Supplementary 2: Genetic identification of *Quercus robur*, *Quercus petraea* and hybrids
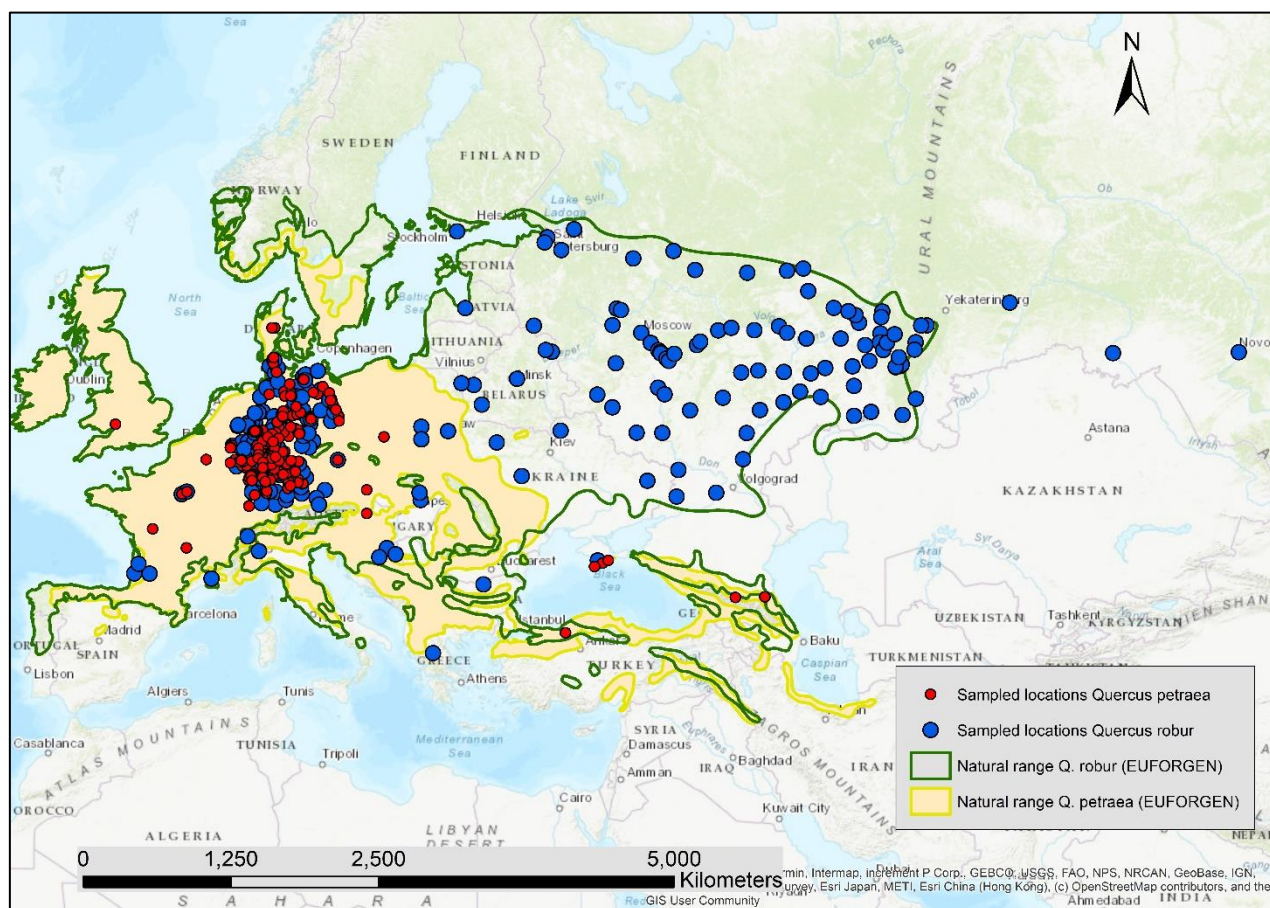
The 1,970 pedunculate oak trees of the presented study were part of a larger data set including in total 4,848 *Quercus robur* and *Quercus petraea* trees. The samples were collected at 440 locations distributed over both species' ranges (Figure S2 1).


Figure S2 1: Distribution of 440 sampled locations and *a priori* species classification

The majority of the samples (>95%) were collected in forest stands and a small proportion was sampled in provenance trials. The sampled trees in the stands had a minimum diameter at breast height of 20 cm and a minimum distance to the next sampled tree of 50 m. The samples at each location were classified according to the registry of the responsible forest administration as *Quercus robur* or *Quercus petraea*. In the further text we call this the "*a priori* species classification".

