Secondary contacts between European white oaks reveal genes underlying reproductive isolation

Thibault Leroy[1], Quentin Rougemont[2], Catherine Bodénès[1], Céline Lalanne[1], Caroline Belser[3], Karine Labadie[3], Grégoire Le Provost[1], Jean-Marc Aury[3], Antoine Kremer[1], Christophe Plomion[1]

1/ BIOGECO, INRA, Univ. Bordeaux, 33610 Cestas, France

2/ Département de biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, G1V 0A6, Québec, Canada

3/ CEA - Institut de Biologie François Jacob, Genoscope, 2 rue Gaston Crémieux, 91057 Evry, France

*** Corresponding author:**

Antoine Kremer

INRA, UMR1202 BIOGECO, F-33610 Cestas, France

Phone number: +33(0)5 57 12 28 32

Fax number: +33(0)5 57 12 28 81

KEYWORDS:

Genome scan, approximate Bayesian computation, demographic inferences, intrinsic and ecological barriers, speciation genes

SUMMARY

Oaks are dominant forest tree species widely distributed across the northern hemisphere, where they constitute natural resources of economic, ecological, social and historical value. Hybridization and adaptive introgression have long been thought to be major evolutionary drivers of the ecological success of oaks. However, the maintenance of species barriers remains a key question, given the extent of interspecific gene flow. We used a demographically-explicit genome scan approach combining approximate Bayesian computation and backward simulations to scan the genomes of four European white oak species for reproductive barriers. We first confirmed earlier findings of interspecific contact since the last glacial maximum. We then made use of the tremendous genetic variation among these four species (31 million SNPs) to identify genomic regions enriched in SNPs deviating from demographic expectations under a hypothesis of recent secondary contact. The outlier-enriched regions detected were of restricted size and widely distributed, over all the chromosomes. We identified 215 regions, containing 227 oak genes. A Literature-based functional annotation of these genes highlighted important functions driving the reproductive isolation of the four species. These functions included tolerance to biotic and abiotic constraints, and intrinsic mating barriers. The functions of the annotated genes were highly consistent with the ecological preferences of the four species. These genes may, therefore, be considered candidate genes in studies of species responses to past and ongoing environmental changes.

INTRODUCTION

Since Classical times, trees have been one of the most powerful cultural and biological symbols common to different human societies and religions (1), to such a point that the current popular view of forest ecosystems remains somewhere between imaginary cultural ideals and reality. The recent worldwide success in book stores of the "hidden life of trees", depicting forests from a panglossian viewpoint, provides a striking illustration of the popular perception of forests and trees. However, the symbolic strength, longevity and stability underpinning these views are being called into question by continual advances in our understanding of population and evolutionary genetics, which have revealed trees to have very high levels of variability over space and time. In particular, reconstruction of the historical trajectories of several tree species has revealed dynamic systems involving rapid migration and extensive gene exchanges between interfertile species occurring in sympatry.

Oaks are a diverse group of about 350 to 500 species (2, 3). The variability in the number of recorded species highlights the challenge of delineating species limits within a genus displaying a high degree of morphological diversity, sometimes described as a "botanic horror" by morphological taxonomists (4–6). Genetic markers have corroborated these taxonomic concerns, particularly in European white oaks, which have been the subject of a large number of surveys. Studies based on nuclear DNA markers have reported unambiguously high levels of admixture between European white oak species, confirming the reported taxonomic issues for oaks (7). Several detailed empirical studies based on chloroplast DNA markers have revealed an absence of private chlorotypes between European white oak species, but congruent associations between chlorotypes and expansion routes during the last postglacial recolonization, suggesting cytoplasmic capture via recurrent hybridization and backcrossing (8, 9). Recent advances in oak genomics (10, 11) have made it possible to investigate interspecific gene flow at the

73 whole-genome scale. Indeed, Leroy et al. (12) have provided evidence suggesting that
74 extensive secondary contacts have occurred between four European white oak species,
75 probably at start of the current interglacial period. These results reconcile earlier
76 findings of population differentiation at the nuclear and organelle levels. The inferences
77 drawn are also consistent with the persistence of genomic regions impermeable to gene
78 flow due to functional reproductive barriers, corresponding to a typical case of semi-
79 isolated species (12). However, the genetic basis of these barriers remains unknown.

80 Controlled pollination trials have provided empirical evidence for the existence of
81 strong reproductive barriers in these four European white oak species (13, 14). Ecological
82 preferences *in situ* have also been widely reported, with tolerance to dry (*Q. petraea*) or
83 wet (*Q. robur*) sites (15), or acidic (*Q. pyrenaica*) or limy (*Q. pubescens*) soils (16), The
84 four species occupy different geographic ranges: extending up to Scandinavia for *Q.*
85 *petraea* and *Q. robur*, whereas the other two species are present mostly in
86 Mediterranean and sub-Mediterranean regions. However, the distribution ranges of
87 these species overlap in some areas, mostly in South-West France, but the four species
88 are rarely found together in the same stand (but see ref. 7), providing further evidence
89 for a role of ecological factors in shaping the spatial distribution of these species. The
90 overlapping species ranges in South-West France thus provide an ideal "natural
91 laboratory" (17) for investigating reproductive barriers between these European white
92 oak species. To sum up, the current species distributions isolate species according to
93 ecological gradients, with a southern distribution of *Quercus pubescens* and *Q. pyrenaica*
94 and a more temperate distribution of *Q. petraea* and *Q. robur*.

95 One widely used conventional means of studying the genetic basis of reproductive
96 isolation is to scan genomes for loci displaying departure from expectations under
97 neutrality based on between-locus variation in fixation indices under an infinite-island
98 (IM) model. Lewontin and Krakauer (18) introduced this method in an article that was

99 initially very badly received (*e.g.* ref. 19), launching heated debate between population
100 geneticists. Four decades later, the performance of these tests has clearly improved, but
101 the debate continues (Hermisson, 2009). The main long-standing criticism of this method
102 relates to the robustness of these tests to departure from the simple underlying IM
103 model. Simulation studies have shown that genome scans are generally robust with
104 respect to demographic events, even for simulations under complex demographic
105 scenarios (*e.g.* ref. 21, 22), with the notable exception of non-equilibrium scenarios, such
106 as recent secondary contact (<1000 generations ago, ref. 23). As previously reported,
107 European white oak genomes display strong signals of a convoluted evolutionary history
108 (12), calling into question the suitability of traditional genome-scan methods.

109 Rather than assuming moderate departures from an IM scenario, an alternative
110 strategy to genome scanning involves the generation of null distributions under realistic
111 scenarios (20). This approach requires identification of the most likely evolutionary and
112 demographic history of species, followed by the simulation of large data sets under this
113 scenario, assuming neutrality. This strategy has been surprisingly little used (24-27),
114 and, to our knowledge, has never been applied to large-scale genome data. We used
115 approximate Bayesian computation (ABC) and backward simulations (Fig. 1) (i) for
116 ascertainment bias-free demographic inferences and to refine estimates of the timing of
117 secondary contacts, (ii) to generate a null distribution for scanning European white oak
118 genomes for reproductive barriers and (iii) to identify key genes involved in intrinsic and
119 extrinsic (ecological) reproductive barriers.

