Latitudinal trends in genetic diversity and distinctiveness of *Quercus robur* rear edge forest remnants call for new conservation priorities

Camilla Avanzi[1], Francesca Bagnoli[1], Edoardo Romiti[1,2], Ilaria Spanu[1], Yoshiaki Tsuda[3], Elia Vajana[1], Giovanni Giuseppe Vendramin[1], Andrea Piotti[1]

[1] Institute of Biosciences and Bioresources, National Research Council of Italy, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Italy

[2] Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Viale delle Idee 30, I-50019 Sesto Fiorentino, Italy

[3] Sugadaira Montane Research Center, University of Tsukuba, 1278-294 Sugadaira-Kogen, Ueda, Nagano 386-2204, Japan

Abstract

Knowledge of the spatial distribution of intraspecific genetic variation is essential for planning conservation actions, designing networks of protected areas, and informing possible assisted migration strategies. Although the Italian peninsula harbours unique genetic variation as a legacy of Quaternary migrations, only sporadic genetic information is available for forest tree species.

Here, we present the first geographically comprehensive genetic characterisation of *Quercus robur*, an iconic broadleaved species of European floodplain forests, in an area which acted as a primary glacial refugium for the species. 745 individuals from 25 populations were sampled and genotyped with 16 microsatellite markers. Their genetic structure was assessed through various metrics of diversity and distinctiveness, as well as by Bayesian clustering and multivariate methods. The demographic history of inferred gene pools was evaluated through Approximate Bayesian Computation analysis.

Genetic distinctiveness showed a decline with increasing latitude, while allelic richness reached its peak in central Italy. A south-to-north trend in the complexity of the genetic structure was observed, with peninsular Italy being characterised by intermingled gene pools in contrast to the relative homogeneity exhibited by northern populations. Demographic inference indicated that the southern gene pool has been genetically isolated since the penultimate interglacial, whereas populations from central Italy persisted locally in a mosaic of small refugia during the last glaciation.

Our results brought out the complexity of the genetic structure of forest trees' populations in southern Europe. At least three *Q. robur* refugia contributed to the genetic layout of extant populations after the last glacial period, but refugial areas were probably even more numerous in central Italy. Such a detailed characterization sheds new light on the priorities to be established for the conservation of highly fragmented *Q. robur* populations in an area rich in diversified genetic lineages.

Keywords

Approximate Bayesian Computation, bottleneck, conservation genetics, genetic structure, intraspecific genetic diversity, pedunculate oak, white oaks

Introduction

The DNA-encoded heritable variation within and between populations of a species (i.e. intraspecific genetic diversity) provides species with the ability to adapt and evolve (Hoban et al. 2020) and influences ecological processes, with significant effects on community structure, primary production, nutrient cycling, and ecosystem resilience (Hughes et al. 2008). In the context of the current global climatic crisis, adequate levels of genetic diversity are the essential prerequisite for species to keep pace with rapidly changing environmental conditions and the spread of new pests and diseases (Sgrò et al. 2011), as well as for ecosystems to keep providing ecological functions and services (Des Roches et al. 2021).

Despite broad scientific *consensus* on the urgency of conserving intraspecific genetic diversity, and hence adaptive potential (DeWoody et al. 2021), conservation strategies often provide vague guidance on how to translate this goal into practical action, particularly at national and subnational levels (Pierson et al. 2016; Cook and Sgrò 2017; Laikre et al. 2020). Hoban et al. (2020) proposed three pragmatic indicators for genetic monitoring, based on the number of populations with an effective population size (*Ne*) above the standard (but controversial) threshold of 500, the proportion of populations conserved for each species and, finally, the number of populations in which genetic diversity is assessed (and tracked) using molecular markers. Indeed, direct knowledge of the spatial distribution of intraspecific genetic diversity and the degree of genetic connectivity among populations is particularly useful for informing translocation strategies and/or designing optimal networks of protected areas (Aitken and Whitlock 2013; Fady et al. 2016; Andrello et al. 2022). Despite the fact that range-wide surveys of the genetic variation of European forest tree species have been carried out for a long time (e.g. Petit et al. 2003; Vendramin et al. 2008; Postolache et al. 2021; Milesi et al. 2023), southern Europe has frequently been under-sampled, despite its disproportionate importance as a long-term repository of genetic diversity (Hampe and Petit 2005). Southern Europe likely hosted a mosaic of refugial areas for tree species during the cold phases of the Quaternary (Tzedakis et al. 1997), resulting in southern European populations often exhibiting complex spatial patterns in the distribution of genetic diversity (e.g. Piotti et al. 2017; Scotti-Saintagne et al. 2021). Robust sampling within putative refugial areas is essential to quantitatively describe such complexity (Piotti et al. 2017) and, in turn, to design effective conservation strategies of extant forest genetic resources (de Vries et al. 2015).

Pedunculate oak (*Quercus robur* L.) is an iconic broadleaved tree species key for the European lowland ecosystems, ranging from Mediterranean peninsulas to Scandinavia, from Scotland to the Ural and Caucasus mountain ranges (Eaton et al. 2016). Although *Q. robur* may appear to be distributed across a continuous range according to the available distribution maps (Caudullo et al. 2017), it regionally suffers from a high degree of fragmentation, as lowland forests have been increasingly replaced by agricultural and urban landscapes in large parts of the species' distribution (Petit et al. 2002c; Vakkari et al. 2006; Belletti et al. 2007; Agnoletti et al. 2018; Degen et al. 2021). In recent decades, and particularly in rear edge populations, fragmentation has been exacerbated by episodes of decline and mortality, fuelled by a complex interplay of

changing biotic and abiotic conditions (Gibbs 1997; Ragazzi et al. 2000; Thomas et al. 2002; de Sampaio e Paiva Camilo-Alves et al. 2013; Stojanović et al. 2015; Gathercole et al. 2021). In this context, a comprehensive and accurate study of the genetic layout of *Q. robur* populations at the rear edge of the species' distribution within a putatively multi-refugial area, such as the Italian peninsula, is still lacking. To our knowledge, despite a few studies at regional scale (Belletti et al. 2005; Ducci et al. 2007), only Fineschi et al. (2002) investigated the genetic diversity of *Q. robur* populations throughout Italy, as part of a larger study of chloroplast DNA variation of European white oaks (Petit et al. 2002a, b). However, this study relied on a very limited *Q. robur* sample size (64 trees), with only a few samples from peninsular Italy. Nevertheless, there is evidence that this area hosted multiple glacial refugia for several forest tree species (Magri et al. 2015) and that it was relevant for post-glacial demographic history of white oaks in Europe (Petit et al. 2002a). Sparse genetic information and the impossibility to distinguish *Quercus* pollen at the species level prevent testing specific hypotheses on *Q. robur* recent demographic dynamics and call for a deep investigation on the genetic structure of the species, and the major events influencing it, in this area.

Our work is the first attempt to comprehensively characterise the forest genetic resources of pedunculate oak throughout Italy. To this end, 25 populations were carefully selected, sampled and genotyped with 16 microsatellite markers. The specific aims of this study were to: i) assess their levels of genetic diversity and distinctiveness, ii) quantify the contribution of each population to the total allelic diversity and its components within and between populations, iii) identify which populations have limited effective population size and show genetic signatures of bottlenecks, iv) decipher the extant genetic structure of the species along the Italian peninsula, and v) date the divergence of extant gene pools. As peninsular Italy is characterised by greater topographic heterogeneity than the Po Valley, and it hosted multiple forest trees' refugial areas, genetic diversity and distinctiveness, as well as the complexity of genetic structure, are expected to decrease with increasing latitude. Uncovering hotspots of genetic diversity and distinctiveness, as well as the existence of potential gene flow constraints among oak forest remnants, will be essential to inform management and conservation actions aimed at mitigating the negative effects of fragmentation genetic drift in such an economically and ecologically important tree species.

Materials and methods

Sampling activities and genotyping

Twenty-five populations of *Quercus robur* L. were sampled throughout Italy (Figure 1). In each population, with few exceptions due to extremely small natural stands (Table S1), leaves were collected from ~30 individuals spaced at least 25 meters to avoid sampling related trees, reaching a total of 745 sampled trees. Each individual was tagged and georeferenced using a GPS tracker. Fresh leaves were dried in silica gel and stored at -20 °C until DNA extraction. To be sure that we had not mistakenly sampled individuals of other

109 white oak species, we also collected leaves from 46, 26 and 8 trees of *Q. petraea*, *Q. pubescens* and *Q.*
110 *frainetto*, respectively (Figure S1a) for successive species discrimination based on the same set of molecular
111 markers.
112 DNA was extracted using the NucleoSpin 96 Plant II Kit (Macherey-Nagel, Germany), following the
113 manufacturer's instructions. Approximately 30 mg of frozen leaves and a 3-mm tungsten bead were added
114 to each well of 96-well plates for grinding leaf material. Then, plates were frozen at -80 °C before two cycles
115 of 30-second disruption at 27 Hz using a Mixer Mill MM300 (Retsch, Germany). All individuals were
116 genotyped with 16 nuclear microsatellite markers (SSRs) (PIE020, PIE223, PIE242, PIE102, PIE239, PIE227,
117 PIE271, PIE267, PIE215: Durand et al. 2010; MsQ13: Dow et al. 1995; QrZAG7, QrZAG112, QrZAG20, QrZAG96:
118 Kampfer et al. 1998; QpZAG15, QpZAG110: Steinkellner et al. 1997). All 16 SSRs were arranged into a 9- and
119 7-plex by slightly modifying the 12-plex and 8-plex developed by Guichoux et al. (2011). Multiplex reactions
120 were carried out with the Type-it Microsatellite PCR Kit (Qiagen, Germany). The final volume of the PCR
121 reactions was optimised to 6 µl. The PCR mix was composed of 3 µl Type-it Multiple PCR Master Mix, 0.6 µl
122 of primer premix, 1.4 µl of H_2O and 1 µl of DNA. The cycling conditions indicated by Guichoux et al. (2011)
123 were used for the 9-plex, while they were slightly changed for the 7-plex, with an initial step at 95°C for 5
124 min, followed by 32 cycles at 95°C for 30 s, 56°C for 90 s and 72°C for 45 s, and a final 30 min extension step
125 at 60°C. PCR products were run on an ABI 3500 automatic sequencer (Applied Biosystems, USA), with LIZ-500
126 (ThermoFisher Scientific) as internal size standard. The obtained results were sized using GeneMarker
127 (Softgenetics).
128
129 *Quality control of the genetic dataset*
130 In some populations of the present study, *Quercus robur* was in close proximity with other related oak species
131 (i.e. *Q. petraea*, *Q. pubescens*, *Q. frainetto*). Although some 'easy-to-use' discriminant morphometric traits
132 were used during fieldwork to identify and sample *Q. robur* trees, molecular markers were subsequently used
133 to further exclude the presence of individuals 'misclassified' as *Q. robur*, which would bias estimates of
134 genetic diversity and structure. Our set of SSRs was indeed proven to discriminate well European white oaks
135 (Guichoux et al. 2011). Such a check was done by applying the Bayesian clustering approach implemented in
136 STRUCTURE (Pritchard et al. 2000), with sampling locations as prior information (Hubisz et al. 2009). Thirty
137 independent runs were run for each value of *K* (the number of putative gene pools), ranging from two to four
138 (i.e. the number of white oak species). Each run consisted of 10,000 burn-in iterations and 50,000 sampling
139 iterations. Runs were aligned and averaged using the software CLUMPAK (Kopelman et al. 2015). Individuals
140 with a probability of belonging to one of the other white oak species >0.8 were considered 'misclassified' and
141 excluded from any further analysis. The Bayesian clustering approach was complemented by a multivariate
142 analysis, i.e. a Principal Coordinate Analysis (PCoA) performed on the matrix of pairwise mean among-
143 population genetic distances (Smouse and Peakall 1999) using GenAlEx (Peakall and Smouse 2012).

144 The presence of null alleles and genotyping failures in the final dataset (i.e. the dataset without 'misclassified'
145 individuals) was assessed at the population level by using INest v2.2 (Chybicki and Burczyk 2009). The model
146 was run using a Gibbs sampler based on 200,000 iterations, with a burn-in of 20,000 and retaining one
147 iteration every 200.
148 To ensure the neutrality of the marker set, first 10,000 coalescent simulations were performed in the Arlequin
149 v3.5.2 environment (Excoffier and Lischer 2010) to get the p-values of locus-specific F-statistic conditioned
150 on the observed levels of heterozygosity (Beaumont and Nichols 1996; Excoffier et al. 2008). A non-
151 hierarchical finite island model was used, with all populations pooled into a single group. Secondly, the
152 hierarchical Bayesian approach implemented in BayeScan v2.1 (Foll and Gaggiotti 2008) was used to estimate
153 the posterior probabilities of two alternative models, one including the effect of selection and one excluding
154 it. The ratio of the posterior probabilities of the two models can be translated, for each locus, into a different
155 level of selection intensity according to Jeffreys' interpretation. BayeScan was run on demographically stable
156 populations identified based on the results of a bottleneck analysis carried out with INest (for details see the
157 'Demographic parameters' paragraph) with 1000 pilot runs of 10,000 iterations each, and a burn-in of
158 500,000, a sample size of 10,000 and a thinning interval of 20. Prior odds for the selection model were set to
159 100.
160
161 *Genetic structure*
162 The Bayesian clustering approach implemented in STRUCTURE was used to identify the most likely number
163 of gene pools (*K*). The analysis was based on the LOCPRIOR model (Hubisz et al. 2009), as well as on the
164 correlated allele frequencies model (Falush et al. 2003). Thirty independent runs were performed for each
165 value of *K*, ranging from 1 to 10. Each run consisted of 10,000 burn-in iterations and 50,000 data collection
166 iterations. These settings for the total length of the MCMC were sufficient to obtain reliable genetic
167 structures by comparing the results of initial runs with those involving longer chains. Different runs for the
168 same *K* were then averaged using the software CLUMPAK. The most likely grouping scenarios were evaluated
169 by using ΔK statistic (Evanno et al. 2005), as well as by a visual assessment of the log-likelihood trend across
170 the various tested scenarios. Both ΔK and log-likelihood values were calculated by using the web-based
171 software StructureSelector (Li and Liu 2018).
172 The genetic structure of Italian oak populations was also investigated with a multivariate approach, running
173 a PCoA on the matrix of pairwise among-population genetic distance (Smouse and Peakall 1999) with
174 GenAlEx. Pairwise G_{ST} values (Meirmans and Hedrick 2011) of genetic differentiation among populations were
175 also calculated with GenAlEx, testing their significance by 999 permutations.
176
177 *Genetic diversity and distinctiveness*

