Tree Genetics & Genomes (2025) 21:11
https://doi.org/10.1007/s11295-025-01695-9

ORIGINAL ARTICLE

Hybridization, spatial genetic structure and potential environmental preadaptation in *Quercus robur* and *Quercus petraea* in Germany–results from the 4th National Forest Inventory

Birte Pakull[1] · Bernd Degen[1] · Hilke Schroeder[1] · Thomas Riedel[2] · Malte Mader[1] · Heike Liesebach[1] · Petra Hoffmann[1] · Susanne Hoppe[1] · Pascal Eusemann[1]

Received: 8 October 2024 / Revised: 14 January 2025 / Accepted: 21 February 2025

Abstract
In the course of the 4th German National Forest Inventory, samples of the two oak species *Quercus robur* and *Quercus petraea* were collected throughout Germany. Genetic analyses were performed based on 403 nuclear, 21 chloroplast and 7 mitochondrial markers. The analyses showed good differentiation between the two species based on genetic data. Both species are connected through hybridization and introgression, but only about 2% of the samples analysed were found to be potential first-generation hybrids. Identical chloroplast and mitochondrial haplotypes with lineage specific distribution patterns were identified in both species. Different haplo- and mitotypes showed a tight linkage. Analysis of nuclear SNPs revealed a clear genetic structure in *Q. robur*, which appears to be largely of natural origin and can be explained by the postglacial recolonization routes through which the species dispersed throughout Germany after the last glacial maximum. Environmental influences, most importantly continentality, also appear to have an impact on the genetic structure of *Q. robur*, possibly caused by preadaptation within the refugial source-populations. For *Q. petraea*, the situation seems more complicated and no clear genetic structure could be identified.

Keywords National Forest Inventory · Oak · SNP marker · Spatial genetic structure · Haplotypes

Introduction

Quercus robur and *Quercus petraea* in Germany

Alongside European beech (*Fagus sylvatica* L.), the two closely related oak species sessile oak (*Quercus petraea* (Matt.) Liebl.) and pedunculate oak (*Quercus robur* L.) are the most common broadleaf tree species in Germany. According to data from the German National Forest Inventory (NFI) 2022 (data available at https://bwi.info/?lang=en), oaks cover about 11.5% of the forest area in Germany. The natural range of both species covers more or less the whole of Germany. Both species differ, though, in their ecological requirements. *Q. robur* prefers heavier, humid soils in the lowland and tolerates periodic flooding, while *Q. petraea* is more drought tolerant and grows preferably on light and well-drained, often rocky, soils (Eaton et al. 2016). However, co-occurrence of both species at the same site is frequent (Eaton et al. 2016; Kleinschmit et al. 1995).

Species discrimination and hybridization/introgression between both species

Differences in, e.g., leaf and acorn morphology allow discrimination between *Q. robur* and *Q. petraea* (Albrecht 2014; Kremer et al. 2002; Neophytou et al. 2014). But high intra-specific morphological variation and potential hybridization/introgression between both species impedes morphological species discrimination, which lead to long-lasting debates about the levels of hybridization, and even

Communicated by J. Chen

✉ Birte Pakull
birte.pakull@thuenen.de

[1] Thünen Institute of Forest Genetics, Sieker Landstr. 2, 22927 Grosshansdorf, Germany

[2] Thünen Institute of Forest Ecosystems, Alfred-Möller-Straße 1, 16225 Eberswalde, Germany

Published online: 18 March 2025

raised doubts about the separation of both species (Kremer et al. 2002; Neophytou et al. 2014).

Although Abadie et al. (2012) reported post-mating reproductive barriers between *Q. robur* and *Q. petraea*, artificial crosses between both species (Kleinschmit et al. 1995; Steinhoff 1998) and backcrosses of hybrids with pure species (Olrik and Kjaer 2007) are possible and hybrid formation can also be found when analysing offspring of wild populations (Curtu et al. 2007, 2009; Gerber et al. 2014; Jensen et al. 2009; Streiff et al. 1999). Identified hybrids show a range of phenotypes with parental and intermediate character, and morphological determination of hybrids is difficult (Kremer et al. 2002; Viscosi et al. 2009).

Species discrimination between *Q. robur* and *Q. petraea* based on genetic markers is quite reliable. Guichoux et al. (2011) used a combination of EST-SSRs and genomic SSRs for species discrimination. Using a set of 58 SNPs from coding regions Reutimann et al. (2020) were able to genetically differentiate between *Q. robur* and *Q. petraea* and detect admixture between both species. Also, a set of 431 nuclear, chloroplast and mitochondrial SNP markers published by Degen et al. (2021a) was able to discriminate between the species and identify admixed individuals. Schroeder and Kersten (2023) published a set of six species-specific SNPs and Indel-variants which allow reliable differentiation between the two species and their hybrids. Kremer et al. (2024) analysed whole genome DNA data and identified a set of 38 diagnostic markers that can be used to distinguish between *Q. robur* and *Q. petraea* as well as *Q. pubescens* and *Q. pyrenaica*, two species with a more southerly distribution range.

Species history and spatial genetic structure in Europe

According to Leroy et al. (2017), the species history of *Q. robur* and *Q. petraea* in Europe is probably characterized by a long lasting period of genetic isolation followed by a recent period of secondary contact during the recolonization of Northern Europe after the last glacial maximum. During the last glacial maximum, oaks survived in isolated refugia on the Iberian and Italian Peninsula, as well as the Balkans. Postglacial recolonization started from these refugia and the different routes of recolonization can be reconstructed by combining the geographic genetic pattern of different maternally inherited chloroplast haplotypes, which have been formed within the genetically isolated refugia, with fossil pollen data (Petit et al. 2002a, b). During the recolonization, *Q. robur* potentially functioned as a pioneer species, and hybridization and backcrossing with *Q. robur* allowed *Q. petraea* to adapt to cooler and wetter climate conditions and migrate together with *Q. robur*. The genetic contact between the two species allowed the sharing of chloroplast haplotypes between both species (Leroy et al. 2020a; Petit et al. 2004).

In Germany, several lineages of chloroplast haplotypes from all three major refugia meet and form a diverse spatial structure of haplotypes. Haplotypes originating from the Iberian Peninsula are found predominantly in the north-western part of Germany. Haplotype 1 from the Italian Peninsula can predominantly be found in the north-west and west. Haplotypes from the Balkans show a more complex pattern and are predominantly distributed in central or central and southern parts of Germany (see figures in Petit et al. (2002a). The chloroplast and mitochondrial SNP haplotypes identified for *Q. robur* by Degen et al. (2021b) also show a spatial structure clearly separating the north-western part from the other parts of Germany. This separation can also be identified within the spatial resolution of a STRUCTURE analysis of the nuclear SNPs analysed by Degen et al. (2021b).

The German National Forest Inventory

Since 1987 Germany has been collecting data about its forests within the German National Forest Inventory (NFI). After 1987, 2002 and 2012, the 4th German National Forest Inventory was implemented in 2022. The NFI is based on a systematic sampling grid of at least 4 × 4 km that covers the whole of Germany (in some federal states the grid density is doubled or quadrupled). Inventory teams visit permanent sampling tracts located at the crossing points of the sampling grid and collect data about, e.g., forest structure, tree species, regeneration as well as diameter and height of selected sample trees (Riedel et al. 2020). Information about the history and design of the NFI can be found in Kändler (2006) and Kleinn et al. (2020). Results of the NFI are available at https://www.bundeswaldinventur.de/en/ and https://bwi.info/?lang=en. In 2022, the collection of plant material (leaves, needles or buds) for DNA analysis has been carried out for the first time within the NFI. For *Q. robur* and *Q. petraea*, 300 sampling tracts per species were selected. The selected tracts were distributed as evenly as possible across Germany. This collection of samples constitutes the largest, most comprehensive and spatially most representative genetic survey of sessile and pedunculate oak in Germany yet assembled.

Objectives of this study

Germany represents a core region of the shared distribution area of *Q. petraea* and *Q. robur* and a region, where several different postglacial recolonization routes converge (Petit et al. 2002a), where hybridization and introgression between

the two species played a key role in the postglacial recolonization process (Leroy et al. 2017, 2020a), and where environmental conditions might have shaped the recolonization process (Leroy et al. 2020a). For these reasons, the region is a prime candidate to study how recolonization, hybridization, introgression, and environmental conditions affected the genetic composition of current central European oak forests. Additionally, the region is characterized by long-lasting intensive human interference (alteration of natural population structures by extraction of timber and potentially large-distance seed transfer (Leroy et al. 2020b), potentially disturbing and diluting any natural genetic structure.

Knowing the extent of current genetic structures and understanding their origin is a prerequisite to conserving any remaining natural structure and the sustainable management of present genetic resources in the two oak species. Taking advantage of the systematic, area-representative approach implemented in the NFI and using the exhaustive dataset, we analyse:

1. the degree of hybridization and introgression,
2. interregional spatial genetic structure and genetic diversity, and.
3. signs of adaptation to local climatic conditions of *Q. robur* and *Q. petraea* throughout their distribution area within Germany.

Materials and methods

Sampling

Based on the results of the previous NFI in 2012, tracts potentially including a sufficient number of *Q. robur/Q. petraea* individuals were identified ($N = 3246$ tracts for *Q. robur* and $N = 2290$ tracts for *Q. petraea*). Then, for each of the two species, a total of 300 sampling tracts, covering the whole of Germany as evenly as possibly, were preselected for DNA sampling. A set of additional sampling tracts were selected as substitutional sampling tracts in cases where sampling of a preselected sampling tract was impossible (e.g., insufficient number of oak trees suitable for sampling). Supp. Figure 1 shows a map of the NFI tracts used for oak sampling. Each sampling tract is defined as a 150×150 m quadrangle. Five different oak individuals were sampled in a radius of about 25 m of one of the corners of the sampling tract by local NFI teams. The aim was to sample leaves (in winter: buds on twigs) of either 5 *Q. robur* or 5 *Q. petraea* individuals. However, due to the possibility of co-occurrence of both species in the same tract in combination with difficult species determination in the field (see Introduction), tracts erroneously containing a mix of both species within the 5 samples could not be ruled out completely.

Sample numbers, tract numbers and geographic coordinates were recorded. Coordinates of the NFI tracts based on the Pan-European INSPIRE reference grid are available at https://bwi.info/Download/de/BWI-Basisdaten/ACCESS2003/. Sampled trees were selected based on accessibility of leaves (or buds) from the ground, leading to the sampling of trees of different age classes. For each sampled tree height class of above or below 4 m has been recorded. Sampled leaves or buds were collected in zip-lock bags containing a 30 g silica gel pad and sent to the Thünen-Institute of Forest Genetics for DNA extraction.