120

121

122

123

124 # RESULTS

125 Strong signatures of secondary contact between European white oaks

126 In total, 31,894,340 SNPs were identified after the filtering of variants with a low
127 minor allele frequency (MAF<0.02) in population samples for the four species (*Q.*
128 *petraea, Q. robur, Q. pubescens, Q. pyrenaica).* Given the haploid genome size of oaks
129 (740 Mb, ref. 11), this considerable genetic variation corresponds to one SNP every 23.2
130 bp, on average. We also used genome-wide data for a *Q. suber* accession described by
131 Leroy *et al.* (12) to root a phylogenetic tree and investigate relationships between species
132 for 9,084,835 of the 31.9 million SNPs. The best maximum-likelihood tree suggested that
133 *Q. robur* initially diverged from the ancestor of the other three species (Fig. S1). Our
134 analysis revealed a second split between *Q. pyrenaica* and the ancestor of *Q. petraea* and
135 *Q. pubescens*. These last two species appear to have diverged more recently.

136 We randomly selected 50,088 SNPs from the entire set of 31.9 million SNPs for
137 ascertainment bias-free ABC inference. We compared two models of divergence with
138 gene flow for each of the six possible species pairs: an isolation-with-migration model
139 assuming constant gene flow since the divergence time (T_{SPLIT}), and a model assuming
140 secondary contact with gene flow starting at T_{SC}, a time point after divergence (T_{SC} <
141 T_{SPLIT}). For all pairs, we obtained strong statistical support for the secondary contact
142 model (>98% posterior probability, Table S1), consistent with our previous findings
143 based on individual data for 3,524 SNPs (12). As in our previous report, our ABC
144 inferences were found to be highly robust to alternative models based on 1000 pseudo-
145 observed datasets, indicating that the probability of an incorrect inference of secondary
146 contact was very small (Fig. S2).

147 We then generated parameter estimates under the best-fitting secondary contact
148 model. After confirming that recent secondary contact had occurred between these four

149 species, we ran additional backward simulations explicitly assuming recent secondary
150 contact (Tsc < 5% Tsplit), to ensure high accuracy for parameter estimates. Based on
151 these two rounds of analyses, posterior distributions of parameters were built from the
152 500 simulations closest to each observed dataset. As expected, this additional simulation
153 improved the precision of some parameter estimates, particularly for T_{SC} (Table S2). For
154 all pairs, T_{SC} was estimated to have occurred between 0.02% and 0.22% of species pair
155 divergence times, for the species pairs *Q.robur*/*Q.pubescens* and *Q.pubescens*/*Q.petraea*,
156 respectively.

157

158 Highly heterogeneous genetic differentiation landscape

159 Based on F_{ST} values calculated over 10-kb genome segments, the genetic
160 differentiation between species pairs differed considerably between chromosomes (Fig.
161 2), and, more interestingly, between segments within chromosomes (Fig. 3)

162 First, with rare exceptions, the three most strongly differentiated species pairs,
163 for all oak chromosomes, were *Q. robur/Q. pubescens*, *Q. robur/Q. petraea* and *Q.*
164 *pyrenaica/Q. petraea*. Lower levels of differentiation were observed for *Q. pubescens/Q.*
165 *pyrenaica* and *Q.pubescens/Q.petraea* (Fig. 2A). These results suggest that the
166 demographic history of these populations shaped the genomic landscape of
167 differentiation such that the ranking of pairs of species according to mean F_{ST} values is
168 conserved over chromosomes. Conversely, regardless of the pair of species considered,
169 the interchromosomal variation of mean F_{ST} values was considerable (Fig. 2B), with
170 significantly higher mean F_{ST} values for chromosomes 2 and 6 and significantly lower
171 values for chromosome 4. For all pairs of species, the relationship between mean F_{ST} and
172 the rate of recombination was significant at the chromosome scale ($P < 0.05$; Fig. S3), for
173 calculations based on the physical and genetic length of the 12 chromosomes (11,28).

174 Second, genome-wide patterns of F_{ST} variation plotted over the entire genome
175 with 10-kb sliding windows (Fig. 3) revealed a highly heterogeneous differentiation
176 landscape, even for the pair of species displaying the lowest level of differentiation
177 according to F_{ST} values (*Q. pubescens/Q. robur*). Our 10-kb F_{ST} estimates ranged from 0
178 to 0.765. For the pairs of species including *Q. petraea*, the values for the window
179 corresponding to the highest level of differentiation were0.998 for *Q. petraea/Q.*
180 *pyrenaica*, 0.999 for *Q. petraea/Q. robur* and 1 for *Q. petraea/Q. pubescens*. On closer
181 inspection, very narrow regions of very high F_{ST} were identified on several chromosomes,
182 for most pairs of species (Fig. 3, chromosomes 1, 7, 9, and 11, for example). These local
183 signatures of strong differentiation suggest a combination of both (i) high levels of
184 selection to maintain such high levels of differentiation in the face of the high level of
185 inferred gene flow and (ii) very high population recombination rates, breaking down
186 associations between the differentiated and flanking loci.

187

188 Narrow regions of non-neutral evolution

189 We developed a genome-scan strategy based on differentiation outlier tests, to
190 account explicitly for secondary contact. Unlike methods based on deviations from
191 neutrality under an IM scenario, our method is based on extensive backward simulations
192 (5,000,000 independent SNPs) under the best inferred scenario for each pair of species,
193 to generate null distributions of G_{ST} conditioned by expected heterozygosity. The
194 application of this method revealed that recent SC led to substantial variation of the null
195 distribution of G_{ST} values, with some SNPs reaching fixation in two pairwise
196 comparisons (dotted lines, Fig. 4). We then performed a genome-scan strategy based on
197 the original approach of Beaumont & Nichols (29), with the aim of identifying SNPs
198 deviating from the distribution of fixation indices conditional on heterozygosity (solid
199 lines in Fig. 4).

200 We generated quantiles at the upper end (99.99%) of the null distribution of G_{ST} and
201 compared the G_{ST} for each of the 31.9 million SNPs for each pair of species to its
202 threshold value. After filtering for SNPs with at least moderate to high levels of
203 heterozygosity (He≥0.2), we identified 1.18% to 4.90% of SNPs as outliers, depending on
204 the pair of species considered. This filtering was used because the proportion of outliers
205 was highly inflated in three pairs of species (2.02 to 18.67%, depending on the pair
206 considered, Fig. S4). However, for all pairs containing *Q. petraea* (Fig. S4), the
207 proportion of outliers at lower diversity values was only slightly high. We decreased the
208 negative impact of false positives, by limiting the genomic scan to SNPs with He values
209 exceeding either 0.3 or 0.2, and assessing the local density of outliers along the genome
210 with a 10-kb sliding window approach (Fig. 5 & S5). For each pair of species, most
211 outlier-enriched windows were retained for the identification of candidate speciation
212 genes.

213 We identified 281 windows containing the highest proportion of outliers (top 0.1% of
214 window enriched in outliers for at least one pair of species). We then analyzed the
215 clustering of these outlier-enriched windows at different genomic scales. We first defined
216 a candidate genomic region (Fig. 1) consisting of the contiguous sequences of two outlier-
217 enriched windows, with interruption by a single undetected window authorized. The 281
218 windows were distributed over 215 candidate genomic regions (Fig. 5). We then zoomed
219 out to chromosomal level, at which we observed the clustering of several different
220 candidate genomic regions close together in the genome, consistent with the presence of
221 longer regions of non-neutral signatures (blue lines, Fig. 5). A comparison of two
222 analyses based on different minimum He threshold values for outliers (He ≥ 0.2 Fig. 5
223 and He ≥ 0.3; Fig. S5) revealed that a high frequency of 10-kb windows was identified in
224 both analyses (58%), with similar footprints at the chromosomal scale (Fig. S5). These

225 findings suggest that the local enrichment strategy is robust to a low-to-moderate excess
226 of false positives.