178 Number of alleles (*Na*), as well as the observed (H_O) and expected heterozygosity (H_E), were calculated for
179 each population using GenAlEx. The inbreeding coefficient (F_{IS}) of each population was assessed by using
180 INEST. Allelic richness (*Ar*) and private allelic richness (*PAr*) were calculated by using HP-RARE (Kalinowski
181 2005). Rarefaction was applied based on a minimum sample size of 44 genes. The contribution of each
182 population to the total allelic diversity (A_T), together with its within-subpopulation (A_S) and among-
183 subpopulation (D_A) components, was assessed using METAPOP2 (López-Cortegano et al. 2019).
184 The genetic distinctiveness of each population was estimated using a population-specific index of
185 differentiation (F_{ST}) between the target population and a common migrant gene pool from which it differs in
186 varying degrees, using BayeScan v2.1. F_{ST} coefficients were estimated through 20 pilot runs of 10,000
187 iterations each, followed by a final MCMC made up of a burn-in of 50,000 and further 50,000 iterations with
188 a thinning of 20.
189 Linear regression was used to explore the relationship between genetic parameters (i.e. Na, H_O, H_E, F_{IS}, *Ar,*
190 *PAr,* A_T, A_S, D_A, F_{ST}) and latitude. To account for non-linear relationships, a quadratic term of latitude was
191 tested in the models. Global *P*-values were adjusted for multiple testing using the Bonferroni correction.
192 Regression models and multiple testing correction were performed using the *lm* and *p.adjust* R functions
193 from the *stats* package (R Core Team 2023)
194
195 *Demographic parameters*
196 Effective population size (*Ne*) was estimated using the linkage disequilibrium method implemented in
197 NeEstimator v2.1 (Do et al. 2014). In forest tree species, due to their peculiar life-history traits, reliable
198 estimates of *Ne* are expected only for isolated and small populations (typically less than 500 reproductive
199 trees) (Santos-del-Blanco et al. 2022). In our dataset, the populations that met such requirements are shown
200 in Table S1.
201 The occurrence of past genetic bottlenecks within each population was evaluated by searching both for an
202 excess of heterozygosity (*H*; Cornuet and Luikart 1996) and a deficiency in the ratio between the number of
203 alleles and the range of allele size (MR; Garza and Williamson 2001) with INEST. The values of *H* and MR
204 expected at equilibrium (H_{eq} and MR_{eq}) were computed with coalescent simulations, assuming a two-phase
205 mutation model with a proportion and average size of multi-step mutations of 0.22 and 3.1, respectively, as
206 recommended by Peery et al. (2012). H_{eq} and MR_{eq} were then tested against observed *H* and MR values using
207 a one-tailed Wilcoxon signed-rank test with 1,000,000 permutations (no normality assumption required).
208
209 *Approximate Bayesian Computation to trace back the divergence among oak gene pools*
210 To infer the past demographic history of Italian *Quercus robur* populations, the Approximate Bayesian
211 Computation (ABC) method implemented in the DIYABC v2.1 software (Cornuet et al. 2008, 2014) was
212 applied. To limit the number of tested scenarios, the genetic information obtained at *K*=3 from the Bayesian

213 clustering analysis was used to guide the ABC analysis. Therefore, three populations were defined: NI
214 (northern Italy), CI (central Italy), and SI (southern Italy) (see Results, Figure 3b). NI and CI were assembled
215 by randomly selecting 60 individuals from the populations assigned to the relative gene pool with a
216 membership ≥0.7, while SI consisted of the entire PAN population (61 individuals). After some pilot runs to
217 determine which was the ancestral population (Appendix 1) and the topology of the split scenarios (results
218 not shown), four demographic scenarios were compared (Figure 2a). In all tested scenarios, t# refers to the
219 time scale measured in generation time, N# stands for the effective population sizes of the corresponding
220 gene pool (i.e. NI, CI, SI, ancestral populations 'a', and changes in the effective population seize 'b'), and r#
221 represents the admixture ratio. In principle, DIYABC does not consider the hypothesis of post-split gene flow
222 within scenarios, but a recent study modified DIYABC to include symmetric admixture between two
223 populations (Chapuis et al. 2020). Kitamura et al. (2022) extended this approach to model directional
224 admixture (e.g. from NI to CI and from CI to NI), which could be considered analogous to gene flow. Thus, the
225 parameter 'r#' related to the admixture rate between two generic populations *a* and *b* can be considered as
226 the relative amount of gene flow at time 't#', while '1-r#' represents the value of the relative amount of gene
227 flow within the reference parent population.
228 All scenarios were based on a *hierarchical split model* which, as suggested by the results of the pilot runs
229 (Appendix 1), assumed that the SI gene pool could be traced back to the ancestral population (Na). The model
230 also predicted a split between SI and the gene pools further north at t6, and a divergence between NI and CI
231 at t5. At t5, changes in population size on both branches of the tree (Na and Nb) were also modelled. The
232 differences between the four scenarios were based on patterns of 'gene flow' (*sensu* Kitamura et al. 2022)
233 modelled using a 'stepping stone' approach, i.e. gene flow was only possible between geographically close
234 gene pools (e.g. from SI to CI, but not from SI to NI):
235 - in *Scenario 1*, gene flow was bidirectional, from north to south (i.e. rb from NI to CI at t2 and rc from
236 CI to SI at t3) and from south to north (i.e. rd from SI to CI at t4 and ra from CI to NI at t1);
237 - in *Scenario 2*, gene flow was unidirectional, from south to north (i.e. rd from SI to CI at t4 and ra from
238 CI to NI at t1);
239 - in *Scenario 3*, gene flow was unidirectional, from north to south (i.e. rb from NI to CI at t2 and rc from
240 CI to SI at t3);
241 - in *Scenario 4*, there was gene flow converging to central Italy, from north (i.e. rb from NI to CI at t2)
242 and from south (i.e. rd from SI to CI at t4).
243 The Generalized Stepwise Mutation model (GSM; Estoup et al. 2002) with Single Nucleotide Indels (SNI) was
244 used for all ABC simulations. The default priors were modified to obtain better posterior distributions based
245 on the results of the pilot runs (Table S2). The minimum and maximum priors for SSR mutation rate were set
246 to 1×10^{-4}–1×10^{-3}. The mean number of alleles, mean genetic diversity, mean allele size variance and mean
247 Garza-Williamson's M were used as summary statistics for individual populations. For each population pair,

248 mean number of alleles, mean genetic diversity, mean allele size variance, F_{ST}, and mean individual
249 assignment likelihood were calculated. One million simulations were run for each scenario. After a total of
250 four million simulations, the most likely scenario was assessed by comparing posterior probabilities, using
251 logistic regression. Goodness-of-fit was estimated for each scenario by model checking using the Principal
252 Component Analysis (PCA) approach implemented in DIYABC, which measures the discrepancy between
253 simulated and observed data.
254
255 **Results**
256
257 *Quality control of the genetic dataset*
258 The Bayesian clustering approach used to detect potential 'misclassified' *Q. robur* individuals showed that *Q.*
259 *petraea*, *Q. pubescens* and *Q. frainetto* trees clustered together at K = 2; at K = 3 the southern population of
260 PAN emerged as a distinct gene pool, while at K = 4, central Italy populations showed distinctive genetic
261 features (Figure S1c). *Q. petraea*, *Q. pubescens* and *Q. frainetto* individuals still clustered together at K = 3
262 and K = 4 but were discriminated against the second axis of the Principal Coordinate Analysis (PCoA) (Figure
263 S1b). Based on STRUCTURE results at K = 4, 16 individuals were identified as 'misclassified' trees and were
264 excluded from the final dataset. They belonged to four populations: FON (one single tree), MON (ten trees),
265 FBN (two trees) and PAN (three trees) (see Table S1).
266 No marker showed consistent and/or significant issues related to genotyping failure and null alleles (Figure
267 S2) and significant signals of selection based on the coalescent simulations performed within the Arlequin
268 environment (Table S3). The Bayesian approach implemented in BayeScan showed a signature of balancing
269 selection for QrZag112, with a F_{ST} value close to zero (Table S4). As the genetic structure and allelic richness
270 with and without this marker was almost identical (results not shown), and the locus was particularly
271 informative for detecting past demographic dynamics (it had the highest number of alleles), QrZag112 was
272 retained in the final dataset.
273
274 *Genetic structure*
275 According to ΔK statistic, the Bayesian clustering analysis revealed an optimal grouping at K = 2 (Figure 3d),
276 which clearly distinguished a gene pool represented by the southernmost population of PAN from another
277 gene pool represented by all the other Italian populations (Figure 3a). Mean pairwise G_{ST} value between PAN
278 and all other populations was 0.044 (Table S5). Based on ΔK, the next strongest level of structuring was K = 3
279 (Figure 3d), which separated populations from central Italy from those from northern Italy (Figure 3b).
280 Although ΔK was the highest when K = 2, the mean Ln P(K) increased with K, reaching its maximum at K = 6
281 (Figure 3d). According to this scenario, the complexity of the genetic structure in central Italy increased while
282 populations from northern Italy mostly kept on clustering together (Figure 3c). Mean r values (averaged over

283 runs for the same *K* value) were equal to 0.71, 0.45 and 0.54 when analysing two, three and six gene pools,
284 respectively, indicating that the sampling locations were informative for unravelling the genetic structure.
285 The results of the Principal Coordinate Analysis (PCoA) on the matrix of pairwise genetic distance among
286 populations confirmed the genetic structure observed with the Bayesian clustering approach (Figure S3). The
287 first three coordinates explained 60.3% of the total variance of the SSRs data. The southernmost population
288 of PAN was well separated from all the others both on the first and second PCo axes. FAB population showed
289 a clear separation from all the other populations on the third PCo axis. Pairwise indexes of genetic
290 differentiation among populations (G_{ST}) were reported in Table S5.
291
292 *Genetic diversity and distinctiveness*
293 The 16 SSRs showed a total and mean number of alleles of 246 and 15.4, respectively. Estimates of population
294 genetic diversity and distinctiveness are reported in Table 1. *Ar* showed a quadratic dependence on latitude
295 (*lat*: F-value = 5.69, $P = 0.026$; lat^2: F-value = 8.88, $P = 0.007$), with populations from central Italy showing the
296 highest *Ar* values (Figure 4a). Genetic distinctiveness (population-specific F_{ST}) had a negative relationship with
297 latitude (*lat*: F-value = 53.65, $P = 2.474\times10^{-7}$; lat^2: F-value = 16.86, $P = 4.65\times10^{-4}$; Figure 4b). Na, *PAr*, H_O, H_E,
298 and F_{IS} showed no significant relationship with latitude (Figure S4).
299 The population that contributed the most to the total allelic diversity was ORV ($A_T = 0.618$) while the one that
300 contributed the least was FAB (A_T = -0.500) (Figure 5). When within-subpopulation (A_S) and among-
301 subpopulation (D_A) components of A_T were disentangled, ORV still had the largest A_S value (0.461), while PAN
302 had the lowest value of A_S (-0.948) but the largest value of D_A (0.593) (Figure 5). As expected considering
303 their collinearity with population-specific F_{ST} and *Ar*, respectively, D_A exhibited a linear negative relationship
304 with latitude (*lat*: F-value = 90.27, $P = 3.038\times10^{-9}$), while A_S a quadratic dependence on latitude (*lat*: F-value
305 = 5.68, $P = 0.026$; lat^2: F-value = 8.89, $P = 0.007$), with populations from central Italy showing the highest
306 values (Figure S4). A_T showed no significant relationship with latitude (Figure S4).
307
308 *Demographic inferences*
309 Estimates of effective population size (*Ne*) based on linkage disequilibrium did not overlap to infinite in ten
310 cases out of 25 (Table 1). In these ten populations, mean *Ne* values were always <500 and, in three cases,
311 <50.
312 No population exhibited a significant heterozygosity excess, while a significant deficiency in the MR value
313 with respect to the values expected at the equilibrium was found in 17 populations out of 25 (Table S6),
314 providing evidence for the presence of bottlenecks alongside most population demographic histories.
315 In the ABC analysis, *Scenario 2* had the highest posterior probabilities (0.5918) and its confidence interval
316 (0.5535-0.6301) did not overlap those of the other three scenarios (Table S7). Only one out of the 30
317 summary statistics calculated on the simulated data for *Scenario 2* was significantly different from the

observed data (Table S8) which, together with the visual inspection by Principal Component Analysis (Figure S5), suggested a high goodness-of-fit of the observed data to the results of this scenario. Median values of the effective population sizes (N#) for *Scenario 2* were 6,280 (95% CI: 2,210-9710), 16,100 (95% CI:8,910-19,700), 2,440 (95% CI:711-6,650), 40,100 (95% CI: 1,990-96,600), and 6,340 (95% CI: 522-18,500) for NI, CI, SI, Nb and Na, respectively (Figure 2b, Table S9). Posterior parameters showed that CI had a larger effective population size compared to NI and SI, even considering the CI values. Median values of t1, t4, t5 and t6 were 111 (95% CI: 21-369), 671 (95% CI:564-2,070), 2,420 (95% CI: 637-7,320), and 5,220 (95% CI:1,380-9,670) generations ago, respectively (Figure 2b, Table S9). Assuming 30 years as generation time (Adams 2003), 'gene flow' from CI to NI (t1) and from SI to CI (t4) were dated to 3,330 (95% CI: 630-11,070) and 20,130 (95% CI: 5,280-62,100) years ago, respectively, while divergence times of NI and CI (t5) and of SI and Nb (t6) were scaled to 72,600 (95% CI: 19,110-219,600) and 156,600 (95% CI: 41,400-290,100) years ago, respectively. Median values of ra and rd were 0.898 (95% CI: 0.429-0.993) and 0.150 (95% CI: 0.006-0.758), respectively (Figure 2b, Table S9). The median values of the mean mutation rate of SSR and mean SNI were 2.94×10^{-4} (95% CI: 1.27×10^{-4}-6.74×10^{-4}) and 1.67×10^{-7} (95% CI: 1.16×10^{-8}-2.69×10^{-6}), respectively (Table S9). The median value of the mean P was 0.280 (95% CI: 0.159-0.300) (Table S9).

Discussion

Quercus robur is a keystone species in lowland ecosystems and it provides a wide range of ecosystem services and wood products. Nonetheless, there is limited knowledge on the genetic diversity of its rear edge populations, despite they reasonably harbour unique genetic variation as a legacy of southward retreatment during the Quaternary glaciations. Our study represents the first, comprehensive characterisation of the nuclear genetic diversity of *Q. robur* genetic resources across the Italian peninsula. We exhaustively covered the entire distribution of the species, sampling 745 trees from 25 populations from both large and extremely small remnant forests. As Fineschi et al. (2002), we observed the highest genetic diversity in central Italy and the lowest in southern Italy. Beyond that, we found that genetic distinctiveness and the contribution to genetic differentiation showed a south-to-north decrease. We observed a similar trend in the complexity of the genetic structure, with peninsular Italy being characterised by a mosaic of different gene pools, in contrast to the relative genetic homogeneity of populations located in northern Italy.