A total of 3172 oak samples (*Q. robur* and *Q. petraea*) from 606 tracts have been collected. The 3172 trees sampled include 58 individuals from 12 tracts which have been sampled twice for quality assurance. Quality assurance of the NFI is carried out by an inventory inspection (Kändler 2006). A minimum of 5% of the tracts are revisited by inspection teams. If the tract was selected for DNA sampling, the inspection team tried to identify the sampled trees and resampled control samples, if possible. The concordance between original and control samples was then verified using SSR marker analysis (see 2.2). Only one of the duplicate individuals was used for further analysis. This reduces the total number of oak individuals sampled to 3114. Out of these 3114 individuals 1916 individuals had a height over 4 m (61,5%), 1137 had a height below 4 m (36.5%). Height specification is missing for 61 trees. For a total of 341 tracts, information was available on whether the associated forest stand originated from natural regeneration (283 tracts) or planting (58 tracts).

DNA extraction and control sample analysis

DNA extraction was carried out based on a tree-adjusted protocol described by Bruegmann et al. (2022). In case of expected sample duplication due to control sampling, individuals with potentially identical genotypes were identified by direct comparison of genotypes using a multiplex of seven SSR markers (QrZAG112, QrZAG96, QpZAG15, QrZAG20, QrZAG11, QrZAG7 (Guichoux et al. 2011; Kampfer et al. 1998), data not shown).

SNP marker analysis and data clean up

SNP marker analysis was based on the markers developed by Degen et al. (2021a) based on sequence data from Schroeder et al. (2016) and Blanc-Jolivet et al. (2020). SNP marker analysis was carried out by LGC Genomics GmbH (Berlin, Germany) using targeted genotyping by sequencing as described in Degen et al. (2021b). Sequencing was

based on Illumina NextSeq, using single primer enrichment technology generating 75 bp single-end reads at an average coverage of 200x per sample. Adapter sequences and low-quality regions were trimmed from the reads. Reads with an average phred score below 30 over a window of 10 bases, reads shorter than 65 bp and reads containing Ns were discarded. Read alignments against self-assembled contigs were carried out using Bowtie2 v2.2.3 (Langmead and Salzberg 2012). Variants were called with Freebayes v1.3.2 (Garrison and Marth 2016). Nuclear variants were called with a ploidy level of 2 and organelle variants were called with a ploidy level of 1. Genotypes were filtered for a minimum coverage of 3 reads using vcftools (Danecek et al. 2011). Initially, a total of 412 nuclear, 22 chloroplast and 7 mitochondrial markers have been analysed. However, a few markers were later removed from the dataset since negative controls included within the analysed samples showed read counts of considerable number or due to other inconsistencies regarding marker scoring (e.g., marker scoring deviations between different sample batches). A total of 403 nuclear, 21 cp and 7 mt SNP-markers were used for the final analyses.

A total of 3028 oak individuals from the NFI were sent to SNP analysis. Additionally, 41 individuals with known cp haplotype according to Petit et al. (2002a) were analysed. Later, individuals with more than 30% missing data were removed from the dataset, with nuclear and plastid SNP datasets considered independently. Individuals with missing geographic information (37 individuals, potentially originating from 9 different tracts without specified geographic information) were not considered in most analyses, but were included in species discrimination analysis. Identified duplicate individuals were removed from the dataset. Duplicate individuals with identical genotypes may occur due to: (1) the analysis of control samples or (2) incorrect sampling of individuals (e.g., caused by intertwined growth of tree crowns). Individuals with identical genotypes were identified by SSR analysis (control samples with expected duplication of individuals, see above) and additionally - after SNP analysis - using an in-house ruby script. A pairwise comparison of all samples was performed by averaging a similarity score over all loci. The score was increased by 1 if both alleles of one locus were identical between two samples and by 0.5 when one allele was identical. The final number of individuals analysed is indicated individually in the description of each analysis. The final datasets consisted of a maximum of 2752 individuals for the nuclear SNPs and 2889 individuals for the cp and mt SNPs, respectively.

Analysis of Chloroplast and mitochondrial haplotypes

Analysis of chloroplast (cp) haplotypes (hereafter referred to as haplotypes) and mitochondrial (mt) haplotypes (hereafter referred to as mitotypes) was based on 21 cp and 7 mt SNP markers. Due to the use of the same marker set, the chloroplast haplotypes can be assigned to the haplotypes described in Degen et al. (2021b). To allow the assignment of the detected multi-locus haplotypes to the haplotype numbers described by Petit et al. (2002a, b), individuals of known haplotype according to Petit et al. (2002a, b) were included into the SNP analysis. If possible, haplotype numbers in this paper were chosen in accordance with Petit et al. (2002a, b).

Structure analysis

Bayesian clustering was applied for analysis of genetic structures, using the software STRUCTURE (Pritchard et al. 2000). Run parameters for the different STRUCTURE runs were: (1) Species discrimination: 2752 individuals from 584 tracts, 403 nuclear markers. $K=1$ to $K=5$, at 15 runs per K, admixture, correlated allele frequencies, 20,000 burn-in replications, 20,000 Markov Chain Monte Carlo (MCMC) replications after burn-in. All other parameters were set to default. (2) Hierarchical run *Q. robur* only: 1551 individuals from 401 tracts that had been assigned to the *Q. robur* cluster with a minimum share of 90% (mean membership coefficient $Q \geq 0.9$) during the previous species discrimination run, 403 nuclear markers, $K=1$ to $K=10$, at 15 runs per K, two separate runs with admixture/no admixture, correlated allele frequencies, 20,000 burn-in replications, 20,000 MCMC replications after burn-in. All other parameters were set to default. (3) Additional run with *Q. robur* with mean membership coefficient $Q \geq 0.99$: 1340 individuals from 397 tracts that had been assigned to the *Q. robur* cluster with a minimum share of 99% (mean membership coefficient $Q \geq 0.99$) during the previous species discrimination run, 403 nuclear markers, $K=1$ to $K=10$, at 15 runs per K, two separate runs with admixture/no admixture, correlated allele frequencies, 20,000 burn-in replications, 20,000 MCMC replications after burn-in. All other parameters were set to default. (4) Hierarchical run *Q. petraea* only: 1024 individuals from 285 tracts that had been assigned to the *Q. petraea* cluster with a minimum share of 90% (mean membership coefficient $Q \geq 0.9$) during the previous species discrimination run, 403 nuclear markers, $K=1$ to $K=10$, at 15 runs per K, two separate runs with admixture/no admixture, correlated allele frequencies, 20,000 burn-in replications, 20,000 MCMC replications after burn-in. All other parameters were set to default.

The optimal number of genetic clusters (K) was estimated with both the ΔK method described in Evanno et al. (2005) and the ln(K)-method described in Pritchard et al. (2000) and Pritchard et al. (2010). The webtool StructureSelector (Li and Liu 2018) was used for determination of the optimal K. The results of the different runs per K were merged using the web tool CLUMPAK (Kopelman et al. 2015). The web tool STRUCTURE PLOT (Ramasamy et al. 2014) was used for graphical display of the results.

Analysis of potential adaptation to environmental conditions

The SNP set used in this analysis contains 403 nuclear SNPs and was specifically developed for species discrimination and analysis of spatial genetic population structure. Accordingly, the SNPs were selected to be polymorphic and geographically informative. Given both the small number of SNPs and the rationale behind marker selection, a classic genotype-environment association analysis aiming at the identification of specific alleles with adaptive significance to certain environmental conditions was not promising. In order to understand whether intraspecific genetic clusters identified during the STRUCTURE analysis were connected to relevant environmental variables, we employed a classic correlative ecological niche model (ENM), treating different genetic clusters as ecotypes (Thuiller 2024). This approach allows us to draw first conclusions whether environmental conditions may contribute to shaping spatial genetic structures in our study species. The ENM was implemented by constructing generalized linear models (GLM), treating mean Q values resulting from the species-specific STRUCTURE runs (proportionately assigning each sampling tract to an identified genetic cluster) as response variable and environmental values at each sampling tract as predictor variables. This approach allowed to analyse whether environmental conditions predict the genetic composition of the population at a given sampling tract.

27 environmental parameters were selected as candidates to be included in the model: elevation; maximum, minimum and average spring, summer, autumn and winter temperature; mean annual temperature; continentality index; extreme minimum temperature over 30 years; number of frost-free days; average spring, winter, summer and autumn precipitation; mean annual precipitation; precipitation as snow between August and July; mean growing season (May to September) precipitation; annual heat: moisture index; summer heat: moisture index; Hargreave's climatic moisture deficit. Climate variables were obtained through the ClimateEU software package (Marchi et al. 2020), covering the 1961–1990 climate reference period for all sampling tracts.

The data distribution of the response variable was first analysed using the descdist() function of the fitdistrplus R-package (Delignette-Muller and Dutang 2015) to identify an appropriate distribution family and link function for the GLM. The GLM was then built with the gamlss() function of the gamlss R-package (Rigby and Stasinopoulos 2005). While developed to be used primarily with complex generalized additive models, gamlss also offers advantages for fitting classic GLMs with specialized data distribution families and link functions. The Q values derived from the STRUCTURE analysis represent proportions from a total, thus reflecting a population's affiliation with a specific genetic cluster. To account for this specific data distribution type, the model was fitted using a beta inflated family and a logit link function.

Selection of the best model was done using the stepGAIC() function of gamlss. This function performs a stepwise forward and backward model selection procedure and is based on the generalized Akaike information criterion (GAIC). Collinearity between predictor variables was addressed by calculating the variance inflation factor (VIF) of all predictor variables contained in the model after the stepGAIC-selection process using the vif() function of the car package (Fox and Weisberg 2018). If the model contained predictor variables with a VIF higher than 3, the variable with the highest VIF was excluded. This process was reiterated until VIFs were smaller than 3 for all remaining predictor variables (Zuur et al. 2010). All R-based analyses were performed in R 4.3.0 (R Core Team 2013).

Calculation of population parameters

In order to compensate for biases caused by species mixture or possible differences in family structure between tracts, individuals from neighbouring tracts were grouped into larger regional groups according to the following criteria: each group should contain 19 to 22 individuals with a maximum of 24 individuals and form a geographically coherent unit which does not overlap with neighbouring groups. The grouping was carried out separately for *Q. robur* and *Q. petraea*. All individuals of a species within a tract were assigned to the same regional group. All genetic diversity calculations and all subsequent analyses were then based on these groups.

As genetic diversity parameters, average (Na) and effective (Ne) number of alleles as well as observed (Ho) and expected (He) heterozygosity were calculated using GenAlEx 6.5 (Peakall and Smouse 2012). Mean kinship (Loiselle et al. 1995), Allelic richness (Ar), proportion of polymorphic loci (Poly), and fixation index (Fis) were calculated with GDA_NT2021 (Degen 2022).

Genetic diversity parameters were correlated against group size to identify possible biases caused by group size variation and against longitude and latitude to test for spatial genetic structuring. We further assessed spatial genetic autocorrelation between groups by calculating Moran's I as a measure of pairwise genetic relationship between all groups and averaging them over distance classes. Distance classes were weighted in order to contain identical numbers of pairwise comparisons. Statistical significance was established by random permutation and construction of 95% confidence intervals.

