide polymorphisms (SNPs), for example from those identified in Lepoittevin et al. (2015) for white oaks. Using a relatively high number of SNPs would not only help in identifying species but also in quantifying levels of backcrossing and introgression. This has recently been exemplified in six stands of *Q. robur* and *Q. petraea* genotyped at 262 SNPs (Guichoux et al. 2013), showing asymmetric introgression toward *Q. petraea*.

5 Conclusions

For the taxonomic delimitation of the European white oaks species described here, varying degrees of phenotypic plasticity, hybridization, backcrossing, and introgression lead to

patterns that are only resolvable with a sufficient number of morphological and/or genetic parameters. Approaches based on pure genetic data are also promising and offer the possibility of a more accurate assignment of hybrids. In contrast to the low taxonomic resolution of classical DNA barcoding (further compromised by uniparental inheritance of the organellar-based markers applied), approaches using highly variable and abundant nuclear markers like nSSRs and SNPs can also give insights into aspects such as mating patterns, backcrossing, and levels of introgression.

Acknowledgments We thank Lorenz Walthert, Fabrizio Cioldi, Urs Beat Brändli, and Thomas Wohlgemuth for their help in sampling site selection; Luzia Oeschger, Dino Wirz, Melody Schmid, Karin Moosbrugger, Marco Gees, and Sabine Brodbeck for their support in the field; Heiko Hauser and Sabine Osterkamp from LGC for DNA extractions; the Genetic Diversity Center (GDC) of ETH Zürich for support in molecular analyses; Nina Roth for statistical advice; forest services and private landowners for sampling permissions; Lorenz Walthert, Christoph Sperisen, and Daniela Csencsics for discussion and suggestions; and two anonymous reviewers for valuable comments on a previous version of this manuscript.

Compliance with ethical standards

Funding This study was carried out in the framework of the WSL/FOEN-supported research program Forest and Climate Change.

References

- Bacilieri R, Ducouso A, Petit RJ, Kremer A (1996) Mating system and asymmetric hybridization in a mixed stand of European oaks. *Evolution* 50:900–908. doi:10.2307/2410861
- Brändli U-B (2010) Schweizerisches Landesforstinventar: Ergebnisse der dritten Erhebung 2004–2006. Eidg. Forschungsanstalt WSL, Birmensdorf
- Bruschi P, Vendramin GG, Bussotti F, Grossoni P (2000) Morphological and molecular differentiation between *Quercus petraea* (Matt.) Liebl. and *Quercus pubescens* Willd. (Fagaceae) in northern and central Italy. *Ann Bot* 85:325–333. doi:10.1006/anbo.1999.1046
- Curtu AL, Gailing O, Finkeldey R (2007a) Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. *BMC Evol Biol* 7:218. doi:10.1186/1471-2148-7-218
- Curtu AL, Gailing O, Leinemann L, Finkeldey R (2007b) Genetic variation and differentiation within a natural community of five oak species (*Quercus* spp.). *Plant Biol* 9:116–126. doi:10.1055/s-2006-924542
- Davies DL, Bouldin DW (1979) A cluster separation measure. *IEEE Trans Pattern Anal* 1:224–227. doi:10.1109/TPAMI.1979.4766909
- R Development Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org>
- Dupouey J, Badaeu V (1993) Morphological variability of oaks (*Quercus robur* L., *Quercus petraea* (Matt) Liebl., *Quercus pubescens* Willd) in northeastern France: preliminary results. *Ann Sci For* 50:35s–40s. doi:10.1051/forest:19930702
- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and