Multi-refugial areas are expected to host forest tree populations with more complex genetic structures than recolonization areas (de Lafontaine et al. 2013), and growing evidence is showing that such pattern holds true for several species in the Italian peninsula (e.g. Piotti et al. 2017; De Dato et al. 2020; Fortini et al. 2023). Although our results were limited to a single multi-refugial area, they showed interesting hints about pedunculate oak dynamics in the last 300,000 years. At the highest hierarchical level, both Bayesian clustering and multivariate analysis revealed the existence of two main gene pools among Italian populations: a first one, at the very rear edge of the species distribution, consisting of only one population (Bosco Pantano, PAN,

353 along the Ionian coast), and a second one made up of all remaining populations that clearly diverged from
354 PAN likely before the Last Glacial Maximum (LGM), even when considering the 95% credible intervals. Even
355 though we need to be aware of the several sources of uncertainty in ABC inferences (Tsuda et al. 2015), this
356 result points towards PAN representing the last remnant of larger southern populations that likely survived
357 (at least) the last two glacial periods in southern Italy, isolated from the other detected gene pools. In fact,
358 since its divergence the southern gene pool had apparently no longer received any gene flow from the north
359 while asymmetrically acting as a source of northward historical gene flow, more intensively at the end of the
360 LGM. Palynological and genetic evidence support the existence of a glacial refugium for pedunculate oak in
361 the surroundings of Laghi di Monticchio, 200 km north of PAN (Watts 1996), or further south (Brewer et al.
362 2002; Fineschi et al. 2002; Petit et al. 2002a; Magri et al. 2015), an area for which, unfortunately, palynological
363 records are scarce (see, for instance, Figure 1 in Magri et al. 2015). This prolonged history of isolation made
364 PAN the most genetically distinct population in Italy, with the highest contribution to genetic differentiation
365 and the second largest private allelic richness value. Nowadays, PAN is extremely small (i.e. 64 remaining
366 trees, all surveyed and genotyped) and isolated, mostly surrounded by cultivated fields and coastal pine
367 forests (Travascia et al. 2022, Pericolo et al. 2023). The population, which has undergone a strong reduction
368 in its census size over the last century due to massive human overexploitation, showed reduced genetic
369 variation, with an average ~25% decrease in allelic richness with respect to the other studied populations.
370 PAN also showed a significantly lower ratio between the number of alleles and the range of allele size than
371 expected at equilibrium, which is usually interpreted as a sign of a severe and prolonged bottleneck
372 (Williamson-Natesan 2005), likely eroding the otherwise high genetic diversity expected in a refugial area
373 (Petit et al. 2003). Its F value, an analogue to F_{ST} between the gene pools retrieved by STRUCTURE and the
374 assumed common ancestral population, was 0.220 for K = 2, which is an independent proof that the
375 population experienced the effects of intense genetic drift (Falush et al. 2003), as expected after a severe
376 demographic reduction. Nevertheless, no inbreeding was observed and the effective population size, albeit
377 low, matched the census size of the population (Ne = 92, $CI_{95\%}$ = 67-135), probably indicating that the local
378 gene pool was well represented in the reproductive dynamics leading to the inspected adult generation
379 (Sandurska et al. 2019) or that the major size reduction was relatively recent, as suggested by anecdotal
380 historical records (e.g. Douglas 1907). In this regard, it is important to stress that the genetic effects of habitat
381 fragmentation are often overlooked in forest trees as studies have mostly focused on pre-fragmentation
382 cohorts (Lowe et al. 2015). A strong reduction in census size due to extensive cutting in the PAN population
383 happened in the '50s (Travascia et al. 2022) and, therefore, deviations in parameters more sensible to very
384 recent demographic dynamics could only be detected in the regeneration established in the last decades,
385 which is virtually absent in PAN.
386 At the next strongest levels of structuring, the gene pool comprising all remaining *Q. robur* populations was
387 further subdivided. At K = 3 and K = 6, one and four differentiated gene pools emerged in central Italy,

respectively, while northern Italian populations kept belonging to a single, rather homogeneous gene pool. We can rule out the possibility that the populations nowadays present in central and northern Italy originated from southern Italy after the Last Glacial Maximum (LGM), which was a recolonization route previously hypothesized (Fineschi et al. 2002; Petit et al. 2002a). Moreover, central Italy was not simply a 'transit area' where multiple recolonization routes got in contact, but it likely acted as a mosaic of small refugia where pedunculate oak persisted during the LGM, overall preserving both high and private genetic diversity. Interestingly, most populations from central Italy have linkage disequilibrium-based *Ne* estimates <500, while the global *Ne* as estimated by ABC analysis was the highest among the three gene pools. This discrepancy is compatible with a scenario where the CI gene pool originated from a mosaic of small refugia exhibiting a higher *Ne* as compared to those of nowadays, highly fragmented, declining populations. Fortini et al. (2023) reached similar conclusions for the European white oak complex in this geographic area by studying chloroplast genetic diversity in southern Europe and reviewing palaeobotanical evidence. Contrary to first syntheses (Brewer et al., 2002), more recent palaeobotanical surveys indeed reported significant percentages of deciduous *Quercus* pollen in peninsular Italy at the onset of the Holocene, with particularly high values in central Italy (Magri et al. 2015). Although a more detailed reconstruction is hampered by difficulties in distinguishing pollen grains at the species level, palynological data indicate that deciduous *Quercus* spp. had limited migrations throughout the Holocene in Italy, and major demographic events were linked to increasing or decreasing their local density (Magri et al. 2015). Our data, showing no significant clines in genetic diversity along possible recolonization routes involving specific gene pools (e.g. longitudinal clines in northern Italy), supported this hypothesis.

Northern populations showed an overall low genetic distinctiveness (average F_{ST} = 0.015), clustered tightly on the first three PCo axes, and showed very similar coancestry coefficients for all relevant levels of structuring in the Bayesian clustering analysis. These results were in contrast with those from a regional-scale study which reported one order of magnitude higher differentiation values (F_{ST} = 0.107) for *Q. robur* populations in the Po Valley based on isozyme markers (Ducci et al. 2007). However, the low levels of genetic differentiation we found in this extensive area (~50,000 km^2) were in line with several other studies reporting low to moderate levels of genetic differentiation at the regional scale (e.g. G_{ST} = 0.003 in Burger et al. 2021; F_{ST} = 0.032 in Kesić et al. 2021; F_{ST} = 0.051 in Ballian et al. 2010). *Q. robur* is indeed a wind-pollinated tree, with a high potential for long-distance dispersal and extensive gene flow (Gerber et al. 2014). Effective pollen dispersal distances up to 80 km in isolated populations are reported for this species (Buschbom et al. 2011). Altogether, available evidence about *Q. robur* gene flow capabilities and the scarce differentiation found in such a large area outlined a scenario in which *Q. robur* populations in the Po Valley constituted a single, large meta-population (*sensu* Hanski and Gilpin 1991), which has been genetically connected despite increasing fragmentation at least until the reproductive events that generated the inspected cohort. The peculiar topography of the Po Valley, with no significant geographic barriers, coupled with the current presence of

423 typical elements of the agroforestry landscape, such as rows of *Q. robur* trees and isolated individuals, may
424 have favoured pollen exchanges among pedunculate oak shrinking populations, leading to the maintenance
425 of the extremely low genetic differentiation detected. Overall, our results and palaeobotanical records,
426 showing almost complete extinction of deciduous forests during the last glacial-interglacial transition, as well
427 as the lack of ancestral chloroplast haplotypes in northern Italy (Pini et al. 2022; Fortini et al. 2023) confirm
428 the absence of a large refugial area in the Po plain (Petit et al., 2002a). These pieces of evidence raise the
429 question about the post-glacial origin of these stands which currently represent by far the largest part of *Q.*
430 *robur* distribution in Italy. Being highly unlikely the hypothesis of their origin from populations further south,
431 and considering the absence of longitudinal clines in genetic diversity in this area, only enlarging the genetic
432 characterization to populations in southern France and northern Balkans would allow to investigate a
433 possible origin from these areas, which were classified as secondary refugia for European white oaks by Petit
434 et al. (2002a).
435 The role of landscape features in shaping the genetic structure of *Q. robur* populations can be discussed also
436 by shifting the focus on populations from central Italy, where the species occurs in much more
437 heterogeneous topographic conditions. For instance, the distance between the two populations of ORV and
438 FAB was ~15 km and their pairwise genetic differentiation (G_{ST}) was equal to 0.026. As a comparison, the
439 easternmost and westernmost populations of the Po Valley (i.e. MER and FAG) were around 450 km apart
440 but were much less differentiated (G_{ST} = 0.006). The differentiation between ORV and FAB was intercepted
441 also by the Bayesian structuring at K = 6, where FAB emerged as a highly genetically peculiar population.
442 Although the two populations were geographically close, they were separated by a small ridge that might
443 have acted as an effective barrier to pollen and seed gene flow. As fine-scale topography may determine
444 contrasting micro-climates across very small distance (Scherrer and Körner 2011), individuals from the two
445 populations could also experiment unsynchronised phenological cycles due to separate post-glacial
446 dynamics, which would further hamper pollen gene flow. Despite the potential for extensive gene flow in
447 this species, large differentiation among populations was observed at range margins (Moracho et al. 2016;
448 Pohjanmies et al. 2016). For instance, at the southwestern margin of the species distribution, Moracho et al.
449 (2016) studied ten *Q. robur* stands located across a 20-km long valley, identifying four distinct gene pools,
450 and finding that most pollination events occurred within each stand. The Authors suggested that the
451 combination of rugged topography, dense canopy and the humid microclimate of riparian forests could have
452 limited long-distance pollen dispersal.
453 Although we found little genetic differentiation, especially among the northern populations, and a complete
454 absence of inbreeding (see Table 1 and Figure S6), concerns about the genetic status of pedunculate oak
455 populations in Italy still arise from their high to extreme degree of fragmentation. A genetic parameter which
456 is potentially intercepting early signals of genetic decline is the effective population size (*Ne*). As it quantifies
457 the rate at which a population loses genetic variation, *Ne* is commonly used in conservation genetics to

evaluate the state and viability of populations. As a rule of thumb, a *Ne* > 50 would be beneficial for avoiding the harmful effects of inbreeding depression, while a *Ne* > 500 would be necessary for maintaining sufficient adaptive potential for coping with changing environmental conditions. (Franklin 1980). We found that ten populations had a *Ne* <500 (three of them <50). These populations represented the 25%, 62.5% and 100% of the northern, central and southern Italian populations, respectively. Low *Ne* estimates for *Q. robur* populations were often reported (Dering and Chybicki 2012; Vranckx et al. 2014; Sandurska et al. 2019). For instance, four Belgian *Q. robur* stands exhibited *Ne* estimates ranging from 23 to 58 which were associated, as in our case, to negligible inbreeding (Vranckx et al. 2014). However, these Authors found a negative relationship between *Ne* and the level of correlated paternity, suggesting that inbreeding might increase in the future because of more frequent mating among relatives. Although it is intensely debated whether *Ne* estimates could be reliable for forest tree species, particularly for large, interconnected populations (Fady and Bozzano 2021; Hoban et al. 2021; Santos-del-Blanco et al. 2022), we may consider our results on small and isolated populations, especially from central and southern Italy, as an early warning of a negative genetic trajectory that may worsen in future generations. Monitoring natural regeneration together with adult cohorts is crucial to investigate how neutral and adaptive genetic variation is transmitted through generations (Lowe et al. 2015; Vajana et al. 2023), and the low to moderate *Ne* values of our populations might predict a significant genetic depauperation in future generations.

Conservation implications and outlook

Although post-LGM recolonization and northward expansion of tree species from Italian refugia have been widely discussed in the last two decades, this is the first study to evaluate the genetic structure and demographic history of *Quercus robur*, one of the key elements of European floodplain forests, in the Italian peninsula based on a robust sampling of extant populations. Our results showed that at least three *Q. robur* refugia contributed to the genetic layout of extant populations after the last glacial period, but refugial areas were probably even more numerous in central Italy. The complexity of the resulting genetic structure can help to identify priorities for the conservation of *Q. robur* forest genetic resources in Italy. For example, at the national scale, the identification of reserves and seed stands for in situ and ex situ conservation programs is generally not based on any knowledge on the genetic layout of forest trees' populations. For *Q. robur*, there are almost 100 registered seed stands in Italy (MASAF 2023), the vast majority located in northern Italy (Figure 6). This means that the distinctive genetic diversity present in peninsular Italy is currently not sampled in national initiatives of seed collection despite, for instance, counting genotypes potentially pre-adapted to changing climatic conditions at northern latitudes. At the European scale, the situation improves with six Genetic Conservation Units (GCUs) of the pan-European network fostered by EUFORGEN (Koskela et al. 2013) distributed throughout the Italian range of the species (Figure 6). Three of them (corresponding to MON, CRC, and PAN) were established throughout the projects leading to the collection of the dataset presented

here. However, given the mosaic of gene pools unveiled in central Italy, our results can help refine the choice of additional conservation units, possibly based on the results of algorithms of spatial conservation prioritization (Vajana et al. 2024), or 'backup' GCUs to further secure the genetic diversity present in each environmental zone, as recommended by EUFORGEN guidelines (de Vries et al. 2015). In addition to providing practical guidance on the conservation of genetic diversity and serving as a basis for future genetic monitoring, our results also invite for a reflection on what is the best strategy for the long-term maintenance of such small and isolated populations, that is treating them as independent conservation units or intervening to increase their genetic connectivity (e.g. through assisted migration or genetic corridors). The key to informing this choice is to integrate our results on the outcome of past demographic dynamics with an assessment of the adaptive potential and risk of non-adaptedness of *Q. robur* populations by population genomics and eco-physiological approaches in future studies.

References

Adams SR (2003) Regulators of growth | Juvenility. In: Thomas B (Eds.), Encyclopedia of Applied Plant Sciences. Academic Press, pp. 1051-1056.

Agnoletti M (2018) Storia del bosco. Il paesaggio forestale italiano. Editori Laterza, Rome, Italy.

Andrello M, D'Aloia C, Dalongeville A, et al (2022) Evolving spatial conservation prioritization with intraspecific genetic data. Trends Ecol Evol 37:553–564. https://doi.org/10.1016/j.tree.2022.03.003

Aitken SN, Whitlock MC (2013) Assisted gene flow to facilitate local adaptation to climate change. Ann Rev Ecol Evol Syst 44:367–388. https://doi.org/10.1146/annurev-ecolsys-110512-135747

Ballian D, Belletti P, Ferrazzini D, et al (2010) Genetic variability of pedunculate oak (*Quercus robur* L.) in Bosnia and Herzegovina. Period Biol 112:353–362.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lond B 263: 1619–1626. https://doi.org/10.1098/rspb.1996.0237

Belletti P, Leonardi S, Monteleone I, Piovani P (2005) Allozyme variation in different species of deciduous oaks from northwestern italy. Silvae Genet 54:9–16. https://doi.org/10.1515/sg-2005-0002

Belletti P, Calvo E, Ducci F, et al (2007) La foresta della Carpaneta e il programma per le dieci grandi foreste lombarde. Studi e ricerche preliminari. In: Ducci F (Eds.), Le Risorse genetiche della farnia della Val Padana. Progetto "Foresta della Carpaneta", Regione Lombardia – ERSAF, Milano, pp. 10-15.