Maps

Maps were drawn using ArcMap 10.8.1 (esri, Redlands, USA) and the Light Grey Canvas basemap. Data interpolation was performed using the IWD interpolation tool included in the Spatial Analyst toolbox, with the standard setting of 12 neighbouring data points.

Continentality index data (TD, defined as the temperature difference between mean warmest month temperature and mean coldest month temperature, reference period 1961–1990) was obtained through the ClimateEU software package (Marchi et al. 2020). Datapoints were located on a 5 km grid covering the whole of Germany.

Results

Species differentiation

For the NFI samples, a first STRUCTURE analysis was carried out using the nuclear SNP markers and the individuals of both species. Based on the ΔK and ln(K)-method the best representation of the data was achieved at K = 2. At K = 2 the individuals were separated into two groups, representing the two different species, together with a relatively low share of hybrids and individuals with a certain level of introgression. Following the criteria defined by Lepais et al. (2009), individuals with mean membership coefficients (Q) ≥ 0.9 in one of the two clusters were classified as pure species and individuals with Q-values < 0.9 as individuals showing signs of hybridization or introgression. We further refined the latter into two more Q-value-classes. Individuals with Q-values in the range of > 0.7-<0.9 for one of the two clusters were classified as individuals showing lower degrees of introgression (introgression levels between 10 and 30%). Individuals with Q-values in the range 0.5–0.7 (introgression levels between 30 and 50%) were classified as individuals with high introgression, potential F1-hybrids or recent backcrosses.

Figure 1A shows the bar plot representing the mean membership coefficients for K = 2 (see Supp. Figure 2A for ΔK and ln(K) plots) merged with CLUMPAK (Kopelman et al. 2015) and sorted according to mean membership coefficient (Q). Each individual is represented by a single vertical line in the bar plot.

Although equal sampling of *Q. robur* and *Q. petraea* was intended, the bar plot shows that more Q. *robur* than *Q. petraea* has been sampled (N = 1597 (58%) for *Q. robur* and N = 1098 (39,9%) for *Q. petraea* (including individuals with introgression levels between 10 and 30%). Individuals with high levels of introgression/potential hybrids account for only 2.1% of the analysed individuals (N = 57). Individuals with low levels of introgression, account for only 3% of all analysed individuals. The percentage of individuals with introgression levels between 10 and 30% differs per species: *Q. robur*: 1.75% (28 out of 1597 *Q. robur* individuals), *Q. petraea*: 5.1% (56 out of 1098 *Q. petraea* individuals).

Figure 1B shows a map of the species composition of the different tracts. Even though species-pure sampling per tract was intended, 19% (N = 111) of 584 tracts contained individuals of both species within the five sampled trees (including individuals with introgression levels between 10 and 30%). Even though co-occurrence of both species in mixed stands occurs quite frequently, the detected number of hybrids and the level of introgression is relatively low.

Chloroplast and mitochondrial haplotypes

Since the used SNP marker set also contains 21 chloroplast and 7 mitochondrial markers, the multilocus genotypes of these markers were used to determine chloroplast and mitochondrial haplotypes (the latter referred to as mitotypes). Table 1 lists all observed haplo- and mitotypes together with the corresponding haplotype names used in previous studies (Petit et al. 2002a, b; Degen et al. 2021b). If possible, haplotype numbers in this paper were chosen in accordance with Petit et al. (2002a).

Figure 2 shows maps of the distribution of the different haplotypes (A) and mitotypes (B) over Germany. Identical haplotypes can be found in both species, forming very similar geographic patterns within Europe (Petit et al. 2002a, b). For this reason, the maps combine the haplotype and mitotype data of both species. Maps for the chloroplast haplotypes, separated by species are given in Supp. Figure 3.

The different haplo- and mitotypes show a clear association. With the exception of a few single individuals (N = 3), which are either misscored or represent potentially imported exotic genotypes, nearly all individuals of one haplotype are fixed to a certain mitotype. Only haplotype 10/11/12 separates into two different mitotypes (C and D). As this haplotype corresponds to 3 different Petit et al. (2002a) haplotypes, this is probably due to a non-sufficient resolution of the used marker set. This haplotype potentially represents a

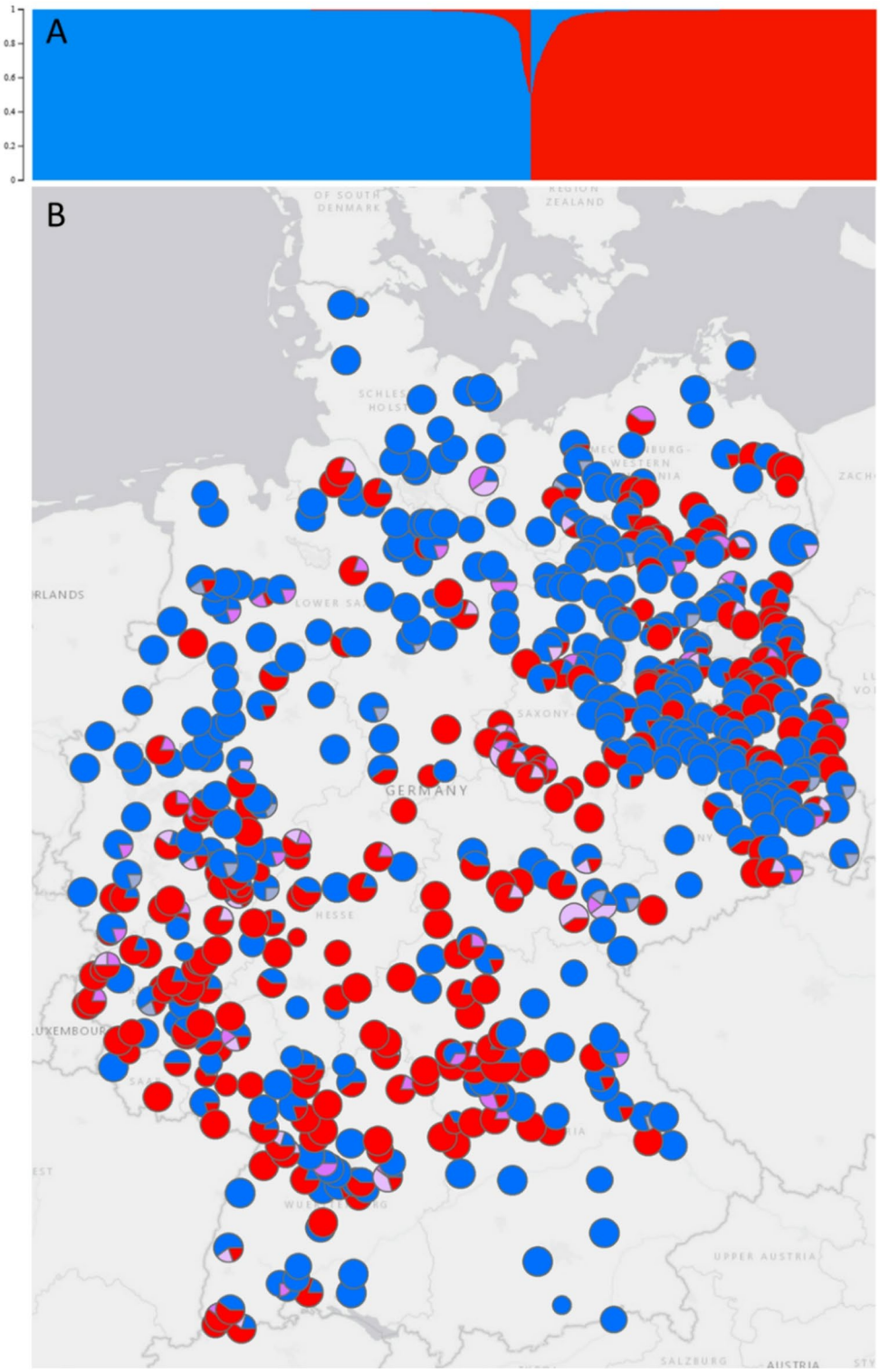

Fig. 1 **A**) STRUCTURE bar plot for K = 2, sorted according to mean membership coefficient (Q). Each individual is represented by a vertical line showing its individual share of both species-clusters. Blue: species-cluster *Q. robur*, red: species-cluster *Q. petraea.***B**) species composition of the different NFI tracts. Each tract is represented by a pie chart located at the publicly available INSPIRE coordinates of the respective tract. Pie chart size varies with the number of analysed individuals per tract. Species affiliation of every individual is colour coded: red: *Q. petraea*, light pink: *Q. petraea* with 10–30% introgression of *Q. robur*, purple: potential hybrids with 30–50% introgression, light blue: *Q. robur* with 10–30% introgression of *Q. petraea*, blue: *Q. robur*

combination of different haplotypes, associated with different mitotypes. Individuals with the rare haplotype 101 and 102 feature the same mitotypes (F or A) as individuals with haplotypes 4 or 1, respectively. Due to the nearly complete linkage of haplo- and mitotypes, the distribution of mitotypes mirrors the distribution of chloroplast haplotypes. Associated haplo- and mitotypes are listed in the same row in Table 1.

Total frequencies of the different haplotypes (Supp. Figure 4) are similar in both species. Haplotype 1 is the most frequent haplotype in both *Q. robur* and *Q. petraea.* The

Table 1 Haplo- and mitotypes. Associated haplo- and mitotypes are listed in the same row

Haplo type	Haplo type	Haplo type	cp2	cp3	cp4	cp5	cp6	cp7	cp8	cp9	cp10	cp11	cp12	cp13	cp14	cp15	cp16	cp17	cp18	cp19	cp20	cp21	cp22	Mito type	mt1	mt2	mt3	mt4	mt5	mt6	mt7
this paper	Petit et al. 2002a, b	Degen et al. 2021b																						this paper							
1	1	H_11, 12	C	C	A	G	T	G	A	T	G	C	C	C	A	T	T	G	T	A	C	T	G	A	G	A	A	G	A	T	A
4	4	H_1	C	A	G	A	C	G	T	G	G	A	A	C	A	T	T	G	T	A	C	A	G	F	T	A	C	G	C	T	C
5.1	5	H_2	C	A	G	A	C	G	T	G	G	A	A	C	A	T	T	G	T	A	C	T	G	B	G	A	A	G	C	G	A
5.2	5	H_6,7, 8	C	A	G	A	T	G	T	T	T	A	A	C	A	T	T	G	T	A	C	T	G	G	G	C	A	C	A	G	A
7	7	H_13	T	A	G	A	C	G	T	G	G	A	A	C	C	T	T	G	T	A	G	T	G	E	T	A	C	G	C	G	C
10, 11, 12	10, 11, 12	H_5	C	A	G	A	T	A	T	G	G	A	A	T	A	C	G	G	T	A	C	T	T	C	T	A	A	G	A	T	A
																								D	T	A	A	G	C	T	A
101		H_3	C	A	G	A	C	G	T	G	G	A	A	C	C	T	T	G	T	A	C	T	G	F	T	A	C	G	C	T	C
102		H_4	C	A	G	A	T	A	T	G	G	A	A	C	A	T	G	G	T	A	C	T	G	A	G	A	A	G	A	T	A