- implementing the Evanno method. *Conserv Genet Resour* 4:359–361. doi:10.1007/s12686-011-9548-7
- Fortini P, Antonecchia G, Di Marzio P, Maiuro L, Viscosi V (2015) Role of micromorphological leaf traits and molecular data in taxonomy of three sympatric white oak species and their hybrids (*Quercus* L.). *Plant Biosyst* 149:546–558. doi:10.1080/11263504.2013.868374
- Gerber S, Chadoeuf J, Gugerli F, Lascoux M, Buiteveld J (2014) High rates of gene flow by pollen and seed in oak populations across Europe. *PLoS ONE* 9:e91301. doi:10.1371/journal.pone.0091301
- Gugerli F, Walser JC, Dounavi K, Holderegger R, Finkeldey R (2007) Coincidence of small-scale spatial discontinuities in leaf morphology and nuclear microsatellite variation of *Quercus petraea* and *Q. robur* in a mixed forest. *Ann Bot* 99:713–722. doi:10.1093/aob/mcm006
- Gugerli F, Brodbeck S, Holderegger R (2008) Utility of multilocus genotypes for taxon assignment in stands of closely related European white oaks from Switzerland. *Ann Bot* 102:855–863. doi:10.1093/aob/mcn164
- 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.). *Mol Ecol Resour* 11: 578–585. doi:10.1111/j.1755-0998.2011.02983.x
- Guichoux E, Garnier-Géré P, Lagache L, Lang T, Boury C, Petit RJ (2013) Outlier loci highlight the direction of introgression in oaks. *Mol Ecol* 22:450–462. doi:10.1111/mec.12125
- Harr B, Weiss S, David JR, Brem G, Schlötterer C (1998) A microsatellite-based multilocus phylogeny of the *Drosophila melanogaster* species complex. *Curr Biol* 8:1183–1186. doi:10.1016/s0960-9822(07)00490-3
- Husson F, Josse J, Pagès J (2010) Principal component methods – hierarchical clustering – partitional clustering: why would we need to choose for visualizing data? Technical report—Agrocampus, Applied Mathematics Department, <http://www.agrocampus-ouest.fr/math/>
- 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* 23:1801–1806. doi:10.1093/bioinformatics/btm233
- Janert PK (2010) Data analysis with open source tools. O'Reilly Media, Sebastopol
- Jørgensen M-H, Carlsen T, Skrede I, Elven R (2008) Microsatellites resolve the taxonomy of the polyploid *Cardamine digitata* aggregate (Brassicaceae). *Taxon* 57:882–892
- Kissling P (1977) Les poils des quatre espèces de chênes du Jura (*Quercus pubescens*, *Q. petraea*, *Q. robur* et *Q. cerris*). *Ber Schweiz Bot Ges* 87:1–18
- Kremer A et al (2002) Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Ann For Sci* 59:777–787. doi:10.1051/forest:2002065
- Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *J Stat Softw* 25:1–18
- Lepais O, Gerber S (2011) Reproductive patterns shape introgression dynamics and species succession within the European white oak species complex. *Evolution* 65:156–170. doi:10.1111/j.1558-5646.2010.01101.x
- Lepais O, Petit RJ, Guichoux E, Lavabre JE, Alberto F, Kremer A, Gerber S (2009) Species relative abundance and direction of introgression in oaks. *Mol Ecol* 18:2228–2242. doi:10.1111/j.1365-294X.2009.04137.x
- Lepoittevin C et al (2015) Single-nucleotide polymorphism discovery and high-density SNP array development for genetic analysis in European white oaks. *Mol Ecol Resour* 15:1446–1459. doi:10.1111/1755-0998.12407
- Lopez-Caamal A, Tovar-Sanchez E (2014) Genetic, morphological, and chemical patterns of plant hybridization. *Rev Chil Hist Nat* 87:16. doi:10.1186/s40693-014-0016-0
- Mallet J (2005) Hybridization as an invasion of the genome. *Trends Ecol Evol* 20:229–237. doi:10.1016/j.tree.2005.02.010
- Mátyás G, Sperisen C (2001) Chloroplast DNA polymorphisms provide evidence for postglacial re-colonisation of oaks (*Quercus* spp.) across the Swiss Alps. *Theor Appl Genet* 102:12–20. doi:10.1007/s001220051613
- Mátyás G, Bonfils P, Sperisen C (2002) Autochthon oder allochthon? Ein molekulargenetischer Ansatz am Beispiel der Eichen (*Quercus* spp.) in der Schweiz. *Schweiz Z Forstwes* 153:91–96
- Muir G, Fleming CC, Schlötterer C (2000) Species status of hybridizing oaks. *Nature* 405:1016. doi:10.1038/35016640
- Neophytou C, Gärtner SM, Vargas-Gaete R, Michiels H-G (2015) Genetic variation of Central European oaks: shaped by evolutionary factors and human intervention? *Tree Genet Genomes* 11:79. doi:10.1007/s11295-015-0905-7
- Pagès J, Camiz S (2008) Analyse factorielle multiple de données mixtes: application à la comparaison de deux codages. *La Revue de Modulad* 38:178–183
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539. doi:10.1093/bioinformatics/bts460
- Petrusek A, Hobaek A, Nilssen JP, Skage M, Cerny M, Brede N, Schwenk K (2008) A taxonomic reappraisal of the European *Daphnia longispina* complex (Crustacea, Cladocera, Anomopoda). *Zool Scr* 37:507–519. doi:10.1111/j.1463-6409.2008.00336.x
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rellstab C, Keller B, Girardclos S, Anselmetti FS, Spaak P (2011) Anthropogenic eutrophication shapes the past and present taxonomic composition of hybridizing *Daphnia* in unproductive lakes. *Limnol Oceanogr* 56:292–302. doi:10.4319/lo.2011.56.1.0292
- Ross-Davis A, Huang Z, McKenna J, Ostry M, Woeste K (2008) Morphological and molecular methods to identify butternut (*Juglans cinerea*) and butternut hybrids: relevance to butternut conservation. *Tree Physiol* 28:1127–1133. doi:10.1093/treephys/28.7.1127
- Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour* 8: 103–106. doi:10.1111/j.1471-8286.2007.01931.x
- Routtu J, Hoikkala A, Kankare M (2007) Microsatellite-based species identification method for *Drosophila virilis* group species. *Hereditas* 144:213–221. doi:10.1111/j.2007.0018-0661.02021.x
- Rushton B (1983) An analysis of variation of leaf characters in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. population samples from Northern Ireland. *Irish For* 40:52–77
- Semerikov L, Glotov N, Zhivotovskii L (1988) Example of effectiveness of analysis of the generalized variance of traits in trees. *Sov J Ecol* 18:140–143
- Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82:561–573. doi:10.1038/sj.hdy.6885180
- Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. *Trends Ecol Evol* 24:110–117. doi:10.1016/j.tree.2008.09.011
- Venables WN, Ripley BD (2002) *Modern applied statistics with S*. Springer, New York
- Viscosi V, Lepais O, Gerber S, Fortini P (2009) Leaf morphological analyses in four European oak species (*Quercus*) and their hybrids: a comparison of traditional and geometric morphometric methods. *Plant Biosyst* 143:564–574. doi:10.1080/11263500902723129
- Walther L, Pannatier EG, Meier ES (2013) Shortage of nutrients and excess of toxic elements in soils limit the distribution of soil-sensitive tree species in temperate forests. *For Ecol Manag* 297: 94–107. doi:10.1016/j.foreco.2013.02.008
- Whitney KD, Ahern JR, Campbell LG, Albert LP, King MS (2010) Patterns of hybridization in plants. *Perspect Plant Ecol Evol Syst* 12:175–182. doi:10.1016/j.ppees.2010.02.002