227

228 Ecological and non-ecological candidate genes underlying reproductive isolation

229 For identification of the best candidate gene for each region, we listed all the
230 *Quercus* genes located within or flanking these 215 genomic regions (Fig. 1 step 5). We
231 identified 227 genes distributed over 133 of the 215 regions, with very few candidates
232 per region (from 1 to 18 genes (mean: 1.71±1.76) per region). After the exclusion of five
233 regions with strong signatures of chloroplast-like DNA probably reflecting different
234 chlorotypes due to cpDNA contamination in the oak assembly, the number of candidate
235 genes per region was even lower (range: 1 to 6; mean: 1.49 ±0.82). We then used a
236 combination of methods for the manual annotation of all these genes (BlastP and
237 extensive literature searches; File S1). Given the narrowness of the candidate genomic
238 regions of non-neutral evolution, we assumed that no more than one gene per candidate
239 genomic region contributed to reproductive isolation. Consequently, for candidate
240 regions containing several oak genes, we manually identified a single best candidate
241 gene from all the gene functions identified by Blast analysis and extensive literature
242 searches (File S1). Gene functions were determined for the 109 best *Quercus* gene
243 candidates from the 133 genomic regions. Additional screening based on the quality and
244 sharpness of the annotation in terms of physiological function in published studies
245 resulted in a set of 32 genes. We restricted our analysis of species differentiation and
246 functional interpretation to these genes (Table 1). However, this approach does not rule
247 out the contribution of other genes. For this reason, we have added a summary of all the
248 available gene annotations to Table S3, and a more detailed version to File S1. With the
249 notable exception of Qrob_P0641650.2 and Qrob_P0430150.2, encoding G-type lectin S-
250 receptor-like serine/threonine-protein kinases, candidate genes have few if any

251 paralogous genes in their corresponding orthogroups, according to the gene models of 16
252 plant species (Ref. 11, Table 1). We detected a large number of genes encoding regulators
253 of gene expression, including transcription factors (Table S3). Several different
254 situations were observed for the enriched regions common to several pairs of species, for
255 these reliable candidates (Table S4: (i) 6 regions enriched in outliers for all pairs of
256 species, (ii) 9 regions enriched in outliers for all but one pair of species (7 regions for all
257 pairs except *Q. robur* – *Q. petraea* and 2 regions for all pairs except *Q. pubescens* – *Q.*
258 *pyrenaica* pairs), (iii) 5 regions specific to all pairs containing the same species (4 for *Q.*
259 *pyrenaica* and 1 for *Q. pubescens*) and (iv) 11 with more complex patterns.

260 An analysis of the annotation of 32 identified genes led to these genes being
261 clustered into three major categories (Table 1). The first of these categories corresponded
262 to genes underlying the ecological preferences of the four species: tolerance of water
263 deficit, cold tolerance, adaptation to alkaline soils. The second category corresponded to
264 genes involved in biotic interactions, such as immune responses, resistance to biotic
265 stresses, and mycorrhization. The third category corresponded to genes probably
266 contributing to intrinsic barriers, and included genes involved in flowering time, pollen
267 recognition, pollen growth and embryo development.

268

269

270 DISCUSSION

271 The identification of reproductive barriers to gene flow is a long-standing
272 challenge in evolutionary biology. The increasing available of genomic resources for
273 different species is helping to improve our understanding of the genomic basis of
274 reproductive isolation, but we also need to update genome-scan methods for application
275 to large sets of genomic data and to account for confounding demographic effects more

276 accurately. In this study, we made use of the recently published *Q. robur* reference
277 genome (10, 11) and description of the evolutionary history of European white oaks (12)
278 to develop a strategy combining an approximate Bayesian computation framework,
279 extensive neutral simulations and the analysis of large amounts of pooled whole-genome
280 resequencing data for scanning white oak genomes for reproductive barriers. We discuss
281 the main findings of this study below: (i) evidence supporting the occurrence of extensive
282 secondary contact following the last postglacial recolonization of Europe, (ii) a
283 description of the genetic landscape of differentiation and (iii) the genetic basis of
284 reproductive isolation between these four species.

285 Our ABC inferences revealed unambiguous signals of secondary contact in the
286 evolutionary history of European white oaks, consistent with our previous report (12).
287 Even if the results reported here are in line with our previous findings, new inferences
288 can be drawn, thanks to the methodology used and the genome-wide exploration
289 performed. First, our ascertainment bias-free analyses provide support for the
290 occurrence of secondary contact. The previous inferences were drawn from a dataset with
291 a strong ascertainment bias. Such a deviation from the true site frequency spectrum is
292 known to have a negative impact on the performance of likelihood-free analyses (30). We
293 overcame this problem, by performing new analyses based on tens of thousands of SNPs
294 randomly sampled from throughout the genome. This strategy yielded very similar
295 results, confirming the occurrence of secondary contacts between these four white oak
296 species. Second, our use of 10 times as many SNPs and a two-round procedure in this
297 study made it possible to narrow down the time frame over which these contacts took
298 place. This strategy improved estimates of the timing of secondary contact by at least
299 one order of magnitude ($T_{SC} < 0.5\%\ T_{SPLIT}$; Table S2). Assuming that T_{SPLIT} occurred
300 between 1 and 5 million years ago (MYA) (3) or between 7 and 10 MYA (31), and taking
301 into account the 95% confidence intervals for each T_{SC}/T_{SPLIT} ratio (Table S2), we

302 estimate that secondary contact occurred between 100 and 62,400 years ago,
303 corresponding to up to 1,225 generations, assuming a generation time of 50 years (32),
304 whereas the upper bound for our previous estimate was 11,200 generations (12). Overall,
305 the median estimates of the timing of secondary contact ranged from 2,450 to 21,760
306 years depending on the species pair considered (49 to 435 generations), consistent with
307 the general hypothesis of a resumption of secondary gene flow at the start of the current
308 interglacial period.

309 Chromosomal F_{ST} estimates differed considerably between pairs of species and
310 chromosomes. For all pairs, inter-chromosomal F_{ST} variation was significantly correlated
311 with variations of mean chromosomal recombination rate, as expected due to the
312 variation of linkage due to positively selected alleles (genetic hitchhiking, ref. 33) or to
313 deleterious alleles (background selection, ref. 34). The recent report of a remarkably high
314 rate of deleterious mutations relative to 28 other plant species (11) highlights the
315 putative role of deleterious mutations in shaping the genomic landscape of species
316 differentiation in oaks. The contribution of linkage to the genomic width of reproductive
317 barriers appears to be limited, as no long stretches of extensive differentiation were
318 observed. We identified very restricted, widely distributed "islands" of high
319 differentiation. The high-resolution genetic landscape of differentiation observed with
320 scattered microislands of high F_{ST} on otherwise poorly differentiated chromosomes
321 contrasts sharply with many other F_{ST} scan studies reporting either continents of
322 differentiation (e.g. ref. 35) or islands of high differentiation in highly structured
323 populations (e.g. ref. 36). Extensive secondary gene flow over the last 20,000 years,
324 together with very high levels of prezygotic and postzygotic selection (13, 14), probably
325 eroded interspecific genetic structures other than those at barrier loci, thereby
326 generating a highly heterogeneous differentiation landscape. European white oaks are

327 thus ideal species for genome scans, providing new opportunities for the identification of
328 genes underlying reproductive isolation.