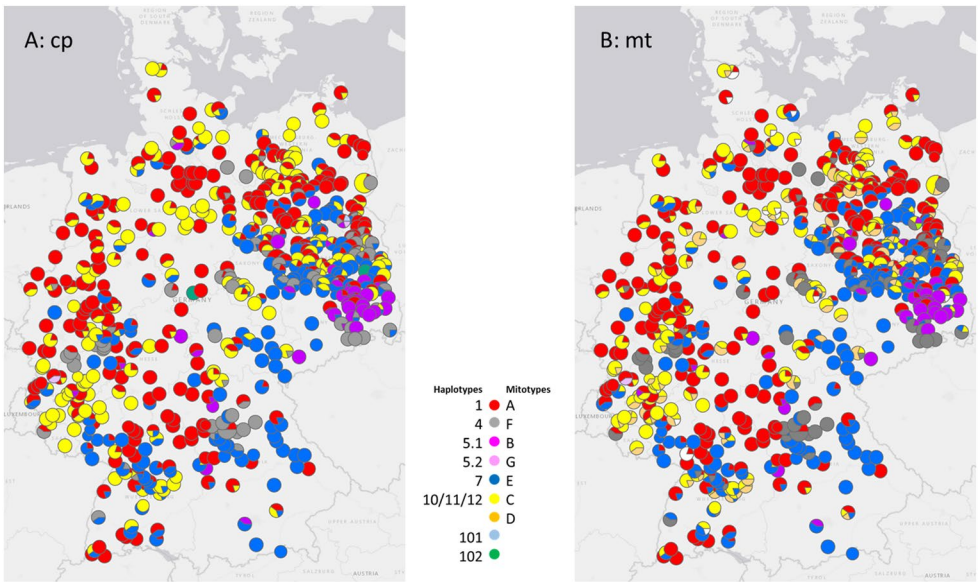

Fig. 2 Maps of the assignment of the individuals of the different NFI tracts to the different chloroplast haplotypes (**A**) and mitotypes (**B**). Each tract is represented by a pie chart located at the publicly available INSPIRE coordinates of the respective tract. Pie chart size varies with the number of analysed individuals per tract. Each haplo-/mitotype is represented by a different colour (see legend) within the pie charts

second most common haplotypes are haplotypes 7 and 10/11/12.

Intraspecific spatial genetic structure

Genetic species assignment from the first STRUCTURE analysis was used to carry out hierarchical STRUCTURE analyses, using the nuclear SNP marker data of the individuals of either *Q. robur* or *Q. petraea* (maximum level of introgression: 10%). STRUCTURE runs were carried out with varying parameters, either assuming genetic admixture or no genetic admixture between the different potential populations. Assuming no genetic admixture between populations will often generate more distinct results and better detects weak genetic structures. However, assuming a certain amount of admixture between different wind-pollinated oak populations is certainly a much more realistic scenario.

Spatial genetic structure and potential adaptation to environmental conditions in Quercus robur

For *Q. robur*, the best representation of the data, based on the ΔK and ln(K) method, was achieved for K = 2 when assuming no admixture and for K = 3 when assuming admixture between the populations (see Supp. Figure 2B and 2C for ΔK and ln(K) plots). The barplots of the mean membership coefficients of the respective STRUCTURE analyses are given in Supp. Figure 5.

Sorted by federal state and latitude (within each federal state), a spatial genetic structure becomes apparent in the barplots: At K = 2 (no admixture), individuals from north-western regions were predominantly assigned to one genetic cluster while individuals from south-eastern regions were predominantly assigned to another cluster. At K = 3 (admixture), the more south-eastern cluster is divided into two subclusters, which do not contribute to further regional differentiation. Projecting the STRUCTURE-results onto a geographic map of Germany gives a clear impression of the spatial genetic structure and its north-west to south-east gradient. Figure 3 shows a map of the average assignment of the different NFI tracts to the different STRUCTURE clusters. As the additional cluster at K = 3 did not contribute any additional differentiation, we proceeded with the results for K = 2. For easier visualisation the map background shows an interpolation of the cluster assignment using the IWD interpolation method included within ArcMap 10.8.1 (esri, Redlands, USA).

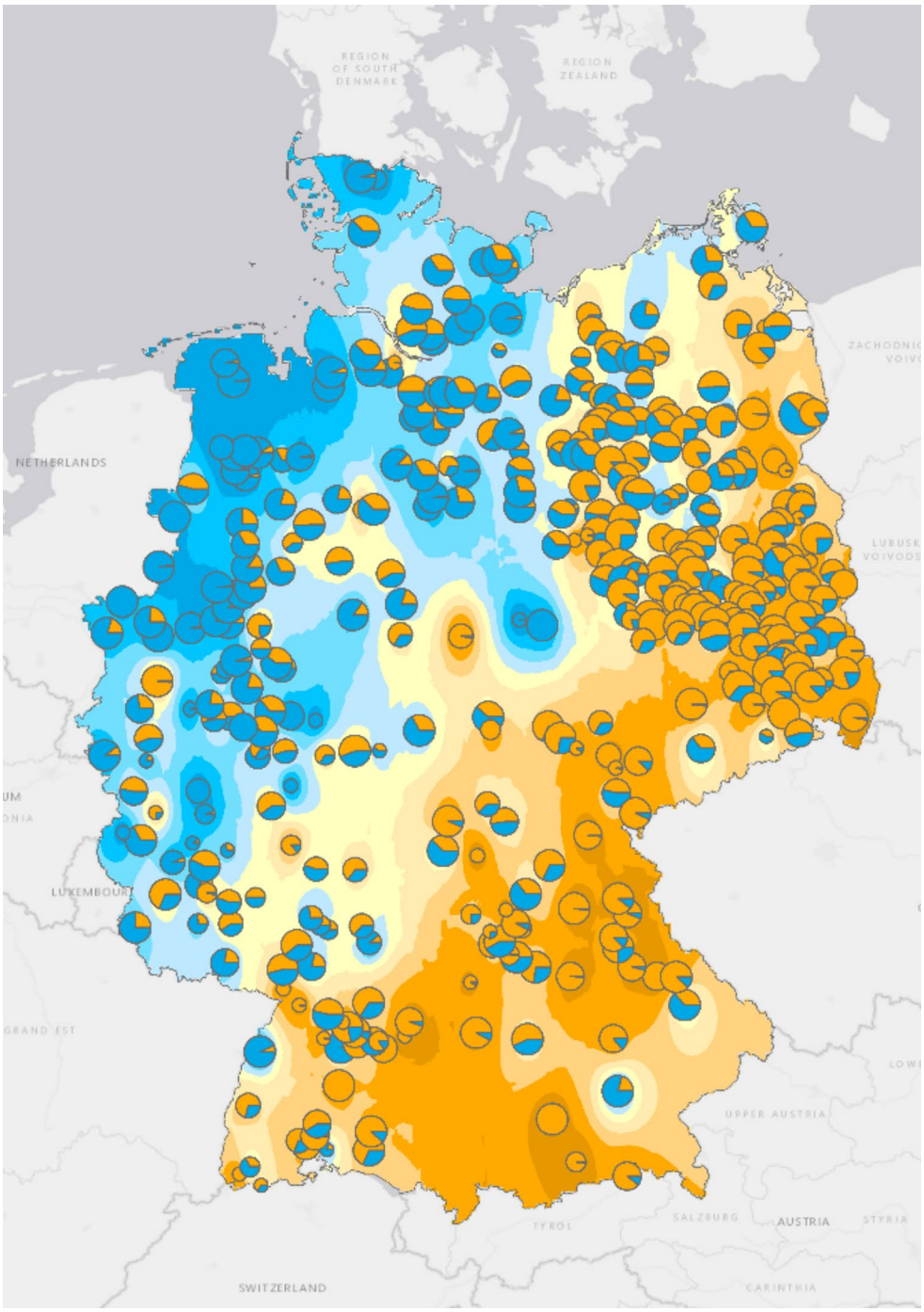

Fig. 3 Map of the average assignment of the *Q. robur* individuals of the different NFI tracts to the STRUCTURE clusters at K2, no admixture model. Each tract is represented by a pie chart located at the publicly available INSPIRE coordinates of the respective tract. Pie chart size varies with the number of analysed *Q. robur* individuals per tract (for species mixed tracts, only the *Q. robur* individuals are included). Each STRUCTURE cluster is represented by a different colour within the pie charts. Colours of the different STRUCTURE clusters are in accordance with the STRUCTURE bar plots in Supp. Figure 5. Map background shows an interpolation of the cluster assignment using the IWD interpolation method included within ArcMap 10.8.1 (esri, Redlands, USA)

To test whether differences in the levels of historic introgression from *Q. petraea* into *Q. robur*. influence the observed genetic north-west to south-east structure (Degen et al. 2021b), we tested our *Q. robur* data for higher introgression levels in the north-western than in the south-eastern STRUCTURE cluster (please note that only individuals with less than 10% detected introgression have been included into the STRUCTURE analysis) and observed a higher average level of introgression in the north-western cluster (⌀ north-west: 0.01, ⌀ south-east: 0,003, $p = 4.61 \times 10^{-28}$). However, a STRUCTURE analysis including only *Q. robur* individuals with Q-values ≥ 0.99 (nearly no introgression of *Q. petraea* in order to test the genetic structure in the absence of any introgression signal) shows a similar pattern separating the north-west from the south-east of Germany (Supp. Figure 6). The two different spatial genetic clusters can thus also be identified, if different levels of introgression probably have no significant influence.

To test if the spatial genetic structure of *Q. robur* could be influenced by a different climate adaption of the different lineages, we analysed whether the genetic clusters identified during the STRUCTURE analysis were connected to relevant environmental variables using an ecological niche model based on GLMs. We found that the spatial distribution of the two clusters is correlated to environmental conditions and seems to be driven mainly by continentality. The final model included continentality ($P = 2*10^{-16}$), summer precipitation ($P = 0.03$) and two variables that did not reach statistical significance, indicating that these had no systematic effect on the response variable: maximum autumn temperature ($P = 0.50$), and minimum spring temperature ($P = 0.61$). With an $R^2 = 0.28$, the final model showed medium explanatory power. Most of the variation was explained by continentality alone ($R^2 = 0.26$).

Figure 4 shows a map of the continentality levels within Germany, illustrating that continentality also follows a north-west to south-east gradient. Comparing Figs. 3 and 4 gives a good visual impression about the congruence between the gradients observed in genetic structure and in continentality.