Brewer S, Cheddadi R, de Beaulieu JL, Reille M (2002) The spread of deciduous *Quercus* throughout Europe since the last glacial period. For Ecol Manag 156:27–48. https://doi.org/10.1016/S0378-1127(01)00646-6

Burger K, Müller M, Rogge M, Gailing O (2021) Genetic differentiation of indigenous (*Quercus robur* L.) and late flushing oak stands (*Q. robur* L. subsp. *slavonica* (Gáyer) Mátyás) in western Germany (North Rhine-Westphalia). Eur J Forest Res 140:1179–1194. https://doi.org/10.1007/s10342-021-01395-8

Buschbom J, Yanbaev Y, Degen B (2011) Efficient long-distance gene flow into an isolated relict oak stand. J Hered 102:464–472. https://doi.org/10.1093/jhered/esr023

Calvo E (2007) Consistenza delle risorse in Italia settentrionale. In: Ducci F (Eds.), Le Risorse genetiche della farnia della Val Padana. Progetto "Foresta della Carpaneta", Regione Lombardia – ERSAF, Milano, pp. 28–41.

Caudullo G, Welk E, San-Miguel-Ayanz J (2017) Chorological maps for the main European woody species. Data Brief 12:662–666. https://doi.org/10.1016/j.dib.2017.05.007

Chapuis MP, Raynal L, Plantamp C, et al (2020) A young age of subspecific divergence in the desert locust inferred by ABC random forest. Mol Ecol 29:4542–4558. https://doi.org/10.1111/mec.15663

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. J Hered 100:106–113. https://doi.org/10.1093/jhered/esn088

Cook CN, Sgrò CM (2017) Aligning science and policy to achieve evolutionarily enlightened conservation. Conserv Biol 31:501–512. https://doi.org/10.1111/cobi.12863

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population

bottlenecks from allele frequency data. Genetics 144:2001–2014. https://doi.org/10.1093/genetics/144.4.2001

Cornuet JM, Santos F, Beaumont MA, et al (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics 24:2713–2719. doi: 10.1093/bioinformatics/btn514

Cornuet J-M, Pudlo P, Veyssier J, et al (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics 30:1187–1189. https://doi.org/10.1093/bioinformatics/btt763

de Dato GD, Teani A, Mattioni C, et al (2020) Genetic analysis by nuSSR markers of silver birch (*Betula* pendula Roth) populations in their southern European distribution range. Front Plant Sci 11: 310. https://doi.org/10.3389/fpls.2020.00310

de Lafontaine G, Ducousso A, Lefèvre S, et al (2013) Stronger spatial genetic structure in recolonized areas than in refugia in the European beech. Mol Ecol 22:4397–4412. https://doi.org/10.1111/mec.12403

de Sampaio e Paiva Camilo-Alves C, da Clara MIE, de Almeida Ribeiro NMC (2013) Decline of Mediterranean oak trees and its association with *Phytophthora cinnamomi*: a review. Eur J Forest Res 132:411–432. https://doi.org/10.1007/s10342-013-0688-z

de Vries SMG, Alan M, Bozzano M, et al (2015) Pan-European strategy for genetic conservation of forest trees and establishment of a core network of dynamic conservation units. European Forest Genetic Resources Programme (EUFORGEN), Bioversity International, Rome, Italy.

Degen B, Yanbaev Y, Ianbaev R, et al (2021) When does habitat fragmentation lead to changes in populations gene pool of pedunculate oak (*Quercus robur* L.)? For Ecol Manag 499:119617. https://doi.org/10.1016/j.foreco.2021.119617

Dering M, Chybicki I (2012) Assessment of genetic diversity in two-species oak seed stands and their progeny populations. Scand J For Res 27:2–9. https://doi.org/10.1080/02827581.2011.631934

Des Roches S, Pendleton LH, Shapiro B, Palkovacs EP (2021) Conserving intraspecific variation for nature's contributions to people. Nat Ecol Evol 5:574–582. https://doi.org/10.1038/s41559-021-01403-5

DeWoody JA, Harder AM, Mathur S, Willoughby JR (2021) The long-standing significance of genetic diversity in conservation. Mol Ecol 30:4147–4154. https://doi.org/10.1111/mec.16051

Do C, Waples RS, Peel D, et al (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol Ecol Resour 14:209–214. https://doi.org/10.1111/1755-0998.12157

Dow BD, Ashley MV, Howe HF (1995) Characterization of highly variable (GA/CT) microsatellites in the bur oak, *Quercus macrocarpa*. Theor Appl Genet 91:137–141. https://doi.org/10.1007/BF00220870

Ducci F, Belletti P, Monteleone I, et al (2007) Gestione di popolazioni di farnia e rovere in Val Padana. In: Ducci F (Eds.), Le Risorse genetiche della farnia della Val Padana. Progetto "Foresta della Carpaneta", Regione Lombardia – ERSAF, Milano, pp. 42–67.

Dumolin-Lapègue S, Demesure B, Fineschi S, et al (1997) Phylogeographic structure of white oaks throughout the European continent. Genetics 146:1475–1487. https://doi.org/10.1093/genetics/146.4.1475

Durand J, Bodénès C, Chancerel E, et al (2010) A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. BMC Genomics 11:570. https://doi.org/10.1186/1471-2164-11-570

Eaton E, Caudullo G, Oliveira S, de Rigo D (2016) *Quercus robur* and *Quercus petraea* in Europe: distribution, habitat, usage and threats. In: San-Miguel-Ayanz J, de Rigo D, Caudullo G, et al (Eds.), European Atlas of Forest Tree Species. Publ. Off. EU, Luxembourg, pp. E01c6df+

Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol Ecol 11:1591–1604. https://doi.org/10.1046/j.1365-294X.2002.01576.x

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

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. Heredity 103:285–298. https://doi.org/10.1038/hdy.2009.74

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Res 10:564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fady B, Cottrell J, Ackzell L, et al (2016) Forests and global change: what can genetics contribute to the major forest management and policy challenges of the twenty-first century? Reg Environ Change 16:927–939. https://doi.org/10.1007/s10113-015-0843-9

Fady B, Bozzano M (2021) Effective population size does not make a practical indicator of genetic diversity in forest trees. Biol Conserv 253:108904. https://doi.org/10.1016/j.biocon.2020.108904

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567–1587. https://doi.org/10.1093/genetics/164.4.1567

Fineschi S, Taurchini D, Grossoni P, et al (2002) Chloroplast DNA variation of white oaks in Italy. For Ecol Manag 156:103–114. https://doi.org/10.1016/S0378-1127(01)00637-5

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. Genetics 180:977–993. https://doi.org/10.1534/genetics.108.092221

Fortini P, Di Pietro R, Proietti E, et al (2023) Dissecting the continuum and unravelling the phylogeographic knot of plastid DNA in European white oaks (*Quercus* sect. *Quercus*): ancient signatures and multiple diversity reservoirs. Eur J For Res. https://doi.org/10.1007/s10342-023-01610-8

Franklin I (1980) Evolutionary change in small populations. In: Soulé M, Wilcox BA (Eds.), Conservation biology: and evolutionary-ecological perspective, Sinauer Associates. Inc., Sunderland, US, pp. 135-150.

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. Mol Ecol 10:305–318. https://doi.org/10.1046/j.1365-294x.2001.01190.x

Gasparini P, Di Cosimo L, Floris A, De Laurentis (2022) Italian National Forest Inventory—Methods and Results of the Third Surve. Springer, Berlin/Heidelberg, Germany.

Gathercole LAP, Nocchi G, Brown N, et al (2021) Evidence for the widespread occurrence of bacteria implicated in Acute Oak Decline from incidental genetic sampling. Forests 12:1683.

https://doi.org/10.3390/f12121683

Gerber S, Chadœuf J, Gugerli F, et al (2014) High rates of gene flow by pollen and seed in oak populations across Europe. PLoS ONE 9:e85130. https://doi.org/10.1371/journal.pone.0085130

Gibbs J (1997) Biotic and abiotic factors affecting the dying back of pedunculate oak *Quercus robur* L. Forestry 70:399–406. https://doi.org/10.1093/forestry/70.4.399

Guichoux E, Lagache L, Wagner S, et al (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. https://doi.org/10.1111/j.1755-0998.2011.02983.x

Hampe A, Petit RJ (2005) Conserving biodiversity under climate change: the rear edge matters: rear edges and climate change. Ecol Lett 8:461–467. https://doi.org/10.1111/j.1461-0248.2005.00739.x

Hanski I, Gilpin M (1991) Metapopulation dyamics: brief history and conceptual domain. Biol J Linn Soc 42:3–16.

Hoban S, Bruford M, D'Urban Jackson J, et al (2020) Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. Biol Conserv 248:108654. https://doi.org/10.1016/j.biocon.2020.108654

Hoban S, Paz-Vinas I, Aitken S, et al (2021) Effective population size remains a suitable, pragmatic indicator of genetic diversity for all species, including forest trees. Biol Conserv 253:108906. https://doi.org/10.1016/j.biocon.2020.108906

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Hughes AR, Inouye BD, Johnson MTJ, et al (2008) Ecological consequences of genetic diversity: ecological effects of genetic diversity. Ecol Lett 11:609–623. https://doi.org/10.1111/j.1461-0248.2008.01179.x

Kalinowski ST (2005) hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol Ecol Notes 5:187–189. https://doi.org/10.1111/j.1471-8286.2004.00845.x

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

Kesić L, Cseke K, Orlović S, et al (2021) Genetic diversity and differentiation of pedunculate oak (*Quercus robur* L.) populations at the southern margin of its distribution range—Implications for Conservation. Diversity 13:371. https://doi.org/10.3390/d13080371

Kitamura K, Namikawa K, Tsuda Y, et al (2022) Possible northern persistence of Siebold's beech, *Fagus crenata*, at its northernmost distribution limit on an island in Japan Sea: Okushiri Island, Hokkaido. Front. Plant Sci 13:990927. doi: 10.3389/fpls.2022.990927

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

Koskela J, Lefevre F, Schueler S, et al (2013) Translating conservation genetics into management: Pan-European minimum requirements for dynamic conservation units of forest tree genetic diversity. Biol Conserv 157:39–49. http://dx.doi.org/10.1016/j.biocon.2012.07.023

Laikre L, Hoban S, Bruford MW, et al (2020) Post-2020 goals overlook genetic diversity. Science 367:1083–1085. https://doi.org/10.1126/science.abb2748

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

López-Cortegano E, Pérez-Figueroa A, Caballero A (2019) metapop 2: Re-implementation of software for the analysis and management of subdivided populations using gene and allelic diversity. Mol Ecol Resour 19:1095–1100. https://doi.org/10.1111/1755-0998.13015

Lowe AJ, Cavers S, Boshier D, et al (2015) The resilience of forest fragmentation genetics—no longer a paradox—we were just looking in the wrong place. Heredity 115:97–99. https://doi.org/10.1038/hdy.2015.40

Magri D, Agrillo E, Di Rita F, et al (2015) Holocene dynamics of tree taxa populations in Italy. Rev Palaeobot Palynol 218:267–284. https://doi.org/10.1016/j.revpalbo.2014.08.012

MASAF (2023) Cartografia dei materiali di base per la produzione di materiali forestali di moltiplicazione.

Meirmans PG, Hedrick PW (2011) Assessing population structure: F_{ST} and related measures. Mol Ecol Res 11:5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x

Milesi P, Kastally C, Dauphin B, et al (2023) Synchronous effective population size changes and genetic stability of forest trees through glacial cycles. BioRXiv. https://doi.org/10.1101/2023.01.05.522822

Moracho E, Moreno G, Jordano P, Hampe A (2016) Unusually limited pollen dispersal and connectivity of pedunculate oak (*Quercus robur*) refugial populations at the species' southern range margin. Mol Ecol 25:3319–3331. https://doi.org/10.1111/mec.13692

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. https://doi.org/10.1093/bioinformatics/bts460

Pericolo O, Camarero JJ, Colangelo M, et al (2023) Species specific vulnerability to increased drought in temperate and Mediterranean floodplain forests. Agric For Meteorol 328:109238. https://doi.org/10.1016/j.agrformet.2022.109238

Petit RJ, Aguinagalde I, de Beaulieu J-L, et al (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300:1563–1565. https://doi.org/10.1126/science.1083264

Petit RJ, Brewer S, Bordács S, et al (2002a) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. For Ecol Manag 156:49–74. https://doi.org/10.1016/S0378-1127(01)00634-X

Petit RJ, Csaikl UM, Bordacs S, et al (2002b) Chloroplast DNA variation in European white oaks. Phylogeography and patterns of diversity based on data from over 2600 populations. For Ecol Manag 156:5–26. https://doi.org/10.1016/S0378-1127(01)00645-4

Petit RJ, Latouche-Hallé C, Pemonge M-H, Kremer A (2002c) Chloroplast DNA variation of oaks in France and the influence of forest fragmentation on genetic diversity. For Ecol Manag 156:115–129. https://doi.org/10.1016/S0378-1127(01)00638-7

Pierson JC, Coates DJ, Oostermeijer JGB, et al (2016) Genetic factors in threatened species recovery plans on three continents. Front Ecol Environ 14:433–440. https://doi.org/10.1002/fee.1323

Pini R, Furlanetto G, Vallé F, et al (2022) Linking North Atlantic and Alpine Last Glacial Maximum climates via a high-resolution pollen-based subarctic forest steppe record. Quatern Sci Rev 294:107759. https://doi.org/10.1016/j.quascirev.2022.107759

Piotti A, Leonarduzzi C, Postolache D, et al (2017) Unexpected scenarios from Mediterranean refugial areas: disentangling complex demographic dynamics along the Apennine distribution of silver fir. J Biogeogr 44:1547–1558. https://doi.org/10.1111/jbi.13011

Pohjanmies T, Elshibli S, Pulkkinen P, et al (2016) Fragmentation-related patterns of genetic differentiation in pedunculate oak (*Quercus robur*) at two hierarchical scales. Silva Fennica 50:1510. http://dx.doi.org/10.14214/sf.1510

Postolache D, Oddou-Muratorio S, Vajana E, et al (2021) Genetic signatures of divergent selection in European beech (*Fagus sylvatica* L.) are associated with the variation in temperature and precipitation across its distribution range. Mol Ecol 30:5029–5047. https://doi.org/10.1111/mec.16115

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1093/genetics/155.2.945

R Core Team (2023)._R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ragazzi A, Dellavalle I, Moricca S, et al (2000) Decline of oak species in Italy, problems and perspectives. Accademia Italiana di Scienze Forestali, Florence, Italy.