329 We made use of the detection of recent secondary contact between white oaks to
330 develop a new demographically explicit genome-scan method. With the exception of a few
331 case studies (24-27), this approach has clearly been neglected in genome-scan studies,
332 despite concerns about demographic confounding effects (20). Basically, rather than
333 considering both expectations under an IM model and assuming only slight departure of
334 our dataset from this IM scenario, we performed extensive neutral simulations directly,
335 assuming inferred parameters under each secondary contact scenario, to generate null
336 distributions for the identification of outlier SNPs. This approach is particularly relevant
337 to European white oaks, in which secondary contact began less than 1,225 generations
338 ago and would therefore be expected to generate substantial deviation from a simple IM
339 model (23). Interestingly, although rare, we identified several neutral simulations under
340 SC scenarios that resulted in fixed private alleles (G_{ST} = 1) for two pairs of species. It is,
341 therefore, the issue of false-positives that cannot be resolved simply by adjusting
342 thresholds or increasing penalization in corrections for multiple testing, because the
343 diagnostic loci would still be expected to have some of the best corrected test scores. An
344 alternative and more straightforward strategy suitable for situations with higher false
345 positive rates involves scanning genomes for outlier-enriched regions in a sliding-window
346 approach. Additional evaluations of our genome-scan strategy are required on larger
347 datasets, including other species, but the high level of consistency between the two
348 enrichment analyses performed here, based on different threshold levels of diversity,
349 suggests alternative methods for decreasing (Fig. S4) the rate of false positives (Fig. 5).
350 In future studies, we recommend i) discarding outliers detected at very low He, as
351 generally assumed for genome scans (ref. 29 Fig. S4) before the enrichment analysis step
352 and ii) checking for unexpectedly high proportions of outlier loci for some classes of

353 heterozygosity (Fig. S4). These pragmatic recommendations should help to refine the
354 discovery of informative SNPs, by preventing the consideration of false positives.

355 The SNPs in candidate genes contributing to reproductive barriers
356 displaying the highest degree of interspecific differentiation included some
357 differentiating between southern (*Q. pyrenaica* & *Q. pubescens*) and northern species (*Q.*
358 *robur* & *Q. petraea*). This trend is incongruent with phylogeny (ref. 12; Fig. S1), and
359 therefore suggests that there are selective factors triggered by ecological drivers. Indeed,
360 our short list of candidate genes (Table 1) includes nine with an "all-versus-one-pair"
361 relationship; seven exclude the *Q. robur*/*Q. petraea* pair and the other two exclude the *Q.*
362 *pubescens* /*Q. pyrenaica* pair. Four of the seven genes for which the *Q. robur*/*Q. petraea*
363 pair was excluded are known to be involved in the drought tolerance or in lateral root
364 growth (Table 1). This pattern is consistent with the higher drought tolerance in *Q.*
365 *pyrenaica* and *Q. pubescens* than in *Q. petraea* or *Q. robur* (37). Reciprocally, we
366 observed an "all vs. one" pattern (undetected for the *Q. robur*/*Q. petraea* pair) for a
367 VRN1 gene involved in responsiveness to vernalization and known to play a key role in
368 cold acclimation in many plant species (Levy et al. 2002). Overall, the genomic variation
369 of these nine genes parallels the northern-southern distribution of the study species,
370 suggesting that the underlying barriers are driven by ecological preferences. Given that
371 dehydration-associated genes currently act as strong barriers between these species in
372 South-West France, drought tolerance alleles from *Q. pubescens* and *Q. pyrenaica* would
373 not be expected to introgress easily into the *Q. petraea* or *Q. robur* genomes and cannot,
374 therefore, drive rapid evolutionary changes through adaptive interspecific introgression.
375 As a consequence, the increase in the number of periods of drought in European forests
376 due to climate change is likely to lead to a northward migration of *Q. petraea* and *Q.*
377 *robur* at the continental scale, and more locally, to the gradual replacement of these
378 species with more drought-tolerant species. In this context, *Q. robur* appears to be the

379 species most under threat, as it is the most susceptible to severe water deficits of the
380 four species studied. A recent analysis of fine-scale species distributions reported
381 significant changes over only two generations in a mixed oak stand composed of *Q. robur*
382 and *Q. petraea,* supporting the trend for *Q. robur* to be replaced by *Q. petraea* following
383 repeated drought events at the juvenile stage (39).

384 We also identified five genes with branch-specific patterns, *i.e.* found in all pairs
385 containing a given species, four for *Q. pyrenaica* and one for *Q. pubescens*. The four *Q.*
386 *pyrenaica*-specific patterns concerned genes involved in plant immunity, three encoding
387 G-type lectin S-receptor-like serine/threonine kinases (LECRKs) and one encoding a
388 transportin (MOS14). *Q. pyrenaica* is extremely sensitive to oak powdery mildew, a
389 pathogen that was introduced into Europe at the start of the 20th century (40). Soon
390 after the first detection of the fungus in Europe, high mortality rates were reported for
391 *Q. pyrenaica* in the humid warm-temperate forests of southwestern and western France
392 (40 and references therein). *Q. pyrenaica*-specific alleles at several genes involved in
393 immunity may, therefore, be the signature of the high susceptibility of *Q. pyrenaica* to
394 biotic stresses in moist environments. Similarly, we identified a *Q. pubescens*-specific
395 pattern for a gene encoding a metal transporter (Nramp5) involved in the assimilation of
396 manganese and cadmium in rice and barley (41). Manganese assimilation is known to be
397 essential for many plant functions, but manganese availability in the soil tends to
398 decrease with increasing pH, and becomes limiting beyond a soil pH of 6.5. This gene
399 probably contributes to the greater ecological preference of *Q. pubescens* than of the
400 other three species for lime-rich soils.

401 In addition to ecological barriers, we found support for both intrinsic premating
402 and postmating barriers. We identified several genes involved in the timing of flowering,
403 including APETALA2 or PRR73. APETALA2 is a key transcription factor for the
404 establishment of the floral meristem (42). Similarly, PRR73 contributes to flowering time

405 variation in barley and wheat (43 and references therein). The genes involved in mating
406 barriers may also include previously described ecological genes with pleiotropic effects,
407 such as *VRN1*. In addition to its primary role in vernalization, VRN1 is involved in the
408 repression of *FLC* (itself a known repressor of flowering) in *Arabidopsis,* through a
409 vernalization-independent floral pathway (38). Several of the genes identified are
410 involved in pollen or embryo development, suggesting that both premating and
411 postmating intrinsic barriers operate in oaks. Three of these genes encode cycloartenol
412 synthases known to be essential for pollen development in *Arabidopsis* (44).