Spatial genetic structure in Quercus petraea

For the analysed *Q. petraea* individuals, the spatial genetic structure analysis remained inconclusive. In north-western and south-eastern parts of Germany only relatively few (north-west) to no (south-east) *Q. petraea* tracts have been sampled. This is not based on incomplete sampling, but on the low distribution level of *Q. petraea* in the respective regions (see, e.g., map with probability of presence in Eaton et al. (2016). For the sampled tracts, the STRUCTURE run assuming admixture between the different potential populations resulted in no distinct regional clusters (data not shown). Using the no admixture model, the best representation of the data, based on the ΔK and ln(K) method, was achieved for K = 2, with a second maximum at K = 5 (see Supp. Figure 2D for ΔK and ln(K) plots). At K = 2 no distinct regional clusters were detected (data not shown). At K = 5 only one of five clusters (shown in orange in Supp. Figure 7) showed a regional differentiated distribution. Figure 5 shows a map of the average assignment of the different NFI tracts to the different clusters at K = 5. Only the cluster highlighted in orange shows a regionally differentiated distribution, limited to the eastern parts of Germany. The remaining four clusters dominate the rest of the distribution area within Germany but do not add any spatial genetic differentiation.

Since the STRUCTURE-analysis remained inconclusive for *Q. petraea*, we performed no further analysis to investigate whether any genetic clusters were correlated to environmental variables for this species.

To check if the regional differentiated eastern cluster that emerged at K = 5 is of potential artificial origin, we checked if the mean proportion of this cluster differs between naturally regenerated and planted stands. For 577 *Q. petrea* individuals from 159 tracts information about the stand history (naturally regenerated or planted) was available. In tracts localized within naturally regenerated stands, the regionally differentiated K5 cluster occurs in a frequency of 11.8% within all of Germany. This increases to 22.6% within the eastern federal states of Mecklenburg-Pomerania, Brandenburg, Saxony, and Saxony-Anhalt and to 29.7% within the state of Brandenburg alone. This illustrates the regional differentiation of the cluster, which is also clearly recognizable in natural regenerated stands.

However, in tracts localized within planted stands the frequency is 32.7% within all of Germany, 33.6% within the eastern states and 41.6% within Brandenburg. This indicates a disproportionately high frequency of the regionally differentiated K5 cluster in planted tracts. The naturally existing regional differentiation of the cluster thus appears to have been reinforced by planting.

Differences in genetic diversity

An initial analysis of genetic diversity parameters at the tract level showed that the presence of family structures as well as species mixture within a tract influenced genetic diversity estimates. The presence of different species, hybrids and individuals with a high degree of introgression within a tract led on average to a higher number of alleles in these tracts, while tracts with family structures, i.e., tracts with a higher degree of kinship between individuals, tended to have a lower number of alleles (Supp. Figures 8 and 9).

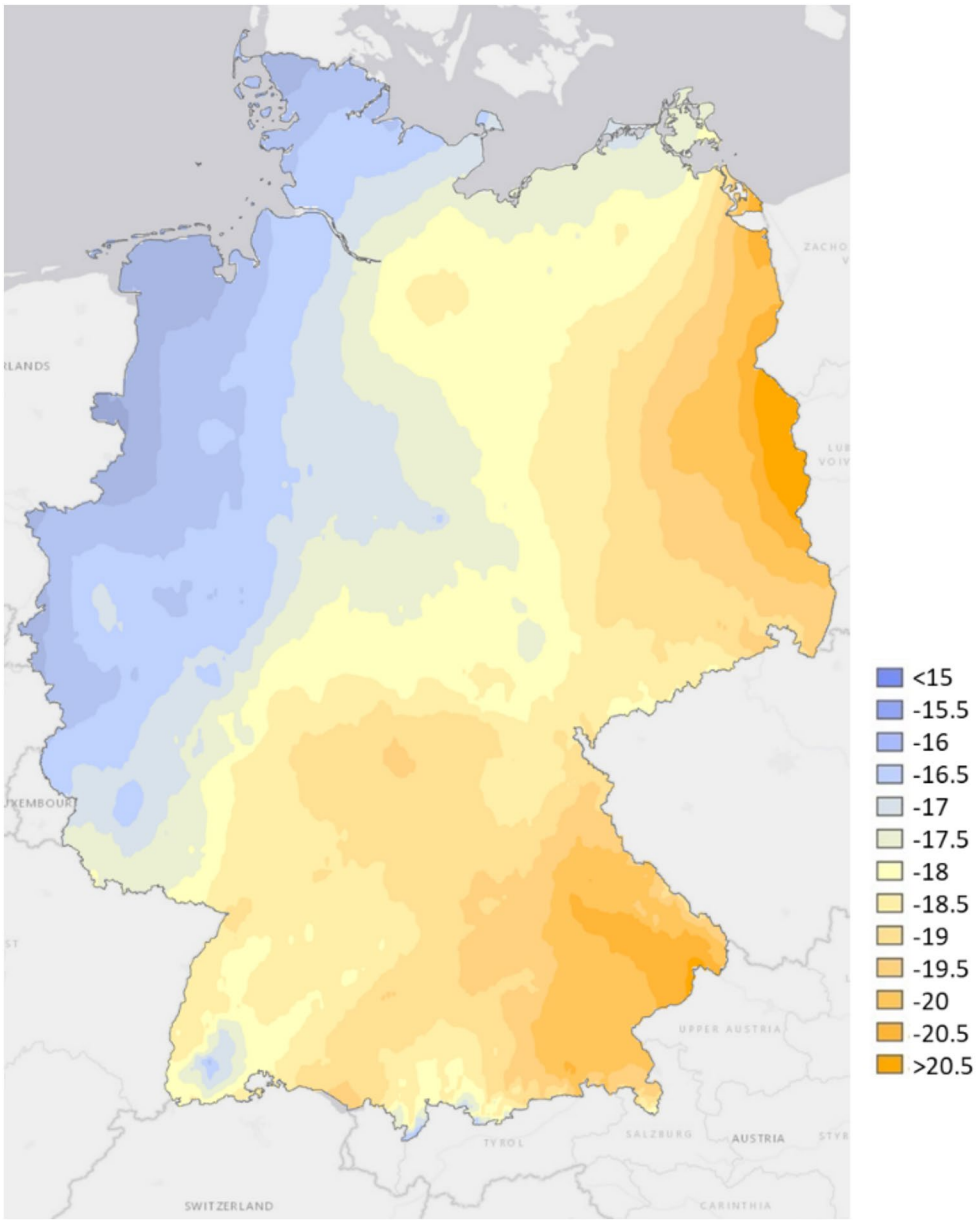

Fig. 4 Continentality/TD [°C] in Germany (reference period 1961–1990). Interpolation is based on datapoints located on a 5 km grid

To alleviate these biases, we combined trees from the same species from neighbouring tracts into larger groups containing 19–24 individuals from four to nine individual tracts. Correlation of genetic diversity parameters showed no significant effect of group size on diversity (for *Q. robur*: Na: $p=0.54994$, Poly: $p=0.77539$, Ar, Ne, Ho, He=>$p=1$, for *Q. petraea*: Na: $p=0.13523$, Poly: $p=0.77539$, Ar, Ne, Ho, He=>$p=1$).

For *Q. robur*, effective number of alleles (Ne) and expected heterozygosity (He) were positively correlated to latitude, declining along a north-south gradient (Ne=1.348–1.387, r(Lat/Ne)=0.52169, $p=0.000117$; He=0.209–0.230, r(Lat/He)=0.52431, $p=0.000102$). Effective number of alleles (Ne), expected heterozygosity (He), and observed heterozygosity (Ho) showed significant positive spatial autocorrelation up to about 200 km and significant negative autocorrelation from about 550 km (Supp. Figure 10).

For *Q. petraea*, the number of alleles (Na) and proportion of polymorphic loci (Poly) were positively correlated to both latitude and longitude, declining both along a north-south and a west-east gradient (Na=1.712–1.816, r (Na/lat)=0.49415, $p=0.021709$; r(Na/lon)=0.5058, $p=0.015119$; poly=69.48 −79.65%, r(Poly/lat)=0.51559, $p=0.011041$; r(Poly/lon)=0.53975, $p=0.004875$). Additionally, allelic richness (Ar) was positively correlated to latitude only, declining from north to south (Ar=1.66–1.73, r(Ar/lat)=0 0.52758, $p=0.007416$). Na, Poly, and Ar showed significant positive spatial autocorrelation up

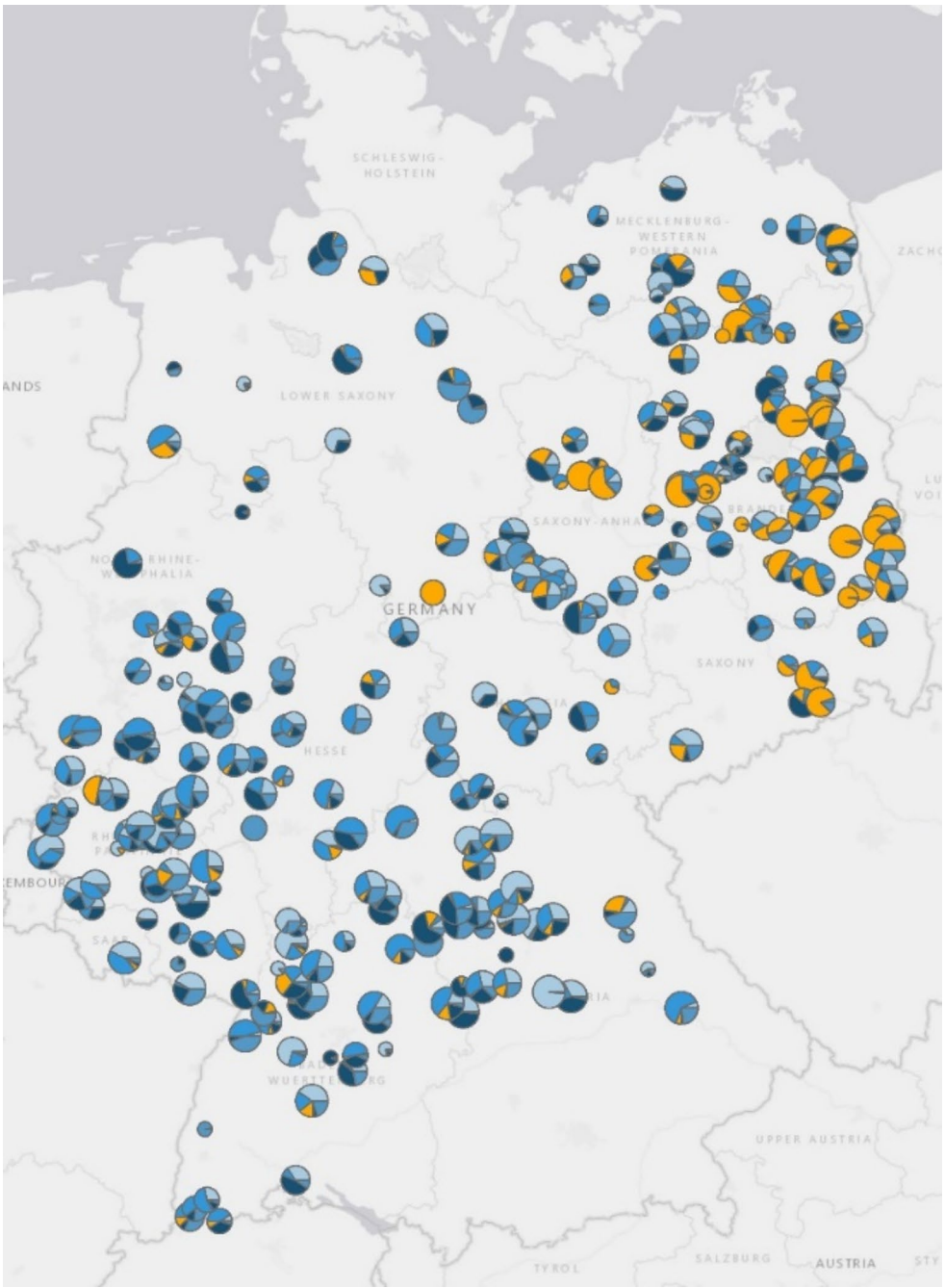

Fig. 5 Map of the average assignment of the *Q. petraea* individuals of the different NFI tracts to the STRUCTURE clusters at K5, no admixture model. Each tract is represented by a pie chart located at the publicly available INSPIRE coordinates of the respective tract. Pie chart size varies with the number of analysed *Q. petraea* individuals per tract. Each cluster is represented by a different colour within the pie charts. Colours of the different clusters are in accordance with the STRUCTURE bar plots in Supp. Figure 7. Due to the less homogenous distribution of Q. petraea and the weak geographical structure, this map is shown without underlying interpolation

to 167 km and significant negative autocorrelation from 339 km (Supp. Figure 11).