Sandurska E, Ulaszewski B, Burczyk J (2019) Genetic diversity and differentiation of coexisting populations of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. Acta Biol Crac Ser Bot 61:17–28. https://doi.org/10.24425/ABCSB.2019.127739

Santos-del-Blanco L, Olsson S, Budde KB, et al (2022) On the feasibility of estimating contemporary effective population size (Ne) for genetic conservation and monitoring of forest trees. Biol Conserv 273:109704. https://doi.org/10.1016/j.biocon.2022.109704

Scherrer D, Körner C (2011) Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. J Biogeog 38:406-416. https://doi.org/10.1111/j.1365-2699.2010.02407.x

Scotti-Saintagne C, Boivin T, Suez M, et al (2021) Signature of mid-Pleistocene lineages in the European silver fir (*Abies alba* Mill.) at its geographic distribution margin. Ecol Evol 11:10984–10999. https://doi.org/10.1002/ece3.7886

Sgrò CM, Lowe AJ, Hoffmann AA (2011) Building evolutionary resilience for conserving biodiversity under climate change: Conserving biodiversity under climate change. Evol Appl 4:326–337. https://doi.org/10.1111/j.1752-4571.2010.00157.x

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity 82:561–73. doi: 10.1038/sj.hdy.6885180Steinkellner H, Lexer C, Turetschek E, Glössl J (1997) Conservation of $(GA)_n$ microsatellite loci between *Quercus* species. Mol Ecol 6:1189–1194. https://doi.org/10.1046/j.1365-294X.1997.00288.x

Stojanović DB, Levanič T, Matović B, Orlović S (2015) Growth decrease and mortality of oak floodplain forests as a response to change of water regime and climate. Eur J Forest Res 134:555–567. https://doi.org/10.1007/s10342-015-0871-5

Thomas FM, Blank R, Hartmann G (2002) Abiotic and biotic factors and their interactions as causes of oak decline in Central Europe. Forest Pathol 32:277–307. https://doi.org/10.1046/j.1439-0329.2002.00291.x

Travascia D, Castellaneta M, Colucci A, et al (2022) Valutazione dei fattori di vulnerabilità della riserva naturale Bosco Pantano di Policoro: l'approccio dinamico SWOT - AHP. Forest@ 19:67-73. https://doi.org/10.3832/efor4133-019

Tsuda Y, Nakao K, Ide Y, Tsumura Y (2015) The population demography of *Betula maximowicziana*, a cool-temperate tree species in Japan, in relation to the last glacial period: its admixture-like genetic structure is the result of simple population splitting not admixing. Mol Ecol 24:1403–1418. https://dor.org/10.1111/mec.13123

Tzedakis PC, Andrieu V, de Beaulieu J-L, et al (1997) Comparison of terrestrial and marine records of changing climate of the last 500,000 years. Earth Planet Sci Lett 150:171–176. https://doi.org/10.1016/S0012-821X(97)00078-2

Vajana E, Bozzano M, Marchi M, Piotti A (2023) On the inclusion of adaptive potential in species distribution models: towards a genomic-informed approach to forest management and conservation. Environments 10:3. https://doi.org/10.3390/environments10010003

Vajana E, Andrello M, Avanzi C, et al (2024) Spatial conservation planning of forest genetic resources in a Mediterranean multi-refugial area. Biol Conserv 293:110599. https://doi.org/10.1016/j.biocon.2024.110599

Vakkari P, Blom A, Rusanen M, et al (2006) Genetic variability of fragmented stands of pedunculate oak (*Quercus robur*) in Finland. Genetica 127:231–241. https://doi.org/10.1007/s10709-005-4014-7

Vendramin GG, Fady B, González-Martínez SC, et al (2008) Genetically depauperate but widespread: the case of an emblematic mediterranean pine. Evolution 62:680–688. https://doi.org/10.1111/j.1558-5646.2007.00294.x

Vranckx G, Jacquemyn H, Mergeay J, et al (2014) Transmission of genetic variation from the adult generation to naturally established seedling cohorts in small forest stands of pedunculate oak (*Quercus robur* L.). For Ecol Manag 312:19–27. https://doi.org/10.1016/j.foreco.2013.10.027

Vries SMG de, Alan M, Bozzano M, et al (2015) Pan-European strategy for genetic conservation of forest trees and establishment of a core network of dynamic conservation units, European Forest Genetic Resources Programme (EUFORGEN). Bioversity International, Rome, Italy.

Watts W (1996) Vegetation history and climate of the last 15,000 years at Laghi Di Monticchio, southern Italy. Quat Sci Rev 15:113–132. https://doi.org/10.1016/0277-3791(95)00038-0

Williamson-Natesan EG (2005) Comparison of methods for detecting bottlenecks from microsatellite loci. Conserv Genet 6:551–562. https://doi.org/10.1007/s10592-005-9009-5**Data Accessibility Statement**

Upon acceptance of the manuscript, the dataset is privately and anonimously available for Reviewers at https://figshare.com/s/c40e3f1e9dc0d9211c0e

TABLES

Table 1. Genetic characteristics of sampled populations. *Na*: mean number of alleles, *Ar*: allelic richness, *PAr*: private allelic richness, H_O: observed heterozygosity, H_E: expected heterozygosity, F_{IS}: inbreeding coefficient, F_{ST}: plot-specific differentiation index according to Foll and Gaggiotti (2008), *Ne* and *Ne* $CI_{95\%}$: effective population size and its parametric 95% confidence interval.

Population	***Na***	***Ar***	***PAr***	H_O	H_E	F_{IS}	F_{ST}	***Ne***	***Ne*** $CI_{95\%}$
MER	7.375	6.852	0.108	0.719	0.679	0.007	0.020	115	66-339
LDS	7.875	7.285	0.051	0.690	0.698	0.033	0.012	655	143-∞
TUR	7.688	7.350	0.008	0.676	0.691	0.026	0.013	∞	244-∞
FGN	7.875	7.241	0.059	0.679	0.690	0.014	0.010	∞	196-∞
GER	7.938	7.295	0.046	0.667	0.675	0.009	0.012	97	61-214
SIR	7.750	7.441	0.064	0.725	0.726	0.015	0.016	1330	139-∞
GRO	7.625	7.043	0.003	0.698	0.684	0.010	0.013	∞	382-∞
PAL	8.188	7.510	0.003	0.735	0.697	0.006	0.012	522	134-∞
CDC	7.375	6.841	0.122	0.671	0.684	0.021	0.023	106	61-315
FON	7.563	7.027	0.017	0.640	0.673	0.027	0.012	272	101-∞
MOR	7.625	7.062	0.086	0.721	0.683	0.004	0.025	46	35-65
CAV	7.938	7.319	0.016	0.688	0.696	0.016	0.008	∞	240-∞
FAG	7.375	6.801	0.026	0.654	0.654	0.011	0.025	1061	144-∞
MES	8.000	7.349	0.094	0.691	0.706	0.021	0.014	738	141-∞
PFL	7.750	7.190	0.047	0.685	0.694	0.019	0.014	285	107-∞
BDC	7.813	7.629	0.030	0.720	0.696	0.013	0.008	∞	256-∞
MON	7.688	7.193	0.073	0.682	0.684	0.016	0.025	325	104-∞
TAV	7.625	7.440	0.228	0.734	0.701	0.006	0.023	1348	144-∞
FBN	7.813	7.746	0.151	0.725	0.702	0.006	0.028	51	37-76
FAB	6.250	6.114	0.100	0.679	0.624	0.006	0.080	22	17-29
ORV	8.188	7.986	0.026	0.732	0.725	0.008	0.021	∞	184-∞
FOG	7.313	6.974	0.150	0.734	0.692	0.009	0.053	36	27-54
CRC	7.625	7.428	0.066	0.732	0.706	0.007	0.038	141	73-965
POL	7.000	6.522	0.115	0.679	0.666	0.011	0.055	60	41-102
PAN	6.375	5.352	0.151	0.655	0.651	0.013	0.165	92	67-135

FIGURES

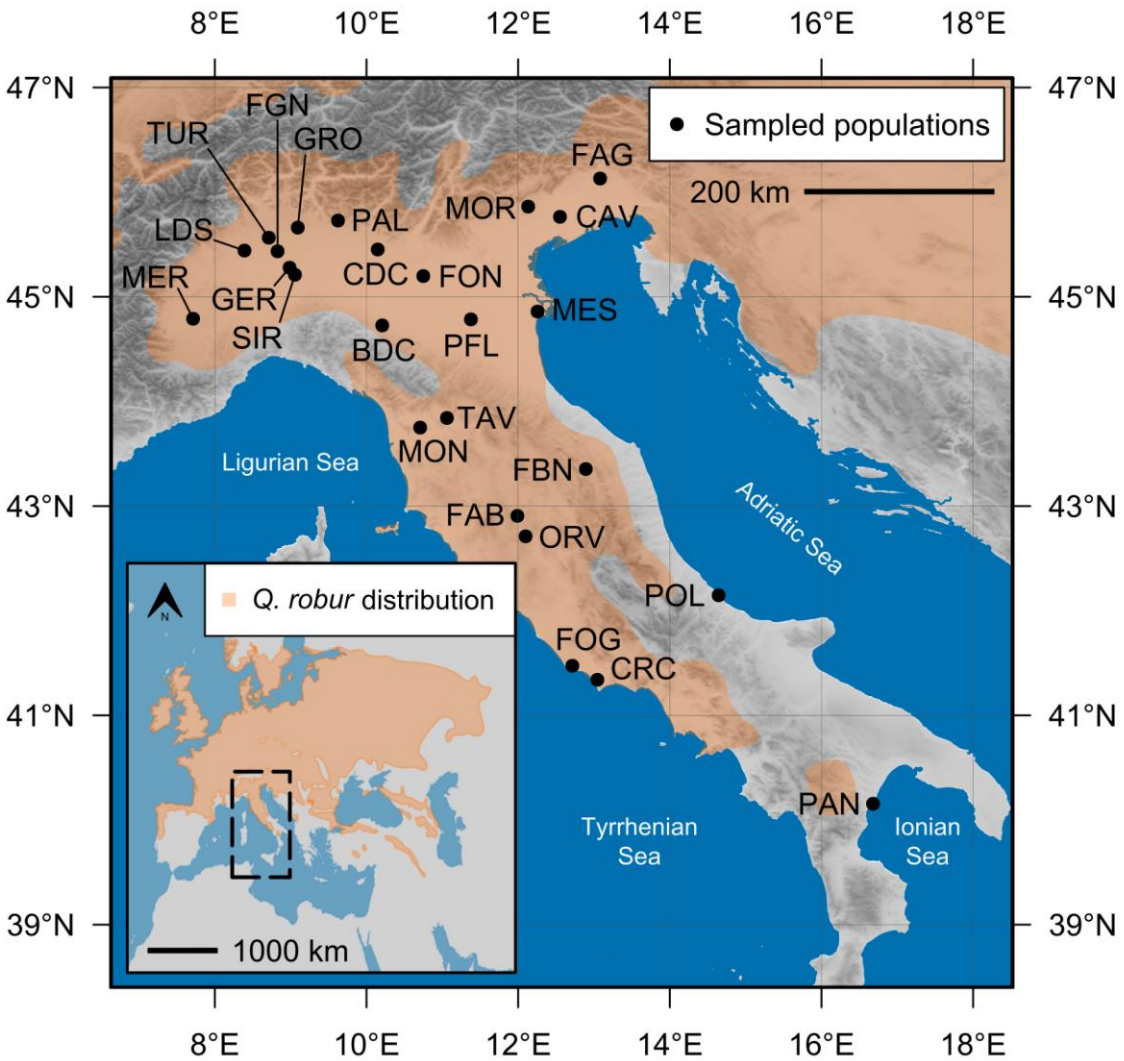


Figure 1. Map of *Quercus robur* populations sampled for the present study. The inset represents the distribution of the species in Europe according to Caudullo et al. (2017). The distribution map of the species strongly overestimates, at least for Italy, the extent and continuity of *Q. robur*-dominated formations, whose distribution is in fact largely fragmented (Belletti et al. 2007).

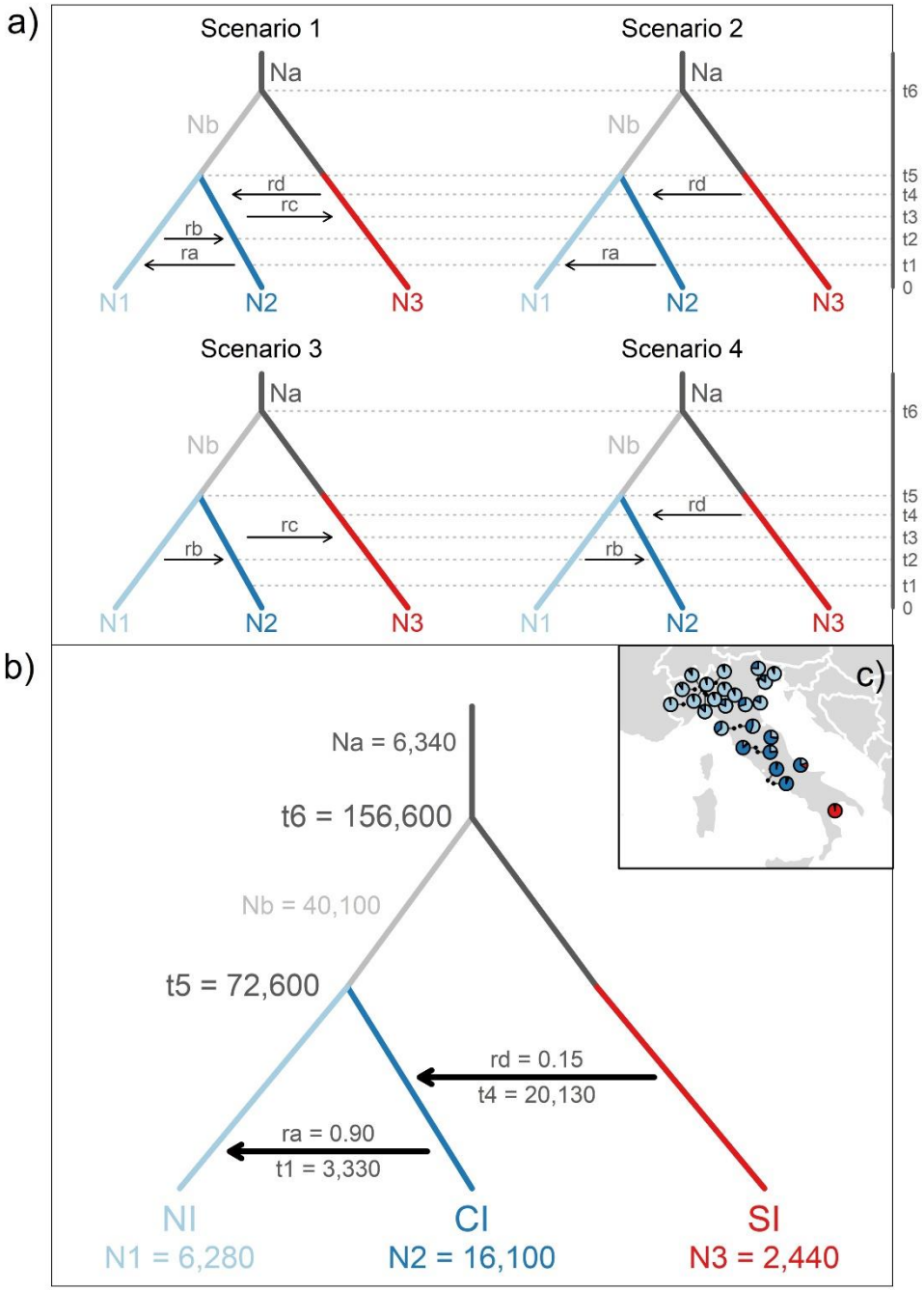


Figure 2. a) Four demographic models tested by using the Approximate Bayesian Computation (ABC) approach implemented in DIYABC. N1, N2 and N3 represent the effective population size of NI, CI and SI, that are the three gene pools resulting from the Bayesian clustering analysis and represent, respectively, individuals from northern Italian populations, central Italian populations and the southernmost stand of PAN. Na and Nb represent the effective population size of the ancestral populations 'a' and 'b'. t# = time-scaled by generation time; r = admixture ratio. b) Median values of posterior parameters for *Scenario 2*, which turned out to be the most likely demographic model based on posterior probabilities. c) Visualization of the geographic positions of the three gene pools, based on the results of the Bayesian clustering approach at *K* = 3 (see Figure 3 for more details).