We genotyped all trees with the same set of 359 nuclear loci as in the present study. On average the individuals had less than 1% missing data. Using the genotypes at the 359 nuclear loci of all 4,848 individuals we did a principal component analysis with the program PAST [1]. The first two principal components (Figure S2 2) showed a clear subdivision of the sampled locations into two

groups. The groups fit in most cases with the *a priori* classification to the two species *Q. robur* and *Q. petraea*. Within each group there was a sub-structure and we observed individuals between both groups.
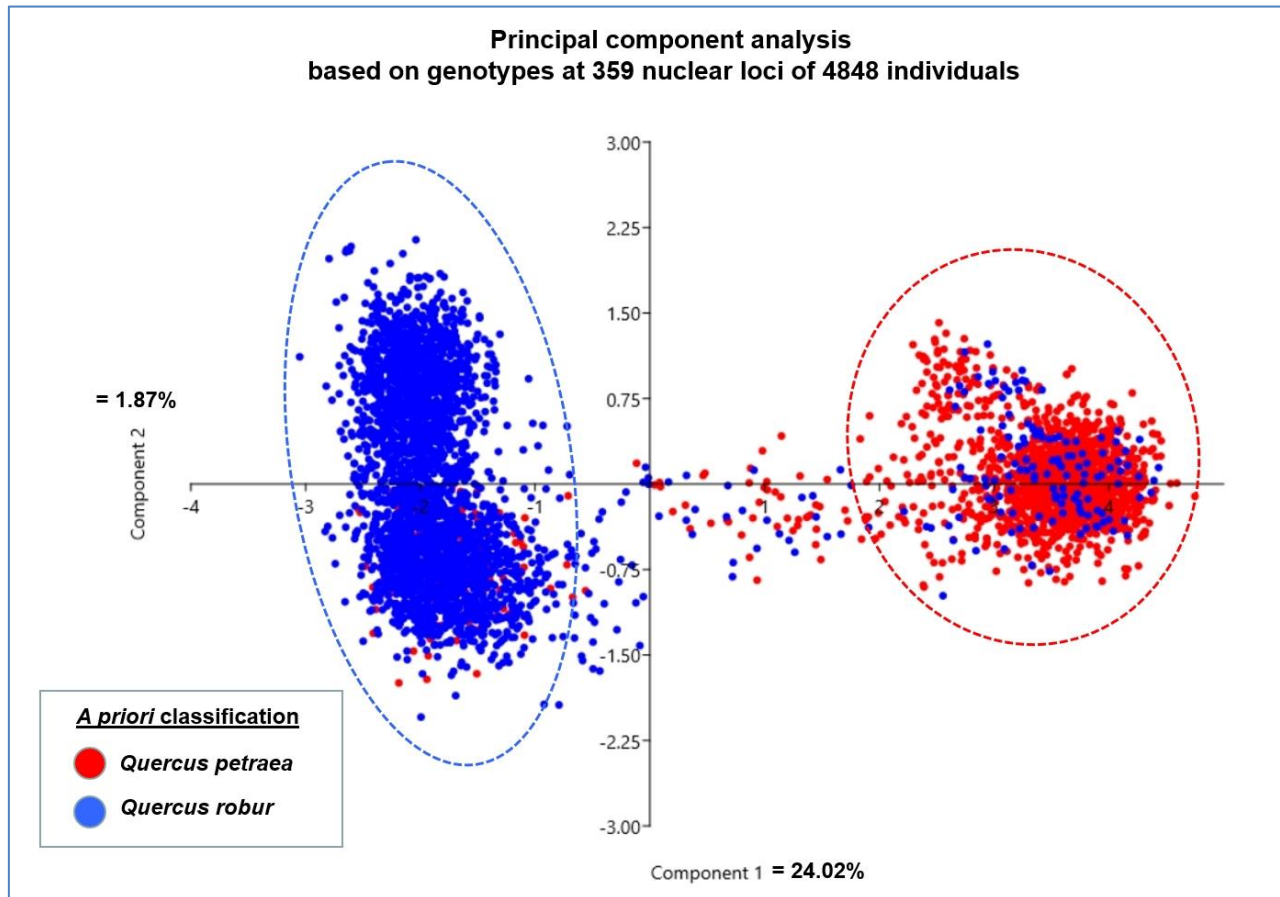
**Figure S2 2**: Principal component analysis with the 4,848 genotypes at 359 nuclear loci

Based on the results of the principal component analysis we used the Bayesian clustering method implemented in the software STRUCTURE v.2.3.4 [2] to check the admixture of all 4,852 individuals for K=2. We set the length of burn-in and Markov chain Monte Carlo simulations to 15,000 and 100,000, respectively. We ran the admixture model with correlated allele frequencies. The data from ten STRUCTURE runs were summarized and graphically presented to obtain the final admixture values for each individual with the program CLUMPAK [3].