All other genetic diversity parameters showed no significant correlation to latitude or longitude and no significant autocorrelation.

Discussion

Species differentiation

The STRUCTURE analysis including the nuclear SNP data of both species, allowed a clear separation of both species. The ability of the used SNP set to differentiate between *Q. robur* and *Q. petraea* has previously been shown by Degen et al. (2021a). Although equal sampling of *Q. robur* and *Q. petraea* was intended, the genetic determination of more *Q. robur* than *Q. petraea* is probably due to erroneous species

identification in the field. This shows that morphological species determination in the field is not always easy to achieve. The detection of a relatively high amount of mixed stands, in which individuals of both species have been sampled, confirms the observation that co-occurrence of both species at the same site is frequent (Eaton et al. 2016; Kleinschmit et al. 1995). However, the non-random sampling and low number of sampled individuals per tract does not allow conclusions about the general share of mixed stands in Germany. The degree of hybridization and introgression found within our analysis is quite low. With a total of 5.1% of individuals showing various degrees of introgression, our results are well within the range described from mixed stands in Western Europe (Curtu et al. 2007, 2009; Lepais et al. 2009; Streiff et al. 1999). Our results indicate that pedunculate and sessile oak in Germany have kept and are still keeping their genetic species integrity despite millennia of introgression and the importance of interspecific gene-flow for adaptation to the Central European climate during recolonization following the last glacial maximum (Leroy et al. 2020a).

Chloroplast and mitochondrial haplotypes

The assignment of the chloroplast multi-locus-haplotypes detected within this study to the haplotypes according to Petit et al. (2000a, b) was possible due to the analysis of individuals of known Petit et al. (2002a)-haplotypes within the same analysis. Such an assignment was also successfully performed by, e.g., Neophytou and Michiels (2013), based on the analysis of 10 chloroplast SSR markers.

The different haplo- and mitotypes found in our dataset are shared by both species. Moreover, we found a clear association between certain haplo- and mitotypes, which is also the same in both species.

Dumolin-Lapègue et al. (1998) report that patterns of chloroplast and mitochondrial lineages in oak show a clear association in Southern France. Our results confirm that this association also exists within Germany. As described by Petit et al. (2004); Leroy et al. (2020a) the co-occurrence of identical haplotypes in both species is probably due to the genetic contact of both species during the period of recolonization after the last glacial maximum. The patterns of chloroplast haplotypes in Germany is in good accordance with the results reported in Petit et al. 2002a, b; Degen et al. (2021b). Haplotype 1 (red), originating from the Italian Peninsula, is the most common haplotype in Germany. It is widely distributed mainly in Northern and Western Germany. Haplotype 10/11/12 (yellow) corresponds to three different Petit et al. (2002a) haplotypes originating from the Iberian Peninsula and is found predominantly in the north-western part of Germany. Haplotype 7 (blue), originating from the Balkan region, is distributed mainly over the southern and eastern regions. Haplotype 4 (grey) shows no distinct pattern of distribution. Haplotypes 5.1 and 5.2 (purple), corresponding to haplotype lineage 5, that can predominantly be found in Italy and Eastern Europe, show a clear main area of distribution in Germany located in the south-eastern region of Saxony. The distribution of the different haplotypes within Germany can therefore be well explained by the different routes of recolonization from different glacial refugia after the last glacial maximum. Some rare/unusual haplotypes (101(light blue) and 102 (green) corresponding to haplotypes H_3 and H_4 in Degen et al. (2021b) can be found in some *Q. petraea* individuals located in Southern Lower Saxony and Brandenburg. The occurrence of these haplotypes of potential Eastern European origin is probably caused by human seed transfer. Also noticeable are tracts with a very heterogenous haplotype composition. One tract in the north-east of Berlin, for example, includes 5 individuals with 5 different haplotypes sampled within the small sampling area. This points towards an artificial origin of the sampled stand, which probably originated from a seed mixture.

Total frequencies of the different haplotypes (Supp. Figure 4) are similar in both species. Observed differences are potentially influenced by the distribution of the samples of each species over Germany and the different haplotype zones.

Intraspecific Spatial genetic structure in *Q. robur*

For *Q. robur* we found a clear genetic spatial genetic structure of two clusters showing a north-west to south-east gradient. Such a north-west to south-east separation of the *Q. robur* spatial genetic structure within Germany has also been observed by Degen et al. (2021b), who analysed the spatial genetic structure of *Q. robur* in Europe. Moreover, it is reminiscent of the north-west and south-east separation of different oak chloroplast haplotypes described by Petit et al. (2002a, b), where the different chloroplast haplotypes represent different postglacial recolonization routes through which the two oak species spread throughout Central Europe from different glacial refugia after the last glacial maximum. The north-western part of Germany has been recolonized mainly by lines originating from the Iberian and Italian Peninsula, while the south-eastern regions are dominated by lines from the Balkan region. The spatial genetic structure observed within the nuclear markers analysed here, could likewise be based on the different paths of recolonization of Europe after the last ice age. This natural genetic structure would then have been largely preserved over time, despite genetic admixture between distant populations and long-term human management, including human seed transfer.

Degen et al. (2021b) suggested historic introgression from *Q. petraea* into *Q. robur* as a possible explanation for the observed genetic north-west to south-east structure. In their study, the recolonization lineage dominating the north-western part of Germany showed stronger influence of historic introgression than the lineage dominating the south-eastern part and these differences in introgression level could possibly cause the nuclear genetic differences between the different recolonization routes. Within our data, we also found higher introgression levels in the north-western than in the south-eastern STRUCTURE cluster. However, an additional STRUCTURE analysis including only pure *Q. robur* individuals with Q-values ≥ 0.99 (Supp. Figure 6) showed a comparable, somewhat less pronounced but still clearly recognizable, north-west to south-east gradient. While indeed causing measurable differences in the genetic background of *Q. robur* populations in Germany, introgression alone is therefore not sufficient to explain the observed genetic structure.

Differences in environmental preadaptation between the two genetic clusters might possibly contribute to the specific distribution of both clusters within Germany. Consisting of only 403 nuclear SNPs that were selected to be geographically informative rather than representing ecologically relevant traits, the SNP set employed in this study is not suited to perform a classic genotype-environment association analysis. To get an impression whether the spatial genetic structure revealed by the STRUCTURE-analysis was caused exclusively by demographic processes or might potentially also be influenced by environmental conditions, we performed an ecological niche model, treating the two main genetic clusters as ecological variants whose spatial distribution might be affected by environmental conditions. We found that the spatial distribution of the two clusters is correlated to environmental conditions and seems to be driven mainly by continentality and to a lesser degree by summer precipitation. The observed explanatory power of the final model ($R^2 = 0.28$) fits well within the range reported for climate-growth relationships for different climate factors as well as different European tree species (Jevšenak 2019; Kolář et al. 2017; Lebourgeois et al. 2005). Continentality and summer precipitation have already been identified as factors affecting growth and survival in introduced Douglas fir in Europe (Niemczyk et al. 2021).

If the identified genetic clusters represent two different postglacial recolonization routes through which *Q. robur* spread throughout Central Europe from different glacial refugia, it seems plausible to assume that source populations from the Iberian Peninsula were preadapted to more Atlantic climate conditions whereas those from the Balkan Peninsula were preadapted to more continental climate conditions. Starting from these source populations, recolonization might then have been guided, at least in part, by clines in continentality throughout Europe, leading to the regional distribution pattern observed in Germany today, where both refugial lines met again (Figs. 3 and 4).

As a word of caution, it has to be mentioned that this analysis is based on a correlative approach, and as such the results do not necessarily indicate causality. Furthermore, our dataset contains considerable nonadaptive population structure caused by demographic processes, in this case postglacial recolonization routes. If these demographic processes are coincidentally correlated to environmental clines, this can mimic an adaptive signal. Given that the demographic signal dominates the entire spatial genetic structure and is most likely more pronounced than any potential adaptive signal, it is not possible to unequivocally discriminate between such an apparent and a true adaptive signal. The ecological niche model therefore serves as a tool to generate hypotheses, indicating whether certain environmental variables might be of adaptive significance. Subsequent studies are needed to experimentally test these hypothesized relationships, ideally incorporating both whole genome data and provenance trials. Our results might be used to inform the selection of populations and trial sites to include in such an experiment in order to better understand any adaptive component in the observed spatial genetic structure of *Q. robur* in Germany.

Intraspecific spatial genetic structure in *Q. petraea*

For the analysed *Q. petraea* individuals, the spatial genetic structure analysis remained inconclusive. How can it be explained, that *Q. robur* and *Q. petraea* species share the same recolonization history and chloroplast and mitochondrial haplotype patterns but for the nuclear markers *Q. robur* shows a clear spatial genetic structure in form of a north-west to south-east gradient, while no comparably spatial genetic signal can be identified for *Q. petraea*?

Generally weaker genetic signals in nuclear markers than in plastid markers are to be expected, since nuclear markers are– in contrast to the purely maternally inherited plastid markers– subject to the mixing influence of pollen-mediated gene-transfer. This however is true for both species.

A lack of genetic structure could result from higher long-distance dispersal capabilities or more pronounced human influence, especially through long-distance seed transfer in *Q. petraea*. In both cases, however, this should influence and distort the spatial genetic structure of both nuclear and plastid markers, which is not the case.