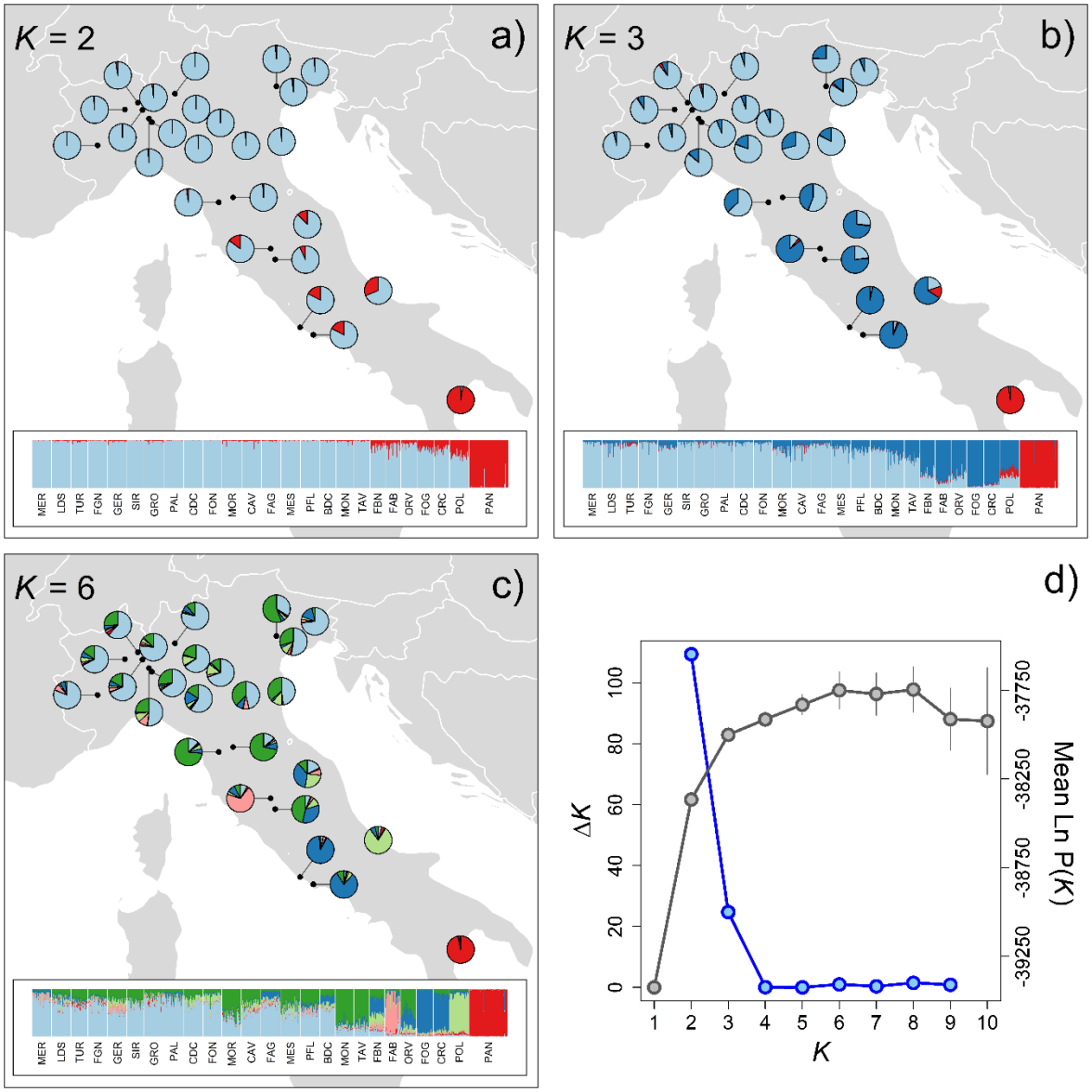


Figure 3. Results of the Bayesian clustering analysis carried out with STRUCTURE. For each population, average ancestry coefficients at a) $K = 2$, b) $K = 3$, and c) $K = 6$ are shown. For each structuring level, individual ancestry coefficients are also shown in the barplot at the bottom of each panel. Evanno et al. (2005)'s ΔK and log-likelihood values used to evaluate the most likely grouping scenarios are presented in panel d).

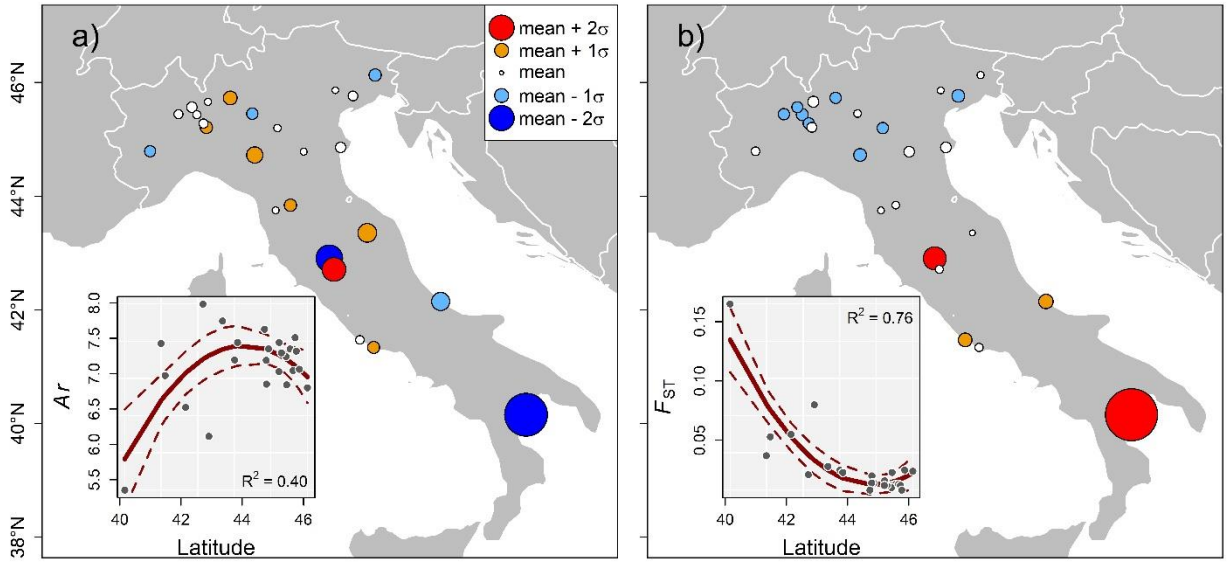


Figure 4. Estimates of (a) allelic richness (*Ar*), and (b) genetic distinctiveness (plot-specific F_{ST}) in the 25 *Quercus robur* populations of the present study. Circle sizes are proportional to the deviation from the mean. Colours indicate the direction of such deviations: white circles represent populations showing around-average values, orange and red circles populations showing above-average values, light blue and blue circles populations showing below-average values. In each panel, the relationship between the genetic parameter and latitude is shown in the inset. Solid line represents the effect of latitude, dashed lines are the limits of 95% confidence interval. The R^2 value of the model is also reported.

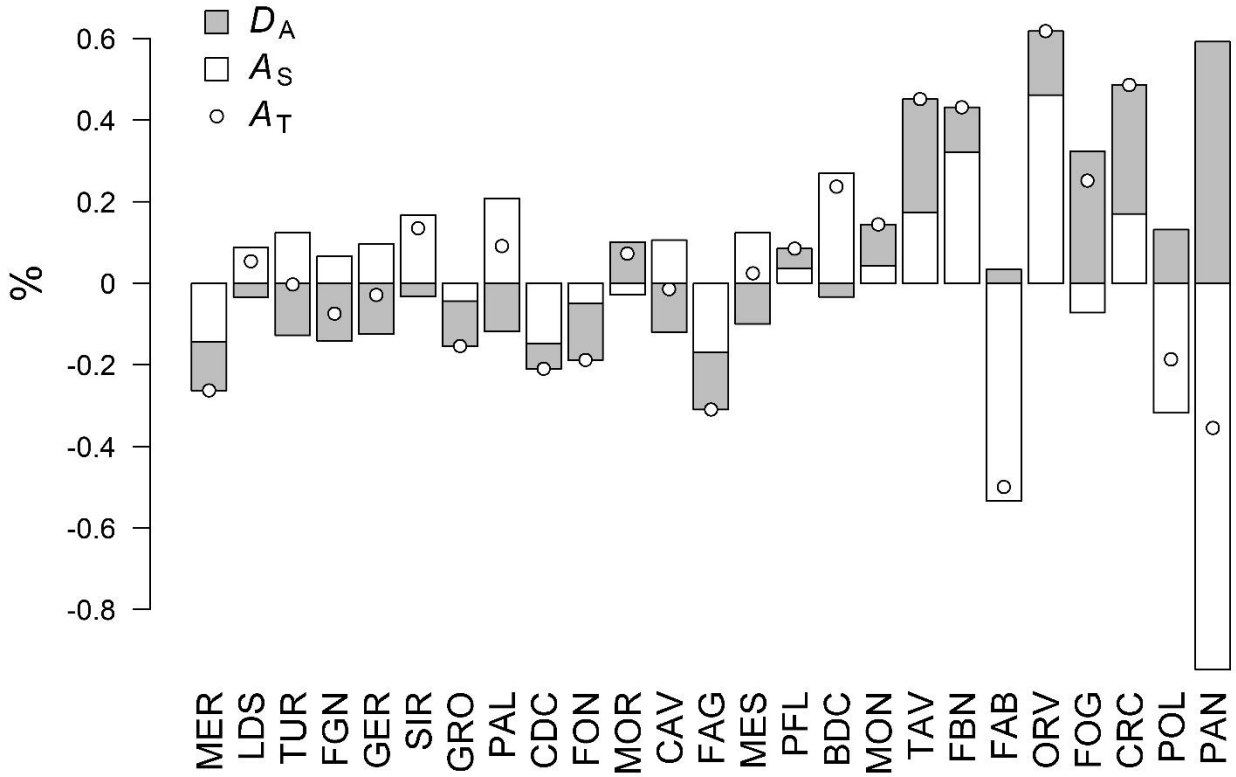


Figure 5. Contributions to within- (A_S; white), between- (D_A; grey) and total (A_T; white dots) allelic diversity of *Quercus robur* populations.

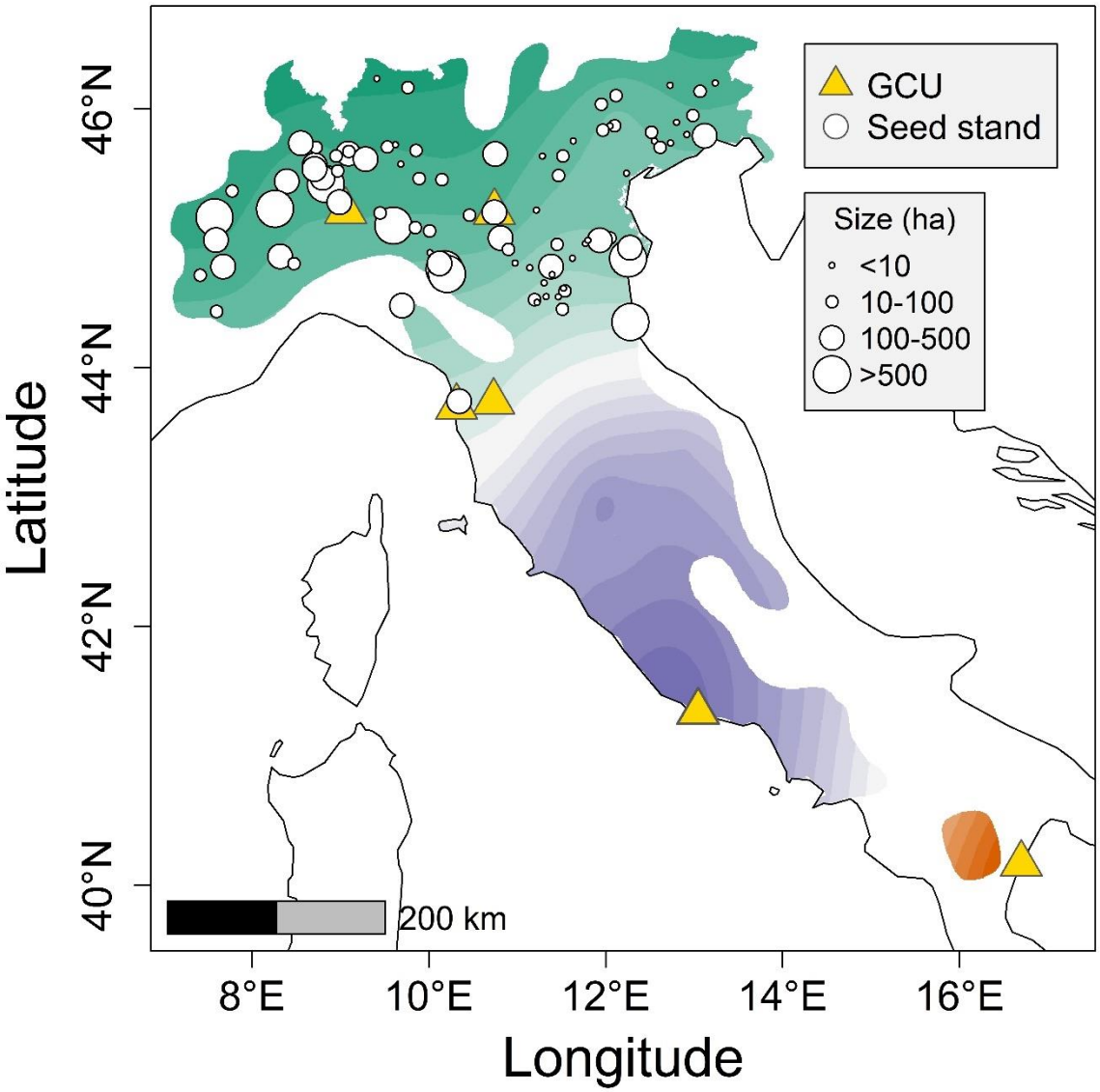


Figure 6. Distribution of Genetic Conservation Units (GCU) (yellow triangles) and registered seed stands (white circles) throughout the Italian peninsula. Information on the distribution of GCUs and seed stands were retrieved from EUFGIS portal (htpp://eufgis.org/) and MASAF (2023), respectively. White circles are proportional to the area of the seed stand (seed stands with an area lower than 1 ha were excluded from the representation). In the background, a spatial interpolation of admixture coefficients estimated with STRUCTURE at *K* = 3 is shown. The colour intensity indicates the probability of each pixel to belong to one of the three gene pools, based on the spatial kriging of the population q-matrix. The interpolation is limited to the distribution range of the species.