The comparison of *a priori* classification and the result of the STRUCTURE analysis is given in table 1. For *Q. robur* 94% of all individuals that were *a priori* classified as *Q. robur* had an admixture value for *Q. robur* between 0.8 and 1. For the *a priori* classified *Q. petraea* this was true for 84%. Four percent of the declared *Q. robur* were genetically assigned to *Q. petraea*, and 11% of the *Q. petraea* were assigned to *Q. robur*. We assume that this is in most cases linked to errors in the records of the forest administration. In total 2.5% of all individuals had an admixture value between 0.2 and 0.8, and thus could be seen as hybrids.

For the present study on the range wide spatial genetic structure of pedunculate oak only individuals with an admixture of *Q. robur* of at least 0.8 were included. For those individuals the average proportion of admixture from the *Q. petraea* gene pool at the sampled 197 locations showed a clear spatial pattern with higher values of admixtures in the west and north-west and absence of admixture in the east of Europe and the southeast of Germany (Figure S2 3).

Table S2 1: Comparison for STRUCTURE results (genetic membership coefficient for K=2) and the *a priori* species declaration

Genetic membership-coefficient of declared species	*A priori Quercus robur*		*A priori Quercus petraea*	
	Absolute	Proportion	Absolute	Proportion
0.95-1.00	2,699	0.874	1,299	0.737
0.95-0.80	212	0.069	188	0.106
0.20-0.80	37	0.012	86	0.049
0.00-0.20	137	0.044	190	0.108
Total	3,085	1.000	1,763	1.000
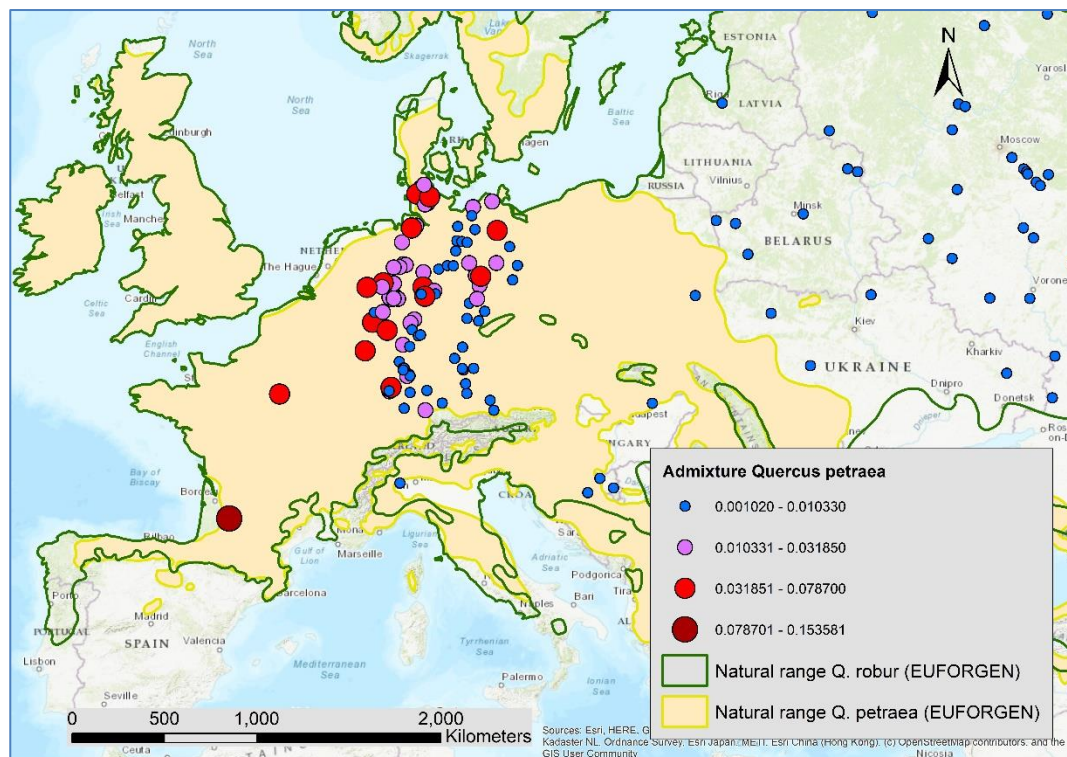
Figure S2 3: Proportion of average admixture with the gene pool from *Q. petraea* for pedunculate oak individuals included in the present study

References

1. Hammer, Ø.; Harper, D.A.; Ryan, P. Paleontological statistics software package for education and data analysis. *Paleontologia Electronica* **2001**, *4*, 1-9.

2. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945-959.

3. Kopelman, N.M.; Mayzel, J.; Jakobsson, M.; Rosenberg, N.A.; Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **2015**, *15*, 1179-1191, doi:10.1111/1755-0998.12387.