Another explanation for the weak geographic nuclear genetic signal in *Q. petraea* could be a lack of resolution power of the marker set. Maybe the markers used are less suited to uncover genetic differences between different *Q.*

petraea populations due to a bias toward species differentiating markers and markers for detecting spatial genetic structures in *Q. robur* during the compilation of the marker set. However, marker testing and selection was based on both *Q. robur* and *Q. petraea* individuals originating from different regions of Europe (see Degen et al. (2021a); Blanc-Jolivet et al. (2020).

Finally, the geographic scale of our study might be too limited to uncover any meaningful spatial genetic structure in *Q. petraea*. Especially the eastern cluster that emerged at K = 5 under the no admixture-model hints at some subtle large-scale structure that possibly could not be resolved by our geographically restricted approach. A high-density, range-wide survey might therefore be worthwhile in order to provide insights into interregional genetic structures in *Q. petraea*. However, a study using the same SNP-set on an European sample set also revealed a lack of spatial genetic structure within *Q. petraea*, which was explained with potential introgression from other white oak species (Degen et al. 2023). This interpretation is in agreement with the findings of Leroy et al. (2020a) that the recolonization history of *Q. petraea* in Central Europe is strongly characterized by introgression events, mainly by *Q. robur*.

The geographical distribution of the weak eastern cluster that emerged at K = 5 seems to be caused by a combination of natural and anthropogenic processes. This cluster occurs almost exclusively in the eastern federal states of Mecklenburg-Pomerania, Brandenburg, Saxony, and Saxony-Anhalt with a pronounced concentration in Brandenburg (Fig. 5). Interestingly, the mean proportion of this cluster differs between naturally regenerated and planted stands. The cluster seems to occur naturally in Eastern Germany and possibly represents the western margin of a Central- or Central- and Eastern European genetic cluster. While showing a strong localization in naturally regenerated tracts, it has a disproportionately high frequency in planted tracts. This increased frequency again is most pronounced in Brandenburg, leading to an even stronger local signal of the cluster. Whether the high frequency in planted tracts is caused by chance or whether plants originating from this cluster show better growth and adaptation and were therefore specifically selected for planting purposes must currently remain an open question.

Differences in genetic diversity

Genetic diversity within single tracts varied with the degree of species mixture, hybridization, introgression and relatedness. On average, tracts which included a mixture of both species, hybrids or individuals with high introgression levels showed higher allele numbers (Supp. Figure 8). This is probably due to the presence of species-specific alleles or different allele frequencies in the two species. Different levels of family structures within the tracts also have an influence on genetic diversity and allele numbers decreased with increasing degrees of relatedness within a tract (Supp. Figure 9). Sampling was performed within a limited radius of 25 m around one tract corner. Both older and younger trees have been sampled (36.6% of sampled trees had a height below 4 m). Both of this favours the sampling of related individuals within a tract. For example, some younger trees might originate from natural regeneration of the surrounding trees, since the relatively high weight of acorns favours short distances of seed dispersal (Ducousso et al. 1993; Eusemann and Liesebach 2021). And even though oaks are wind pollinated, a high percentage of pollination events also occurs within a limited radius, especially in forest areas with high coverage (Eusemann and Liesebach 2021; Streiff et al. 1999).

In order to compensate for these biases, we combined neighbouring tracts to obtain groups of about 19–22 (in rare cases up to 24) individuals and recalculated all genetic diversity parameters. Based on these groups, we observed a small but significant decline in the effective number of alleles (Ne) and the expected heterozygosity (He) along a north-south gradient in *Q. robur* and a corresponding decline in the number of alleles (Na) and the proportion of polymorphic loci (Poly) along a north-south gradient and also along a west-east-gradient in *Q. petraea*. Additionally, in *Q. petraea*, the allelic richness (Ar) also declined slightly from north to south. In both species and in all diversity parameters, the absolute values and differences are small (difference between minimum and maximum observed diversity: *Q. robur*– Ne = 0.039, He = 0.021, *Q. petraea*– Na = 0.104, Poly = 10.17%, Ar = 0.07). However, given that even strong differences in population structure can be difficult to detect using these genetic diversity indices (Liesebach et al. 2024), the observed geographic pattern can be regarded as a strong indication of pronounced underlying differences in population structure between northern and southern German stands in both species.

To pinpoint the reason behind the observed pattern, however, is not easy. Interestingly, Neophytou et al. (2015) observed similar differences in allelic richness and heterozygosity on a much smaller geographic scale and using a different marker type. They identified differences in genetic diversity between older (> 120 years) and younger (< 70 years) stands of *Q. robur* in the Upper Rhine Valley in France and Germany. They interpreted this result as the effect of a historically recorded switch from long-term natural regeneration by local seed sources to artificial regeneration using long-distance seed transfer around the mid-twentieth century. These two contrasting sylvicultural

systems are supposed to have led to the observed higher genetic diversity in the cohort of young stands.

A similar effect underlying the north-south cline in genetic diversity observed in our samples is possible but could not be tested directly. Information on regeneration history and age is available for some sampling tracts, but is too sparse to allow for reliable statistical inference. We can, however, rule out interregional long-distance seed transfer as the single, deciding factor behind the pattern. The significant similarity between tracts of up to 200 km in *Q. robur* and 167 km in *Q. petraea* along with the significant dissimilarity above 550 km in *Q. robur* and 339 in *Q. petraea* clearly shows the absence of pronounced long-distance gene flow and indicates that both species have been regenerating mostly on a local or regional scale so far. Still, genetic diversity within a population could possibly also be increased by mixing and planting seeds obtained from several different local source populations. We therefore cannot rule out that an effect similar to that observed by Neophytou et al. (2015) also acts on the pattern observed in our study. If this were the case, this pattern could possibly illustrate historic differences in oak management systems in Northern and Southern Germany with a prevalence of artificial regeneration in northern and natural regeneration in southern parts of the country.

Conclusions

Our study represents a high-density genetic survey of *Q. robur* and *Q. petraea* in Germany, an area where the species' distribution and genetic structure is expected to be shaped by complex interactions between long-lasting recolonization processes, hybridization and introgression, as well as human interference.

In regard to our initial research questions, our results indicate that.

1. Species identification based on morphological traits in the field remains challenging and a significant proportion of the samples were found to be misidentified.
2. *Q. robur* and *Q. petraea* are still connected though hybridization and introgression, but both processes remain on low levels and the majority of both individuals can be clearly assigned to either one or the other species throughout Germany.
3. Despite several centuries of human use and management, the genetic structure of *Q. robur* still seems largely natural and can be explained by the postglacial recolonization routes through which the species dispersed throughout Germany after the last glacial maximum. For *Q. petraea*, the situation seems more complicated and no clear genetic structure could be identified.
4. Adding to this natural genetic structure, a pattern of declining genetic diversity following a north-south gradient could be indicative of regional differences in silvicultural systems with more pronounced artificial regeneration in Northern Germany compared to a prevalence of natural regeneration in southern regions.
5. The distribution of *Q. robur* seems to be shaped not only by postglacial recolonization processes but also by environmental factors, most importantly continentality, possibly caused by preadaptation within the refugial source-populations. However, due to the method employed in this analysis, further studies on this aspect are needed in order to produce a more complete picture.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11295-025-01695-9.

Acknowledgements We would like to thank all planers, organizers and inventory teams of the 4th German National Forest Inventory.

Author contributions Birte Pakull: Investigation, Methodology, Formal analysis, Data Curation, Writing - Original Draft. Bernd Degen: Methodology, Formal analysis, Writing - Review & Editing, Conceptualization, Funding acquisition. Hilke Schroeder: Resources, Methodology, Writing - Review & Editing. Thomas Riedel: Resources, Data Curation, Methodology. Malte Mader: Software, Writing - Review & Editing. Heike Liesebach: Writing - Review & Editing. Petra Hoffmann: Investigation. Susanne Hoppe: Investigation. Pascal Eusemann: Methodology, Formal analysis, Project administration, Writing - Original Draft.

Funding Open Access funding enabled and organized by Projekt DEAL.
This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors. The authors have no competing interests to declare that are relevant to the content of this article.

Data Availability Upon acceptance of the manuscript all genotype data will be made available on OSF (https://osf.io/5uxd8/). Coordinates of the NFI tracts based on the Pan-European INSPIRE reference grid are available at https://bwi.info/Download/de/BWI-Basisdaten/ACCESS2003/.

Declaration

Competing interest No competing interests are declared here.

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). J Evol Biol 25(1):157–173. https://doi.org/10.1111/j.1420-9101.2011.02414.x

Albrecht L (2014) Trauben-oder Stieleiche: Wer ist wer? LWF Aktuell 102:42–43

Blanc-Jolivet C, Bakhtina S, Yanbaev R, Yanbaev Y, Mader M, Guichoux E, Degen B (2020) Development of new SNPs loci on Quercus robur and Quercus petraea for genetic studies covering the whole species' distribution range. Conserv Genet Resour 12(4):597–600. https://doi.org/10.1007/s12686-020-01141-z

Bruegmann T, Fladung M, Schroeder H (2022) Flexible DNA isolation procedure for different tree species as a convenient lab routine. Silvae Genetica 71(1):20–30. https://doi.org/10.2478/sg-2022-0003

Curtu AL, Gailing O, Finkeldey R (2007) Evidence for hybridization and introgression within a species-rich oak (Quercus spp.) community. BMC Evol Biol 7(1):1–15. https://doi.org/10.1186/1471-2148-7-218

Curtu AL, Gailing O, Finkeldey R (2009) Patterns of contemporary hybridization inferred from paternity analysis in a four-oak-species forest. BMC Evol Biol 9(1):1–9. https://doi.org/10.1186/1471-2148-9-284

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. Bioinf 27(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Degen B (2022) GDA-NT 2021–a computer program for population genetic data analysis and assignment. Conserv Genet Resour 14(4):347–350. https://doi.org/10.1007/s12686-022-01283-2

Degen B, Blanc-Jolivet C, Bakhtina S, Ianbaev R, Yanbaev Y, Mader M, Nürnberg S, Schröder H (2021a) Applying targeted genotyping by sequencing with a new set of nuclear and plastid SNP and indel loci for Quercus robur and Quercus petraea. Conserv Genet Resour 13(3):345–347. https://doi.org/10.1007/s12686-021-01207-6

Degen B, Yanbaev Y, Mader M, Ianbaev R, Bakhtina S, Schroeder H, Blanc-Jolivet C (2021b) Impact of gene flow and introgression on the range wide genetic structure of Quercus robur (L.) in Europe. Forests. 12(10):1425. https://doi.org/10.3390/f12101425

Degen B, Blanc-Jolivet C, Mader M, Yanbaeva V, Yanbaev Y (2023) Introgression as an important driver of geographic genetic differentiation within European white oaks. Forests 14(12):2279. https://doi.org/10.3390/f14122279

Delignette-Muller ML, Dutang C (2015) Fitdistrplus: an R package for fitting distributions. J Stat Softw 64:1–34. https://doi.org/10.18637/jss.v064.i04