APPENDIX 1

To understand which of the three populations should be set as the ancestor, we performed some pilot runs to compare three split model scenarios with northern Italy (NI), central Italy (CI) and southern Italy (SI) gene pools set as ancestral populations, respectively (Figure A). A change in population size on both branches of the tree was modelled at t1 Prior distribution of the estimated parameters are shown in Table A. The best supported scenario was Scenario A (Table B), that is the one with SI as the ancestor population, that was consequently used as ancestral population in all subsequent analyses.

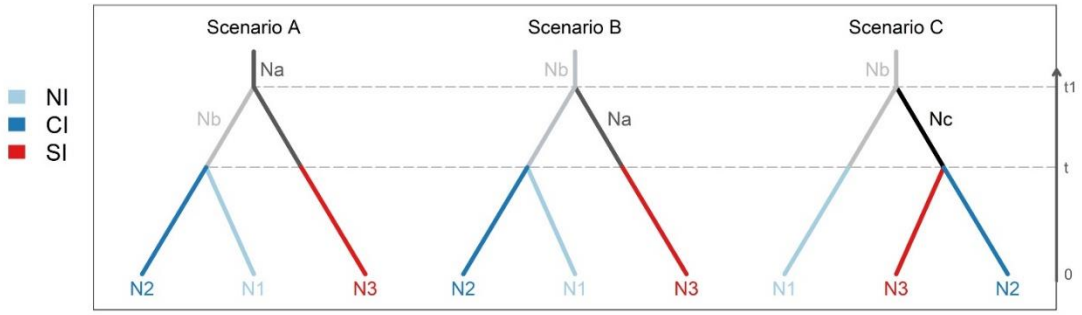


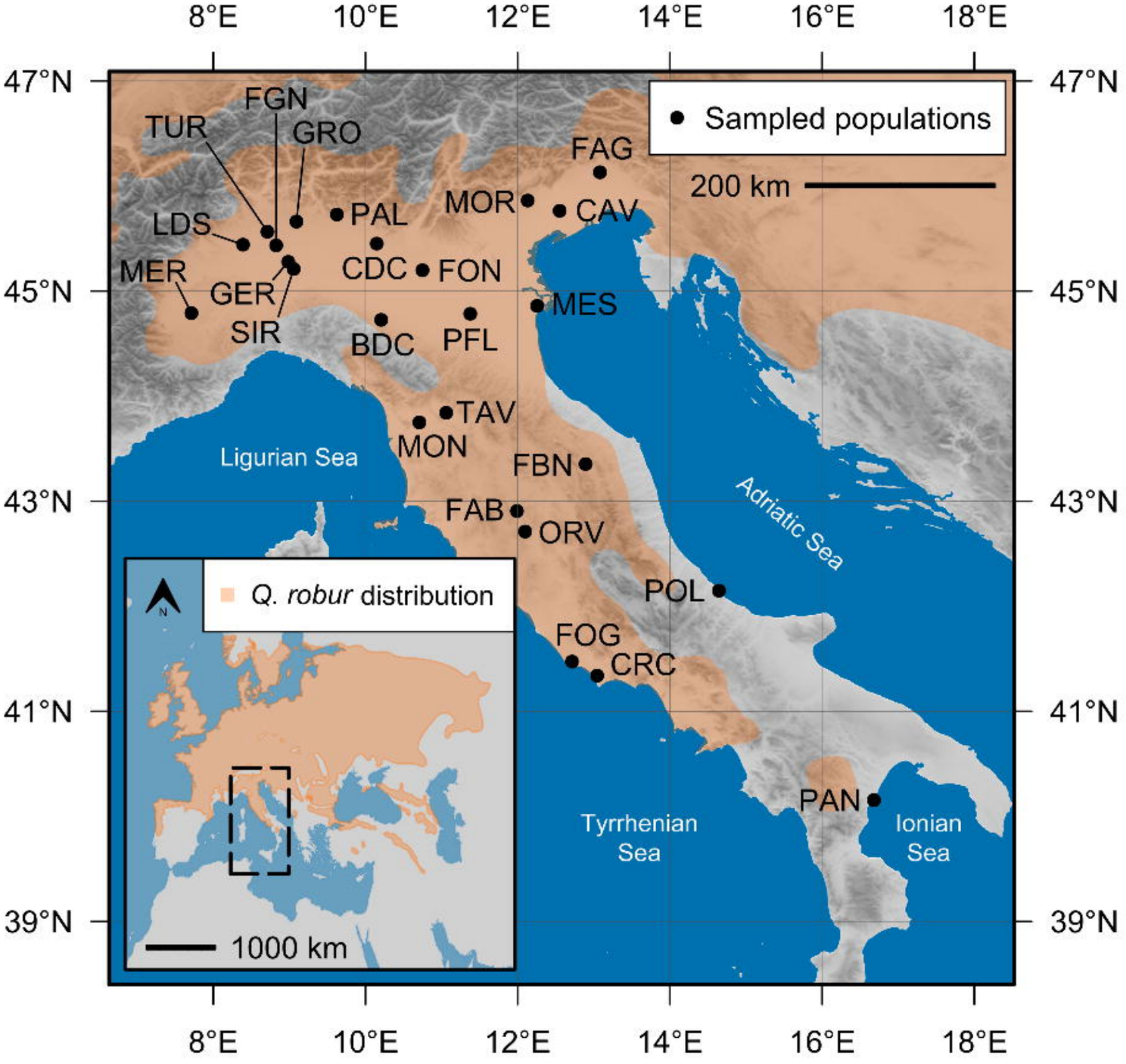
Figure A. Three simple split scenarios tested to understand which populations should be set as the ancestor.

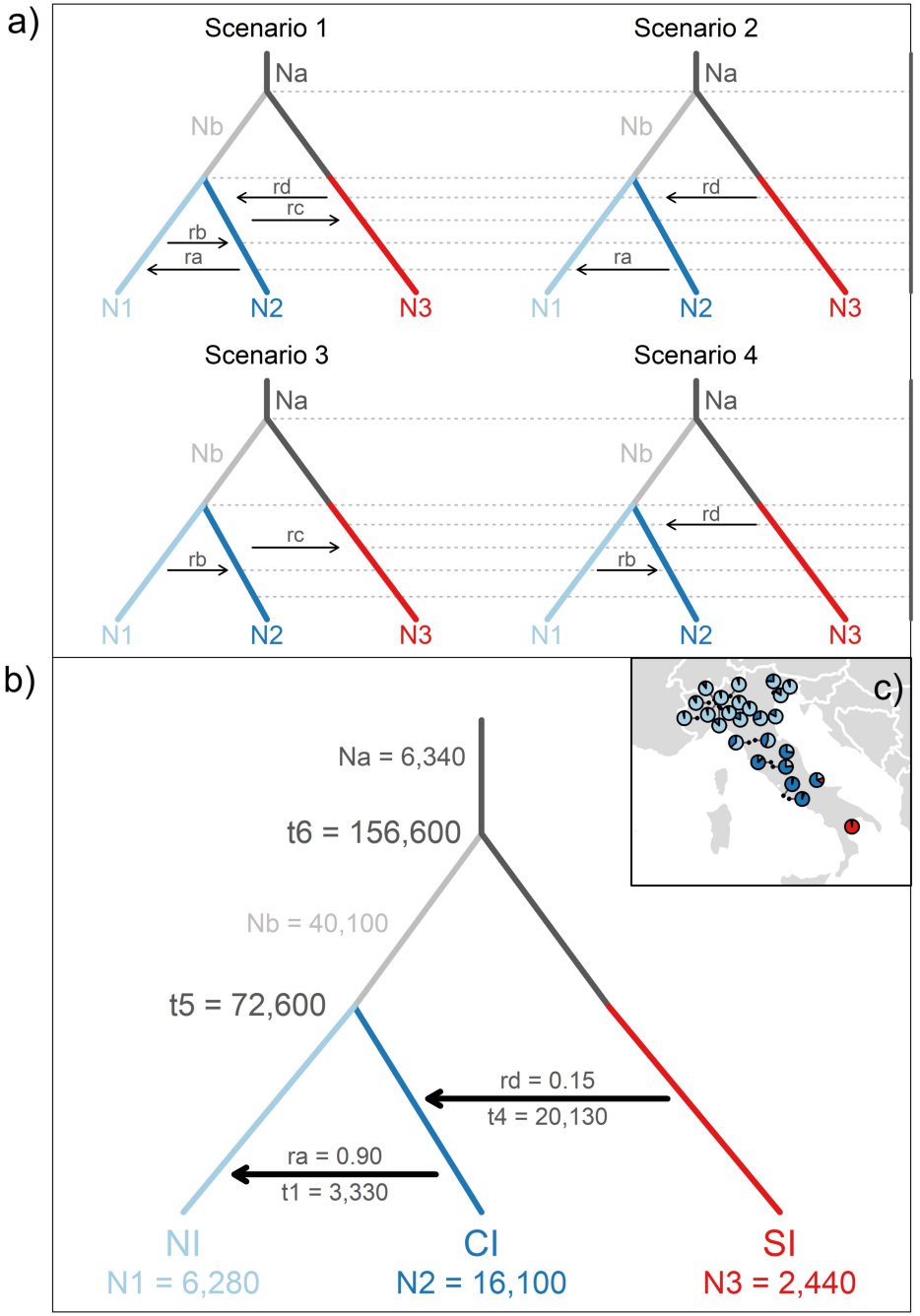
Table A. Prior distributions used in the DIYABC simulations

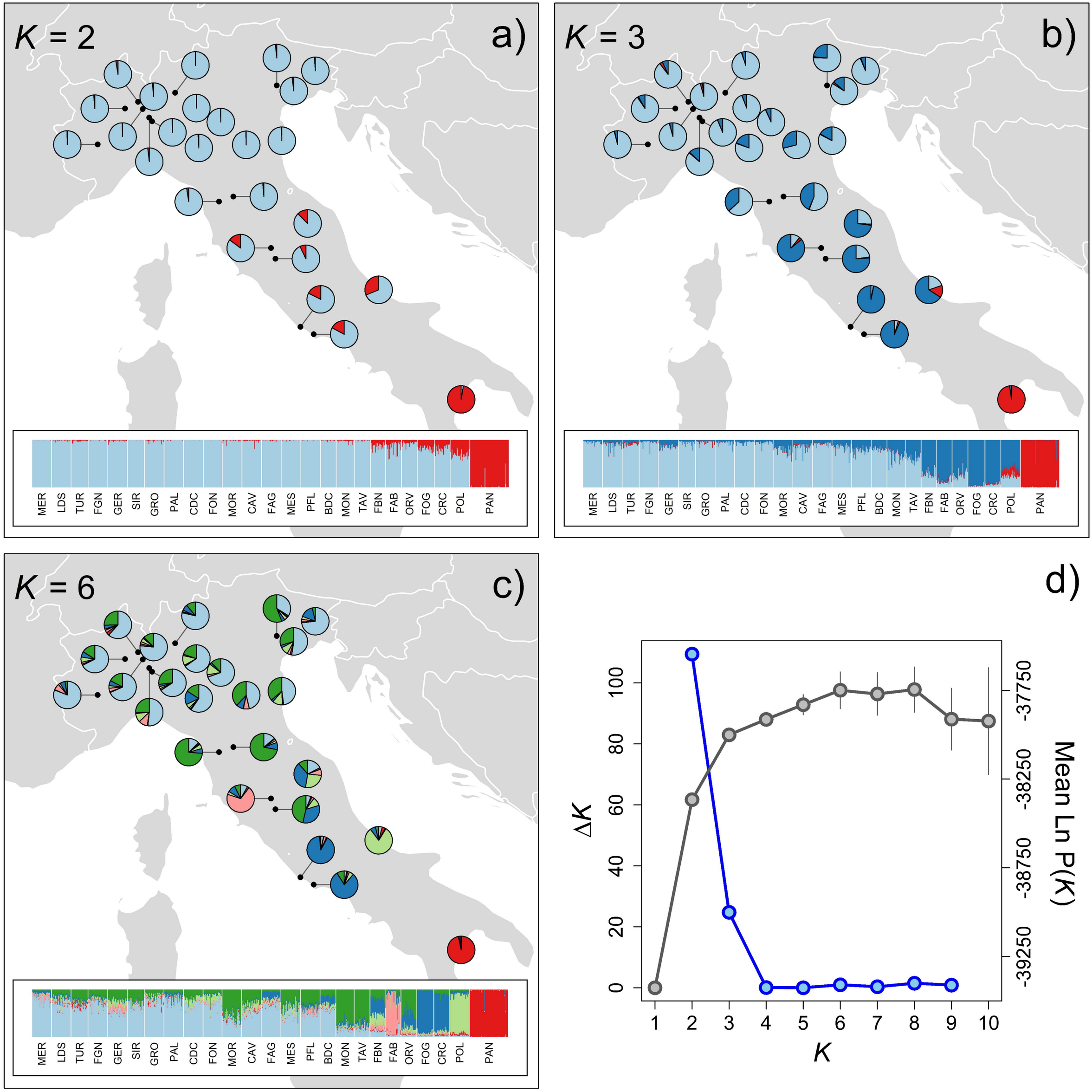
Parameter	Minimum	Maximum
N1	10	10000
N2	10	15000
N3	10	8000
Na	10	20000
Nb	10	100000
Nc	10	100000
t1	10	10000
t	10	1000
Mean mutation rate	1×10^{-4}	1×10^{-3}
Individual locus mutation rate	1×10^{-5}	1×10^{-2}
Mean coefficient P #	1×10^{-1}	3×10^{-1}
Individual locus coefficient P	1×10^{-2}	9×10^{-1}
Mean SNI rate	1×10^{-8}	1×10^{-5}
Individual locus SNI rate	1×10^{-9}	1×10^{-4}