413 Thus, European white oaks, with their extensive recent secondary gene flow,
414 constitute an ideal complex of semi-isolated species for investigations of the genetic basis
415 of reproductive isolation. By applying a population genomics approach based on
416 demographically explicit genome scans in four oak species, we have moved beyond the
417 current limits of genome scans for species with non-equilibrium evolutionary histories.
418 This pioneering work focuses on the strongest signatures of non-neutral evolution (*i.e.*
419 without aiming to catalog all the barriers in operation), but this strategy nevertheless
420 identified key genes associated with both intrinsic and ecological barriers, including
421 genes involved in flowering and tolerance to biotic and abiotic stresses. These genes are
422 consistent with the ecological preferences of the species, including their
423 northern/southern distribution in particular, and, therefore, provide new opportunities
424 for exploring adaptation to ongoing environmental changes. This work constitutes an
425 important step towards prediction of the effects of global warming and management of
426 the renewal of European oak forests in this context.

427

428

429

430 # MATERIALS & METHODS

431 ***Sampling and DNA extraction***

432 We sampled populations of the four *Quercus* species in stands of natural origin
433 located in South-West France. We sampled 13 *Q. petraea* trees in Laveyron (Landes,
434 France), and 20 *Q. robur* and 20 *Q. pyrenaica* trees from the Landes EVOLTREE
435 "Intensive Study Site" (ISS). We also sampled 18 *Q. pubescens* trees from two sites in
436 Gironde: 12 in Branne and six in Blaignan (Gironde, France) (see Table S5 for details).
437 DNA was extracted from individual trees with the Invisorb Spin Plant Mini kit,
438 according to the manufacturer's specification (Startec Molecular, GmbH, Berlin,
439 Germany). DNA yields were evaluated with a NanoDrop 1000 spectrophotometer
440 (NanoDrop Technologies, Wilmington, DE, USA) and DNA samples were mixed in
441 equimolar amounts to obtain a single pool for each species.

442

443 ***Library preparation and sequencing***

444 For each pool, a paired-end DNA genomic library was generated with the Paired-
445 End DNA Sample Preparation Kit (Illumina, San Diego, CA, USA). The library was then
446 sequenced on a HiSeq2000 sequencer (Illumina, San Diego, CA, USA) with 2x100 paired-
447 end reads. For each pool, we used nine to ten sequencing lanes. Raw reads were trimmed
448 to remove low-quality bases (<20) from the ends, and sequences between the second
449 unknown nucleotide and the end of the reads were removed. Reads less than 30
450 nucleotides long after trimming were discarded. Overall, between 1,617,465,418 and
451 1,813,403,677 reads per pool were retained for analysis, corresponding to 313 Gb (425X)
452 to 356 Gb (483X) of raw data. Raw data have been deposited in the Sequence Read
453 Archive (SRA): ERP105626.

454

455 **Mapping and calling**

456 All reads were then mapped against the v2.3 oak haplome assembly (11), with
457 bowtie2 v. 2.1.0 (45), using standard parameters for the "sensitive end-to-end" mode.
458 PCR duplicates were removed with Picard v. 1.106
459 (http://broadinstitute.github.io/picard/). Samtools v.1.1 (46) and Popoolation2 v. 1.201
460 (47) were then used to call biallelic SNPs with at least 10 alternate alleles and a depth
461 between 50 and 2000X at each position. To ensure a reasonably low rate of false positives
462 due to Illumina sequencing errors, all SNPs with a MAF lower than 0.02 were discarded.
463 We obtained allele counts for a total of 31,894,340 SNPs. Sliding-window F_{ST} were
464 calculated from allele frequencies with the popoolation2 bioinformatics software suite
465 (47).

466

467 ***Approximate Bayesian computation (ABC) analysis***

468 We used a strategy combining ABC and backward simulations to scan genomes for
469 reproductive barriers (Fig. 1).

470

471 *Observed dataset*

472 For all subsequent ABC analyses, we randomly selected 50,088 of the 31 million
473 SNPs and, for each position, we multiplied allele frequencies by the number of sets of
474 chromosomes in each pool (26 for *Q. petraea*, 36 for *Q. pubescens* and 40 for both *Q.*
475 *pyrenaica* and *Q. robur*). At this stage, we assumed that any departure from
476 equimolarity due to bias in the mixing of DNA samples would have a negligible effect on
477 our summary statistics calculated for a set of 50 thousand SNPs. Seventeen summary
478 statistics were computed by *mscalc:* (1) the number of polymorphic sites specific to each
479 gene pool, (2) the number of polymorphic sites existing in both gene pools, (3) Tajima's pi

480 for each gene pool and, between gene pools, the mean value and standard variation for
481 (4) gross divergence (Dxy), (5) net divergence (Da), (6) F_{ST} and (7) Pearson's R^2
482 correlation coefficient in Tajima's pi test (see also ref. 48).

483 *Simulations*

484 We used an ABC procedure similar to that described by Leroy and coworkers
485 (2017). Briefly, we compared two different model scenarios: isolation-with-migration (IM)
486 and secondary contact (SC). Both model scenarios assumed the subdivision of an
487 ancestral panmictic population (PopAnc) into two daughter populations (Pop1 & Pop2) at
488 time T_{SPLIT}, with population sizes remaining constant over time (N_{PopAnc}, N_{Pop1}, N_{Pop2}).
489 The IM model assumed uninterrupted gene flow since T_{SPLIT}. The SC model assumed
490 that populations initially evolved in strict isolation, with secondary gene flow beginning
491 at some time before the present (at time T_{SC}, see ref. 12).

492 *Coalescent simulations*

493 For coalescent simulations, we adapted the pipeline described by Leroy et al. 2017
494 for the calculation of summary statistics to large datasets (up to 100,000 SNPs). This
495 pipeline includes a modified version of the random prior generator *priorgen* (12, 49), and
496 the *msnsam* (49, 50) and *mscalc* (49) programs. For each of the two scenarios (IM and
497 SC), 1,000,000 random multilocus simulations were performed. Both models made use of
498 genomic heterogeneity in effective migration rates (M) and population sizes (N_e) to take
499 into account the occurrence of genomic barriers to gene flow and the confounding effect of
500 linked selection. For both sources of heterogeneity, we used the strategy described by
501 Leroy et al. (12).

502

503 *Model selection*

504 The 500 best replicate simulations closest to the observed values were selected, and the
505 posterior probabilities for each of the two scenarios (IM or SC) were estimated with a
506 feed-forward neural network, by nonlinear multilocus regression (ref. 12 for details).
507 ABC computations were performed with 20 feed-forward trained neural and eight hidden
508 networks.

509

510 *Parameter estimation*

511 Posterior distributions of the parameters were estimated with a two-step
512 hierarchical procedure. We first evaluated the parameters under the best model, to check
513 the consistency of our estimates, particularly concerning our previous support for very
514 recent secondary contacts (12). We then ran an additional set of 1,000,000 coalescent
515 simulations under an SC model, assuming that T_{SC} occurred in the last 5% of the
516 divergence time, using a modified version of *priorgen*. This strategy was used to obtain
517 more accurate parameter estimates. For both rounds of estimation, we used a logit
518 transformation of the parameters on the 500 best simulations providing the smallest
519 Euclidean distance. The posterior probability of parameters was then estimated by the
520 neural network procedure, from the means of weighted nonlinear multivariate
521 regressions of the parameters on the summary statistics for 25 feed-forward trained
522 neural and 10 hidden networks.