Ducousso A, Michaud H, Lumaret R (1993) Reproduction and gene flow in the genus Quercus L. Ann For Sci 50:91s–106s. https://doi.org/10.1051/forest:19930708

Dumolin-Lapègue S, Pemonge M, Petit R (1998) Association between chloroplast and mitochondrial lineages in oaks. Mol Biol Evol 15(10):1321–1331. https://doi.org/10.1093/oxfordjournals.molbev.a025860

Eaton E, Caudullo G, Oliveira S, De Rigo D (2016) Quercus robur and Quercus petraea in Europe: distribution, habitat, usage and threats. European atlas of forest tree species. Public Office of the European Union, Publi., pp 162–163

Eusemann P, Liesebach H (2021) Small-scale genetic structure and mating patterns in an extensive sessile oak forest (Quercus petraea (Matt.) Liebl). Ecol Evol 11(12):7796–7809. https://doi.org/10.1002/ece3.7613

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

Fox J, Weisberg S (2018) An R companion to applied regression. USA, Sage publications

Garrison E, Marth G (2016) Haplotype-based variant detection from short-read sequencing. arXiv 1207.3907. https://doi.org/10.48550/arXiv.1207.3907

Gerber S, Chadœuf J, Gugerli F, Lascoux M, Buiteveld J, Cottrell J, Dounavi A, Fineschi S, Forrest LL, Fogelqvist J (2014) High rates of gene flow by pollen and seed in oak populations across Europe. PLoS ONE 9(1):e85130. https://doi.org/10.1371/journal.pone.0085130

Guichoux E, Lagache L, Wagner S, Léger P, Petit R (2011) Two highly validated multiplexes (12-plex and 8-plex) for species delimitation and parentage analysis in oaks (Quercus spp). Mol Ecol Resour 11(3):578–585. https://doi.org/10.1111/j.1755-0998.2011.02983.x

Jensen J, Larsen A, Nielsen LR, Cottrell J (2009) Hybridization between Quercus robur and Q. petraea in a mixed oak stand in Denmark. Ann for Sci 66(7):706–706. https://doi.org/10.1051/forest/2009058

Jevšenak J (2019) Daily climate data reveal stronger climate-growth relationships for an extended European tree-ring network. Q Sci Rev 221:105868. https://doi.org/10.1016/j.quascirev.2019.105868

Kampfer S, Lexer C, Glössl J, Steinkellner H (1998) Characterization of (GA) nmicrosatellite loci from Quercus robur. Hereditas 129(2):183–186. https://doi.org/10.1111/j.1601-5223.1998.00183.x

Kleinn C, Kändler G, Polley H, Riedel T, Schmitz F (2020) The National forest inventory in Germany: responding to forest-related information needs. Allg Forst Jagdzeitg 191(5/6):97–118. https://doi.org/10.23765/afjz0002062

Kleinschmit JR, Bacilieri R, Kremer A, Roloff A (1995) Comparison of morphological and genetic traits of pedunculate oak (Q. robur L.) and sessile oak (Q. petraea (Matt.) Liebl). Silvae Genetica 44(5):256–268

Kändler G (2006) The design of the second German national forest inventory. Proceedings of the Eighth Annual Forest Inventory and Analysis Symposium, Monterey, CA, USA

Kolář T, Čermák P, Trnka M, Žid T, Rybníček M (2017) Temporal changes in the climate sensitivity of Norway Spruce and European Beech along an elevation gradient in central Europe. Agric for Meteorol 239:24–33. https://doi.org/10.1016/j.agrformet.2017.02.028

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

Kremer A, Delcamp A, Lesur I, Wagner S, Rellstab C, Guichoux E, Leroy T (2024) Whole-genome screening for near-diagnostic genetic markers for four Western European white oak species identification. Ann for Sci 81(1):21. https://doi.org/10.1186/s13595-024-01236-9

Kremer A, Dupouey JL, Deans JD, Cottrell J, Csaikl U, Finkeldey R, Espinel S, Jensen J, Kleinschmit J, Van Dam B (2002) Leaf morphological differentiation between Quercus robur and Quercus petraea is stable across Western European mixed oak stands. Ann for Sci 59(7):777–787. https://doi.org/10.1051/forest:2002065

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. Nat Methods 9(4):357–359. https://doi.org/10.1038/nmeth.1923

Lebourgeois F, Bréda N, Ulrich E, Granier A (2005) Climate-tree-growth relationships of European beech (Fagus sylvatica L.) in the French permanent plot network (RENECOFOR). Trees 19:385–401. https://doi.org/10.1007/s00468-004-0397-9

Lepais O, Petit R, Guichoux E, Lavabre J, Alberto F, Kremer A, Gerber S (2009) Species relative abundance and direction of introgression in oaks. Mol Ecol 18(10):2228–2242. https://doi.org/10.1111/j.1365-294X.2009.04137.x

Leroy T, Louvet JM, Lalanne C, Le Provost G, Labadie K, Aury JM, Delzon S, Plomion C, Kremer A (2020a) Adaptive introgression as a driver of local adaptation to climate in European white oaks. New Phytol 226(4):1171–1182. https://doi.org/10.1111/nph.16095

Leroy T, Plomion C, Kremer A (2020b) Oak symbolism in the light of genomics. New Phytol 226(4):1012–1017. https://doi.org/10.1111/nph.15987

Leroy T, Roux C, Villate L, Bodénès C, Romiguier J, Paiva JA, Dossat C, Aury JM, Plomion C, Kremer A (2017) Extensive recent secondary contacts between four European white oak species. New Phytol 214(2):865–878. https://doi.org/10.1111/nph.14413

Liesebach H, Eusemann P, Höltken AM, Tröber U, Kuchma O, Karopka M, Becker F, Kätzel R, Fussi B (2024) Effective population size of adult and offspring cohorts as a genetic monitoring tool in two stand-forming and wind-pollinated tree species: Fagus sylvatica L. and Picea abies (L.) Karst. Conserv Genet 25:739–753. https://doi.org/10.1007/s10592-024-01600-2

Li YL, Liu JX (2018) StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour 18(1):176–177. https://doi.org/10.1111/1755-0998.12719

Loiselle B, Sork V, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). Am J Bot 82(11):1420–1425. https://doi.org/10.1002/j.1537-2197.1995.tb12679.x

Marchi M, Castellanos-Acuña D, Hamann A, Wang T, Ray D, Menzel A (2020) ClimateEU, scale-free climate normals, historical time series, and future projections for Europe. Sci Data 7(1):428. https://doi.org/10.1038/s41597-020-00763-0

Neophytou C, Fussi B, Konnert M, Luckas M (2014) Traubeneiche und Stieleiche–zwei ungleiche Schwestern. LWF Wissen 75:14–20

Neophytou C, Gärtner SM, Vargas-Gaete R, Michiels HG (2015) Genetic variation of central European oaks: shaped by evolutionary factors and human intervention? Tree Genet Genomes 11:1–15. https://doi.org/10.1007/s11295-015-0905-7

Neophytou C, Michiels HG (2013) Upper Rhine Valley: A migration crossroads of middle European oaks. For Ecol Manag 304:89–98. https://doi.org/10.1016/j.foreco.2013.04.020

Niemczyk M, Chmura DJ, Socha J, Wojda T, Mroczek P, Gil W, Thomas BR (2021) How geographic and climatic factors affect the adaptation of Douglas-fir provenances to the temperate continental climate zone in Europe. Eur J for Res 140(6):1341–1361. https://doi.org/10.1007/s10342-021-01398-5

Olrik DCED, Kjaer (2007) The reproductive success of a Quercus petraea X Q. robur F1-hybrid in back-crossing situations. Ann for Sci 64(1):37–45. https://doi.org/10.1051/forest:2006086

Peakall R, Smouse P (2012) GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. Bioinformatics 28(19):2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Petit RJ, Bodénès C, Ducousso A, Roussel G, Kremer A (2004) Hybridization as a mechanism of invasion in oaks. New Phytol 161(1):151–164. https://doi.org/10.1046/j.1469-8137.2003.00944.x

Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, van Dam B, Deans JD (2002a) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. For Ecol Manage 156(1–3):49–74. https://doi.org/10.1016/S0378-1127(01)00634-X

Petit RJ, Csaikl UM, Bordács S, Burg K, Coart E, Cottrell J, van Dam B, Deans JD, Dumolin-Lapègue S, Fineschi S (2002b) Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. For Ecol Manage 156(1–3):5–26. https://doi.org/10.1016/S0378-1127(01)00645-4

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype. Data Genet 155(2):945. https://doi.org/10.1093/genetics/155.2.945

Pritchard JK, Wen X, Falush D (2010) Documentation for structure software: Version 2.3. Univ. Chicago 10

R Core Team (2013) R: a language and environment for statistical computing

Ramasamy RK, Ramasamy S, Bindroo BB, Naik VG (2014) STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. SpringerPlus 3(1):1–3. https://doi.org/10.1186/2193-1801-3-431

Reutimann O, Gugerli F, Rellstab C (2020) A species-discriminatory single-nucleotide polymorphism set reveals maintenance of species integrity in hybridizing European white oaks (Quercus spp.) despite high levels of admixture. Ann Botany 125(4):663–676. https://doi.org/10.1093/aob/mcaa001

Riedel T, Hennig P, Polley H, Schwitzgebel F (2020) Aufnahmeanweisung für die vierte Bundeswaldinventur (BWI 2022) 1. edition, November 2020 (Version 1.11)

Rigby RA, Stasinopoulos DM (2005) Generalized additive models for location, scale and shape. J Royal Stat Soc Ser C: Appl Stat 54(3):507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Schroeder H, Kersten B (2023) A small set of nuclear markers for reliable differentiation of the two closely related oak species Quercus robur and Q. petraea Plants 12(3):566. https://doi.org/10.3390/plants12030566

Schroeder H, Cronn R, Yanbaev Y, Jennings T, Mader M, Degen B, Kersten B (2016) Development of molecular markers for determining continental origin of wood from white oaks (Quercus L. Sect. Quercus). PloS one 11. 6e0158221. https://doi.org/10.1371/journal.pone.0158221

Steinhoff S (1998) Kontrollierte Kreuzungen zwischen Stiel-und Traubeneiche:Ergebnisse und Folgerungen. Allgemeine Forst- und Jagdzeitung 169(9):163–168

Streiff R, Ducousso A, Lexer C, Steinkellner H, Gloessl J, Kremer A (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of Quercus robur L. and Q. petraea (Matt.) Liebl. Mol Ecol 8(5):831–841. https://doi.org/10.1046/j.1365-294X.1999.00637.x

Thuiller W (2024) Ecological niche modelling. Curr Biol 34(6):R225. https://doi.org/10.1016/j.cub.2024.02.018

Viscosi V, Lepais O, Gerber S, Fortini P (2009) Leaf morphologicalanalyses in four European oak species (Quercus) and their hybrids: a comparisonof traditional and geometric morphometric methods. Plant Biosyst 143(3):564–574. https://doi.org/10.1080/11263500902723129

Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol 1(1):3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x