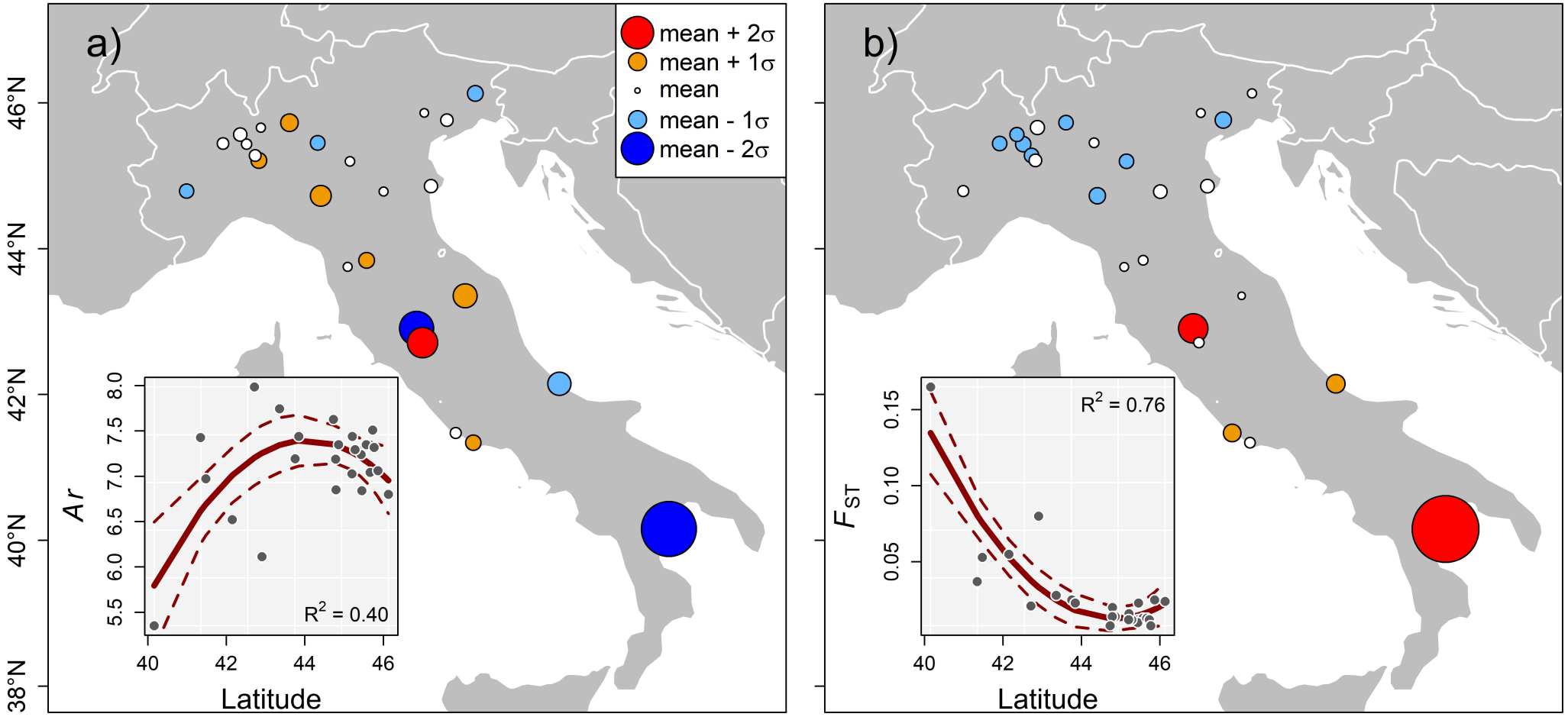
Table B. Posterior probabilities of the three simple split scenarios tested, and their 95% confidence intervals based on the DIYABC logistic estimate.

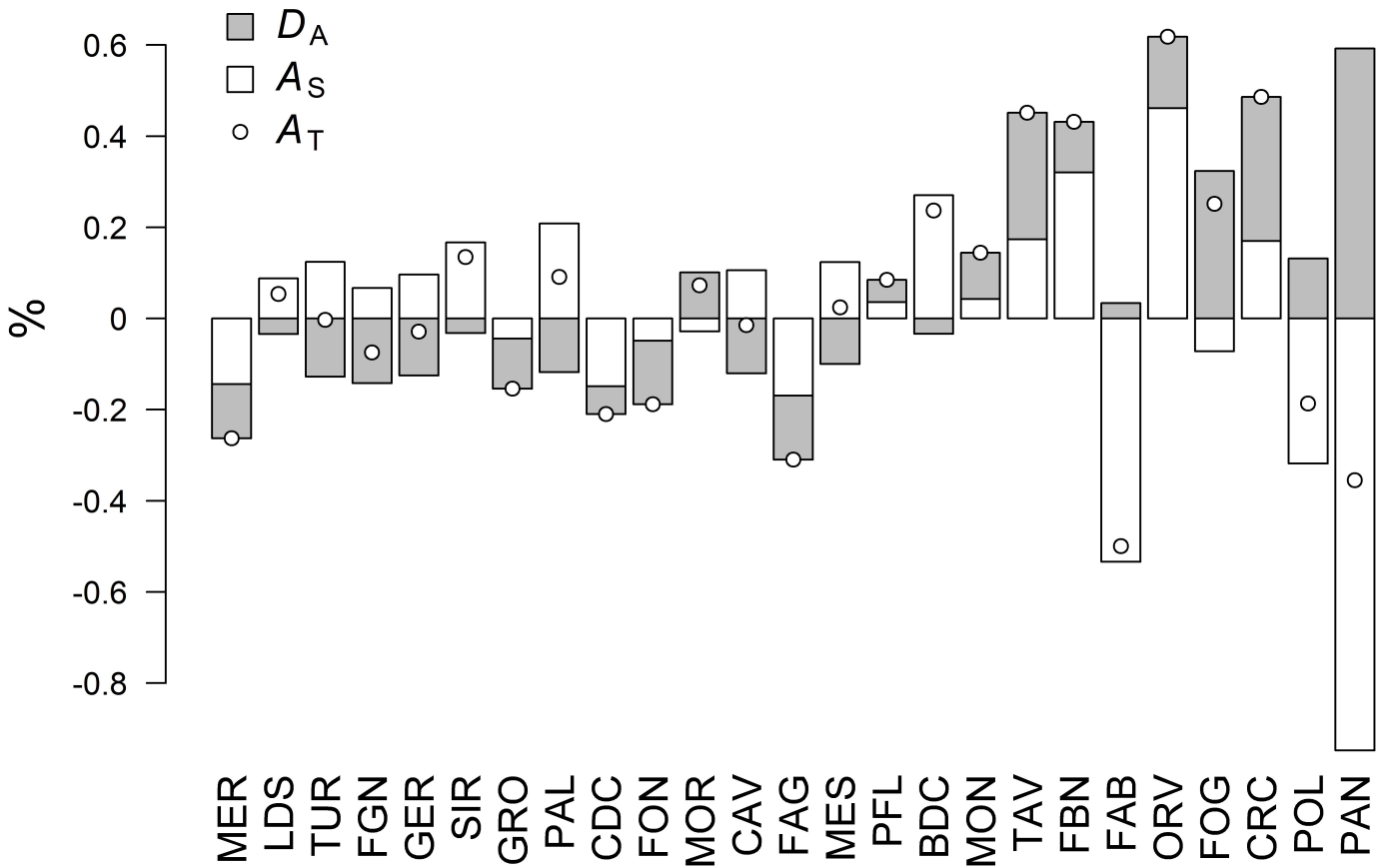
Scenario	Posterior probability	95% CI (lower - upper)
1	0.9096	0.8925 - 0.9266
2	0.0599	0.0462 - 0.0736
3	0.0305	0.0200 - 0.0411

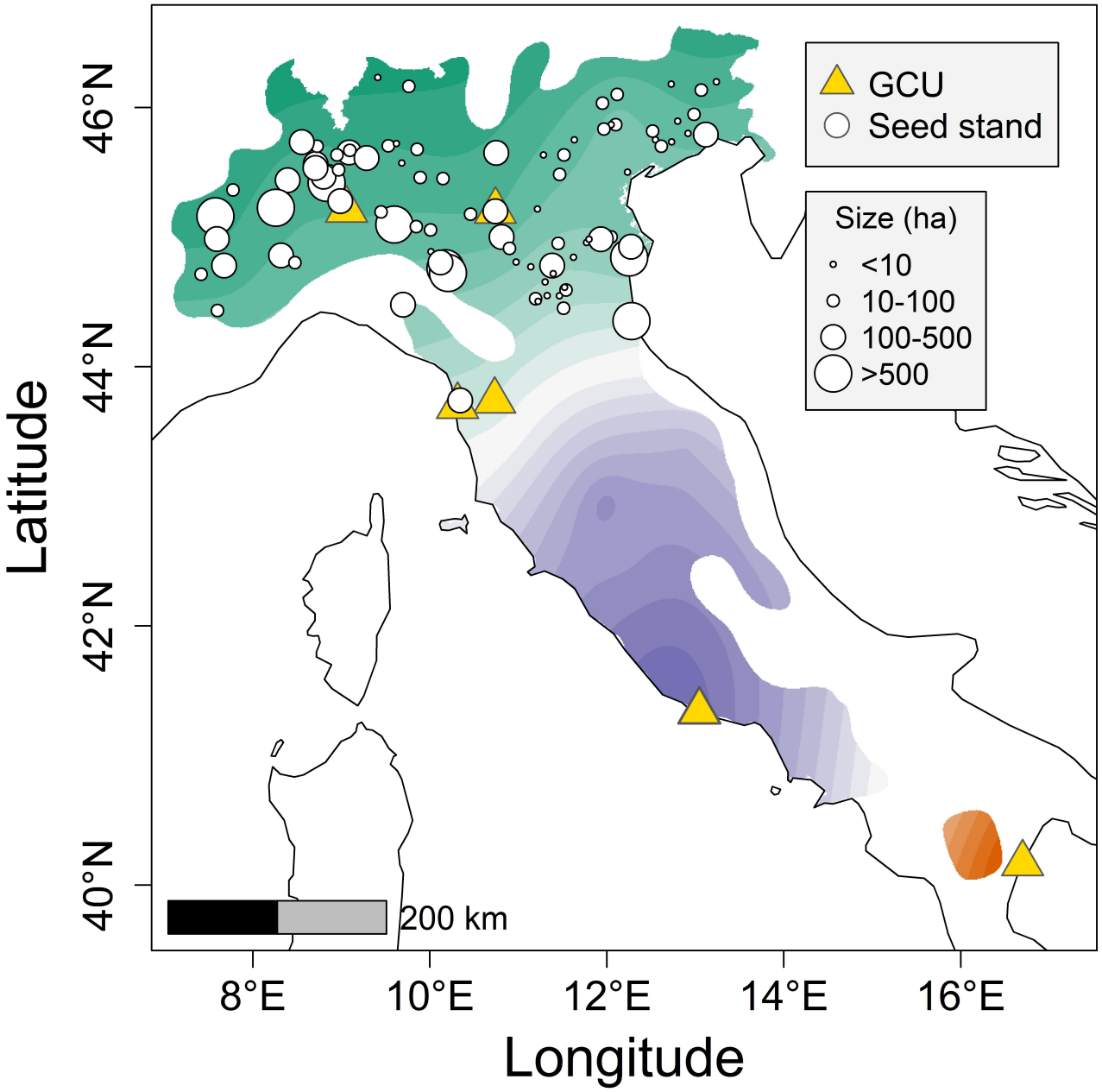

8°E
10°E
12°E
14°E
16°E
18°E
47°N
45°N
43°N
41°N
39°N
Sampled populations
200 km
FGN
TUR
GRO
FAG
MOR
CAV
PAL
LDS
CDC
FON
MER
GER
SIR
MES
BDC
PFL
TAV
MON
Ligurian Sea
FBN
FAB
ORV
Adriatic Sea
POL
FOG
CRC
PAN
Tyrrhenian Sea
Ionian Sea
Q. robur distribution
N
1000 km


a)
Scenario 1
Na
Nb
rd
rc
rb
ra
N1
N2
N3
Scenario 2
Na
Nb
rd
ra
N1
N2
N3
t6
t5
t4
t3
t2
t1
0
Scenario 3
Na
Nb
rc
rb
N1
N2
N3
Scenario 4
Na
Nb
rd
rb
N1
N2
N3
t6
t5
t4
t3
t2
t1
0
b)
c)
Na = 6,340
t6 = 156,600
Nb = 40,100
t5 = 72,600
rd = 0.15
t4 = 20,130
ra = 0.90
t1 = 3,330
NI
N1 = 6,280
CI
N2 = 16,100
SI
N3 = 2,440


K = 2
a)
K = 3
b)
K = 6
c)
d)
MER
LDS
TUR
FGN
GER
SIR
GRO
PAL
CDC
FON
MOR
CAV
FAG
MES
PFL
BDC
MON
TAV
FBN
FAB
ORV
FOG
CRC
POL
PAN
ΔK
K
Mean Ln P(K)
0
20
40
60
80
100
1
2
3
4
5
6
7
8
9
10
-39250
-38750
-38250
-37750


a)
mean + 2σ
mean + 1σ
mean
mean - 1σ
mean - 2σ
46°N
44°N
42°N
40°N
38°N
Ar
8.0
7.5
7.0
6.5
6.0
5.5
R2 = 0.40
40
42
44
46
Latitude
b)
FST
0.15
0.10
0.05
R2 = 0.76
40
42
44
46
Latitude


D_A
A_S
A_T
%
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
MER
LDS
TUR
FGN
GER
SIR
GRO
PAL
CDC
FON
MOR
CAV
FAG
MES
PFL
BDC
MON
TAV
FBN
FAB
ORV
FOG
CRC
POL
PAN


GCU
Seed stand
Size (ha)
<10
10-100
100-500
>500
200 km
46°N
44°N
42°N
40°N
Latitude
8°E
10°E
12°E
14°E
16°E
Longitude