523

524 ***Genome scans***

525 *Coalescent simulations and outlier detection*

526 For each pair of species, we ran 5,000,000 coalescent simulations, using
527 parameter values sampled from the 95% confidence interval (CI) of the posterior
528 distribution of all parameters for the pair considered (Table S2). The simulations

529 assumed genomic homogeneity for effective migration rates (M1 & M2) but heterogeneity
530 for population size. Random values were generated with a modified version of *priorgen*.
531 For each simulated locus, we then calculated He and G_{ST} with a custom-developed script.
532 All scripts and datasets are publicly available for download from a GitHub repository
533 (https://github.com/ThibaultLeroyFr/GenomeScansByABC). Extreme quantiles of the
534 distribution of G_{ST} (99.99% of simulated values) relative to the expected heterozygosity of
535 the locus were then calculated, with a strategy similar to that used for Fdist (29). More
536 specifically, a null envelope was computed from G_{ST} quantiles, with heterozygosity
537 intervals of 2%. This strategy has the advantage of providing a null distribution of G_{ST}
538 under much more complex scenarios (*i.e.* the best inferred scenario) than other state-of-
539 the art genome-scan methods.

540 For each of the 31 million SNPs, we computed the same summary statistics as for
541 neutral markers. The observed G_{ST} values conditioned by heterozygosity at the locus
542 were compared to the previously generated null envelope. Markers with a G_{ST} value
543 above this envelope were considered to be outliers. We then used a non-overlapping
544 sliding-windows approach to estimate the proportion of outliers per 10-kb window.
545 Windows containing fewer than 10 SNPs were discarded.

546

547 **Functional annotations**

548 For all genes found within regions enriched in outliers, we conducted BlastP
549 searches in both the SwissProt and nr protein databases. Only BlastP results with e-
550 values lower than 10e-5 were considered for gene annotation analyses. After
551 identification of the gene by BlastP analyses, extensive manual literature searches were
552 performed. We also reported information from a previous identification of orthologous
553 and paralogous genes in 16 plant species, including *Q. robur,* performed with OrthoMCL
554 (see ref. 11, for details).

555 # ACKNOWLEDGMENTS

556 This research was funded by the French ANR (GENOAK project, 11-BSV6-009-021) and
557 by the European Research Council under the European Union's Seventh Framework
558 Programme (TREEPEACE project, FP/2014-2019; ERC Grand Agreement no. 339728).
559 We thank the Genotoul Bioinformatics Platform Toulouse Midi-Pyrenees (Bioinfo
560 Genotoul) and the Biogenouest BiRD core facility (Université de Nantes) for providing
561 computing and storage resources. We also thank Jorge A. P. Paiva for providing access to
562 *Q. suber* data and Camille Roux for fruitful discussions concerning ABC. We would like
563 to thank fellow members of the oak genome consortium for helpful advice and
564 suggestions.

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579 1. Logan WB (2005) *Oak: The Frame of Civilization* (W.W. Norton & Company, New York)

580 2. Denk T, Grimm GW, Manos PS, Deng M, Hipp AL (2017) An Updated Infrageneric
581 Classification of the Oaks: Review of Previous Taxonomic Schemes and Synthesis of Evolutionary
582 Patterns. *Oaks Physiological Ecology. Exploring the Functional Diversity of Genus Quercus L.*,
583 eds Gil-Pelegrín E, Peguero-Pina JJ, Sancho-Knapik D (Springer International Publishing,
584 Cham), pp 13–38.

585 3. Hubert F, et al. (2014) Multiple nuclear genes stabilize the phylogenetic backbone of the genus
586 Quercus. *Syst Biodivers* 12:405–423.

587 4. Darwin C (1869) *On the Origin of Species by Means of Natural Selection: Or the Preservation*
588 *of Favoured Races in the Struggle for Life* (Murray, London).

589 5. Palmer EJ (1948) HYBRID OAKS OF NORTH AMERICA. *J. Arnold Arbor* 29:1–48.

590 6. Rieseberg LH, Wood TE, Baack EJ (2006) The nature of plant species. *Nature* 440:524–527.

591 7. Lepais O, et al. (2009) Species relative abundance and direction of introgression in oaks. *Mol*
592 *Ecol* 18:2228–2242.

593 8. Petit RJ, et al. (2002) Chloroplast DNA variation in European white oaks: Phylogeography and
594 patterns of diversity based on data from over 2600 populations. *Forest Ecol Manag* 156:5–26.

595 9. Petit RJ, et al. (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proc*
596 *Natl Acad Sci USA* 94:9996–10001.

597 10. Plomion C, et al. (2016) Decoding the oak genome: public release of sequence data, assembly,
598 annotation and publication strategies. *Mol Ecol Resour* 16:254–265.

599 11. Plomion C, et al. (submitted) Oak genome reveals facets of long lifespan.

600 12. Leroy T, et al. (2017) Extensive recent secondary contacts between four European white oak
601 species. *New Phytol* 214:865–878.

602 13. Abadie P, et al. (2012) Strength, diversity and plasticity of postmating reproductive barriers
603 between two hybridizing oak species (*Quercus robur* L. and *Quercus petraea* (Matt) Liebl.). *J Evol*
604 *Biol* 25:157–173.

605 14. Lepais O, Roussel G, Hubert F, Kremer A, Gerber S (2013) Strength and variability of
606 postmating reproductive isolating barriers between four European white oak species. *Tree Genet*
607 *Genomes* 9:841–853.

608 15. Eaton, E, Caudullo, G, Oliveira, S, de Rigo, D (2016) *Quercus robur* and *Quercus petraea* in
609 Europe: distribution, habitat, usage and threats. *European Atlas of Forest Tree Species,* eds San-
610 Miguel-Ayanz, J, de Rigo D, Caudullo G, Houston Durrant T, Mauri A (Publications Office of the
611 European Union. Luxembourg).

612 16. Timbal, J, Aussenac, G (1996) An overview of ecology and silviculture of indigenous oaks in
613 France. *Ann For Sci* 53:649–661.

614 17. Hewitt GM (1988) Hybrid zones-natural laboratories for evolutionary studies. *Trends in*
615 *Ecology & Evolution* 3:158–167.

616 18. Lewontin RC, Krakauer J (1973) Distribution of Gene Frequency as a Test of the Theory of
617 the Selective Neutrality of Polymorphisms. *Genetics* 74:175–195.

618 19. Nei M, Maruyama T (1975) Lewontin-Krakauer test for neutral genes. *Genetics* 80:395–395.

619 20. Hermisson J (2009) Who believes in whole-genome scans for selection? *Heredity* 103:283–284.

620 21. De Mita S, et al. (2013) Detecting selection along environmental gradients: analysis of eight
621 methods and their effectiveness for outbreeding and selfing populations. *Mol Ecol* 22:1383–1399.

622 22. Vitalis R, Gautier M, Dawson KJ, Beaumont MA (2014) Detecting and Measuring Selection
623 from Gene Frequency Data. *Genetics* 196:799–817.

624 23. Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral
625 parameterization on the performance of F_{ST} outlier tests. *Mol Ecol* 23:2178–2192.

626 24. Fraïsse C, Roux C, Welch JJ, Bierne N (2014) Gene-Flow in a Mosaic Hybrid Zone: Is Local
627 Introgression Adaptive? *Genetics* 197:939–951.

628 25. Le Moan A, Gagnaire P-A, Bonhomme F (2016) Parallel genetic divergence among coastal–
629 marine ecotype pairs of European anchovy explained by differential introgression after secondary
630 contact. *Mol Ecol* 25:3187–3202.

631 26. Rougemont Q, et al. (2017) Inferring the demographic history underlying parallel genomic
632 divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Mol Ecol* 26(1):142–162.

633 27. Takuno S, et al. (2015) Independent Molecular Basis of Convergent Highland Adaptation in
634 Maize. *Genetics* 200(4):1297–1312.

635 28. Bodénès C, Chancerel E, Ehrenmann F, Kremer A, Plomion C (2016) High-density linkage
636 mapping and distribution of segregation distortion regions in the oak genome. *DNA Res* 23:115–
637 124.

638 29. Beaumont, MA, Nichols RA. (1996) Evaluating loci for use in the genetic analysis of
639 population structure. *Proc R Soc Lond B Biol Sci* 263:1619–1626.

640 30. Albrechtsen A, Nielsen FC, Nielsen R (2010) Ascertainment Biases in SNP Chips Affect
641 Measures of Population Divergence. *Mol Biol Evol* 27:2534–2547.

642 31. Hipp AL, et al. (2018) Sympatric parallel diversification of major oak clades in the Americas
643 and the origins of Mexican species diversity. *New Phytol* 217:439–452.

644 32. Gregorius H-R, Degen B, König A (2007) Problems in the Analysis of Genetic Differentiation
645 Among Populations – a Case Study in Quercus robur. *Silvae Genet* 56:190.

646 33. Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet Res*
647 23:23–35.

648 34. Charlesworth B, Morgan MT, Charlesworth D (1993) The Effect of Deleterious Mutations on
649 Neutral Molecular Variation. *Genetics*:1289–1303.

650 35. Tine M, et al. (2014) European sea bass genome and its variation provide insights into
651 adaptation to euryhalinity and speciation. *Nat Commun* 5:5770.

652 36. Renaut S, et al. (2013) Genomic islands of divergence are not affected by geography of
653 speciation in sunflowers. *Nat Com* 4:1827.

654 37. Fonti P, Heller O, Cherubini P, Rigling A, Arend M (2013) Wood anatomical responses of oak
655 saplings exposed to air warming and soil drought. *Plant Biol* 15:210–219.

656 38. Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C (2002) Multiple Roles of *Arabidopsis*
657 *VRN1* in Vernalization and Flowering Time Control. *Science* 297:243.

658 39. Truffaut L, et al. (2017) Fine-scale species distribution changes in a mixed oak stand over two
659 successive generations. *New Phytol* 215:126–139.

660 40. Desprez-Loustau M-L, Feau N, Mougou-Hamdane A, Dutech C (2011) Interspecific and
661 intraspecific diversity in oak powdery mildews in Europe: coevolution history and adaptation to
662 their hosts. *Mycoscience* 52(3):165–173.

663 41. Wu D, et al. (2016) The HvNramp5 Transporter Mediates Uptake of Cadmium and
664 Manganese, But Not Iron. *Plant Physiol* 172:1899–1910.

665 42. Irish VF, Sussex IM (1990) Function of the apetala-1 gene during *Arabidopsis* floral
666 development. *Plant Cell* 2:741–753.

667 43. Higgins JA, Bailey PC, Laurie DA (2010) Comparative Genomics of Flowering Time Pathways
668 Using *Brachypodium distachyon* as a Model for the Temperate Grasses. *PLoS One* 5:e10065.

669 44. Babiychuk E, et al. (2008) Allelic mutant series reveal distinct functions for *Arabidopsis*
670 cycloartenol synthase 1 in cell viability and plastid biogenesis. *Proc Natl Acad Sci USA* 105:3163–
671 3168.

672 45. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient
673 alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.

674 46. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping
675 and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–
676 2993.

677 47. Kofler R, Pandey RV, Schlötterer C (2011) PoPoolation2: identifying differentiation between
678 populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27:3435–3436.

679 48. Roux C, et al. (2016) Shedding Light on the Grey Zone of Speciation along a Continuum of
680 Genomic Divergence. *PLoS Biol* 14:e2000234.

681 49. Ross-Ibarra J, et al. (2008) Patterns of Polymorphism and Demographic History in Natural
682 Populations of *Arabidopsis lyrata*. *PLoS One* 3:e2411.

683 50. Hudson RR (2002) Generating samples under a Wright–Fisher neutral model of genetic
684 variation. *Bioinformatics* 18:337–338

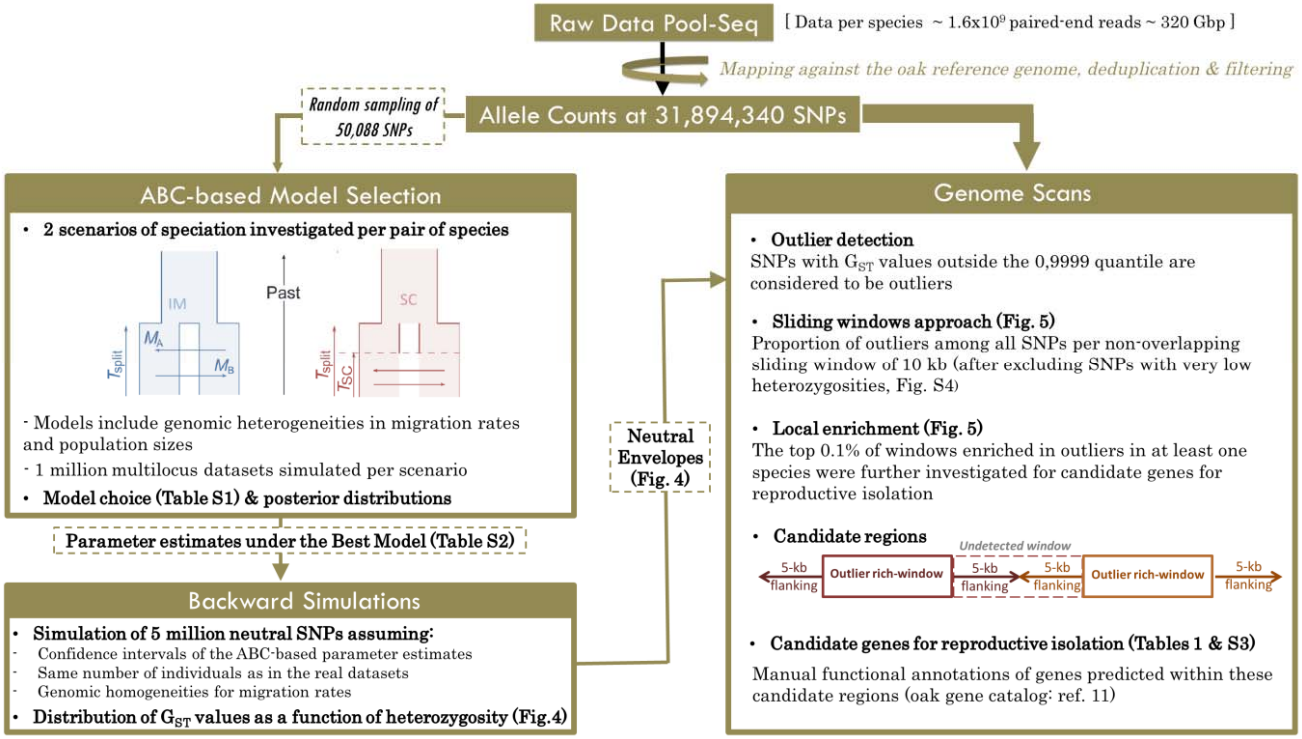


Figure 1: Workflow used to identify genes contributing to reproductive isolation between four European white oak species. A subset of 50 thousand of the called SNPs was selected at random and used for model selection under an ABC framework and the generation of parameter estimates under the best model. Large neutral datasets were then simulated to create null envelopes for the identification of SNPs displaying significant departure from expectations under neutrality. We searched for candidate genes in regions enriched in outliers.

691

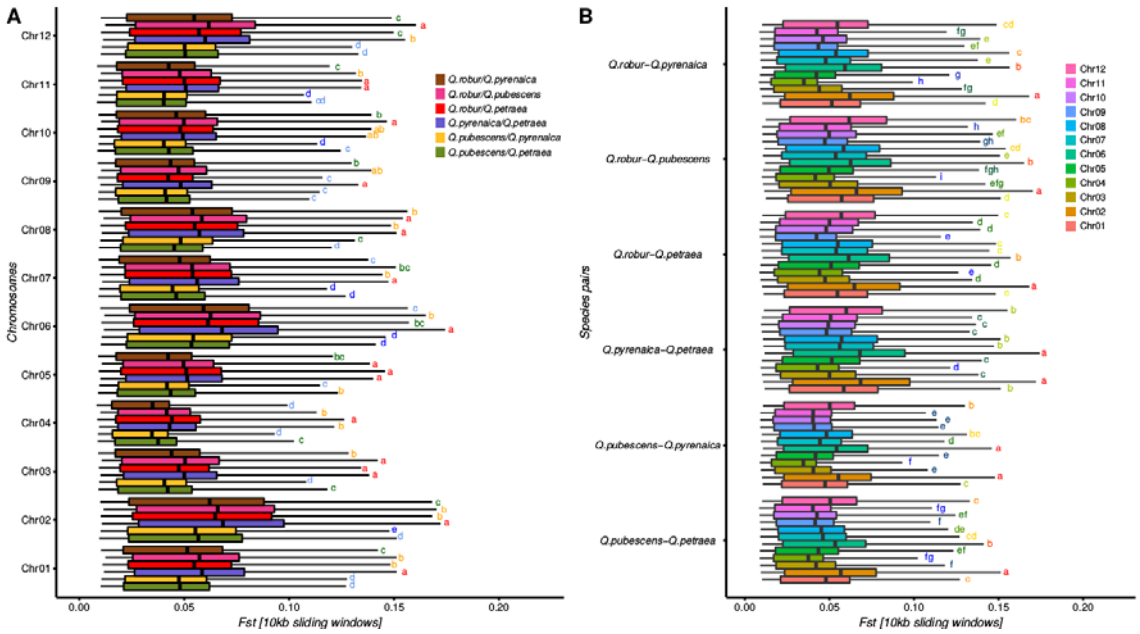
692

693

694

695

696



697

698 *Fig. 2: Boxplots of pairwise F_{ST} values per chromosome and per pair of species. F_{ST}*

699 *values were calculated from SNP allele frequencies within non-overlapping sliding*

700 *windows of 10 kb. Tukey's honestly significant difference (HSD) criterion at a*

701 *significance level of α=0.05 is reported for analyses between pairs of species for each*

702 *chromosome (A) and between chromosomes for each pair of species (B).*

703

704

705

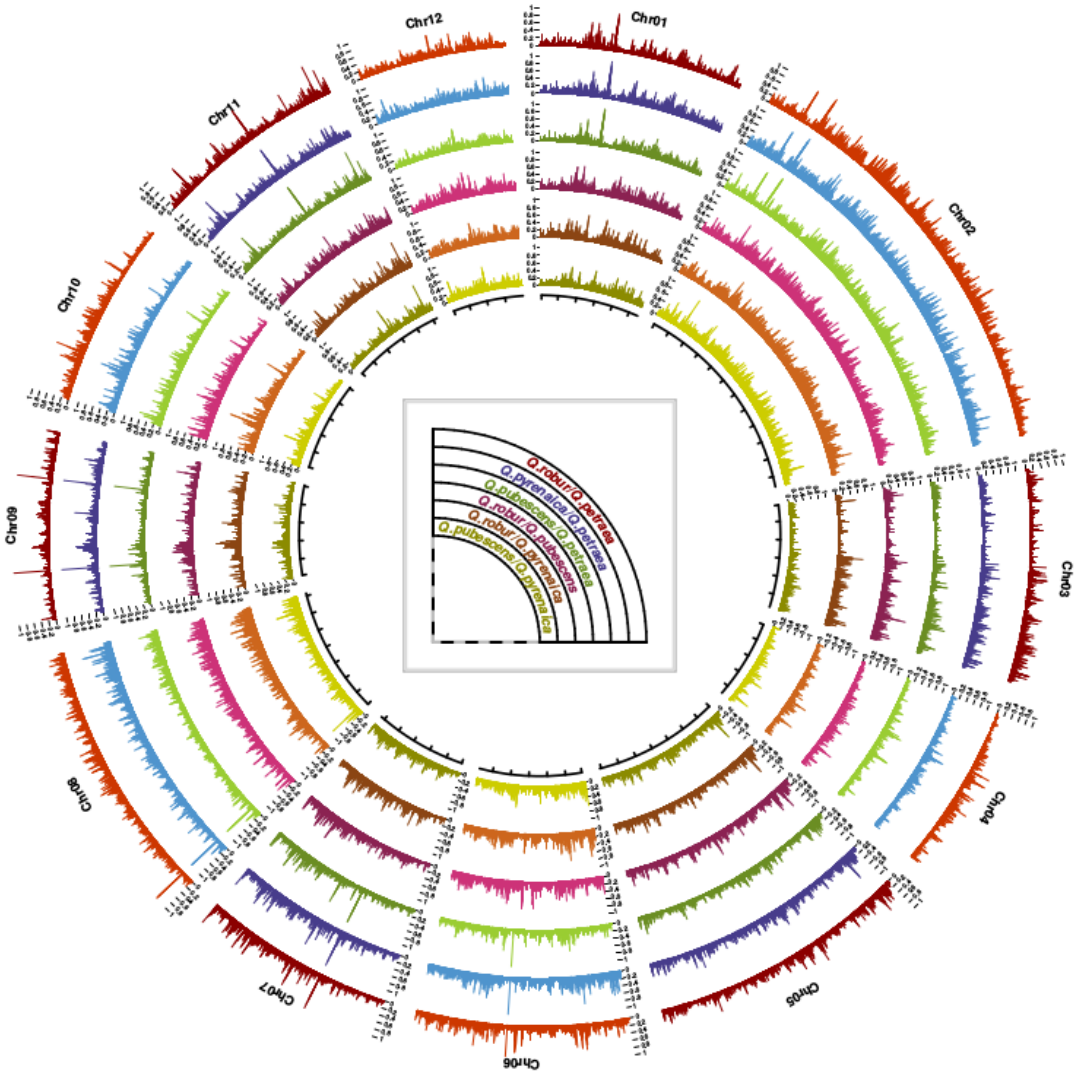


Fig. 3: Circular representation of F_{ST} values along the 12 oak chromosomes. F_{ST} was calculated from allele frequencies, using non-overlapping 10 kb sliding windows (detailed patterns are accessible from: https://github.com/ThibaultLeroyFr/GenomeScansByABC/).

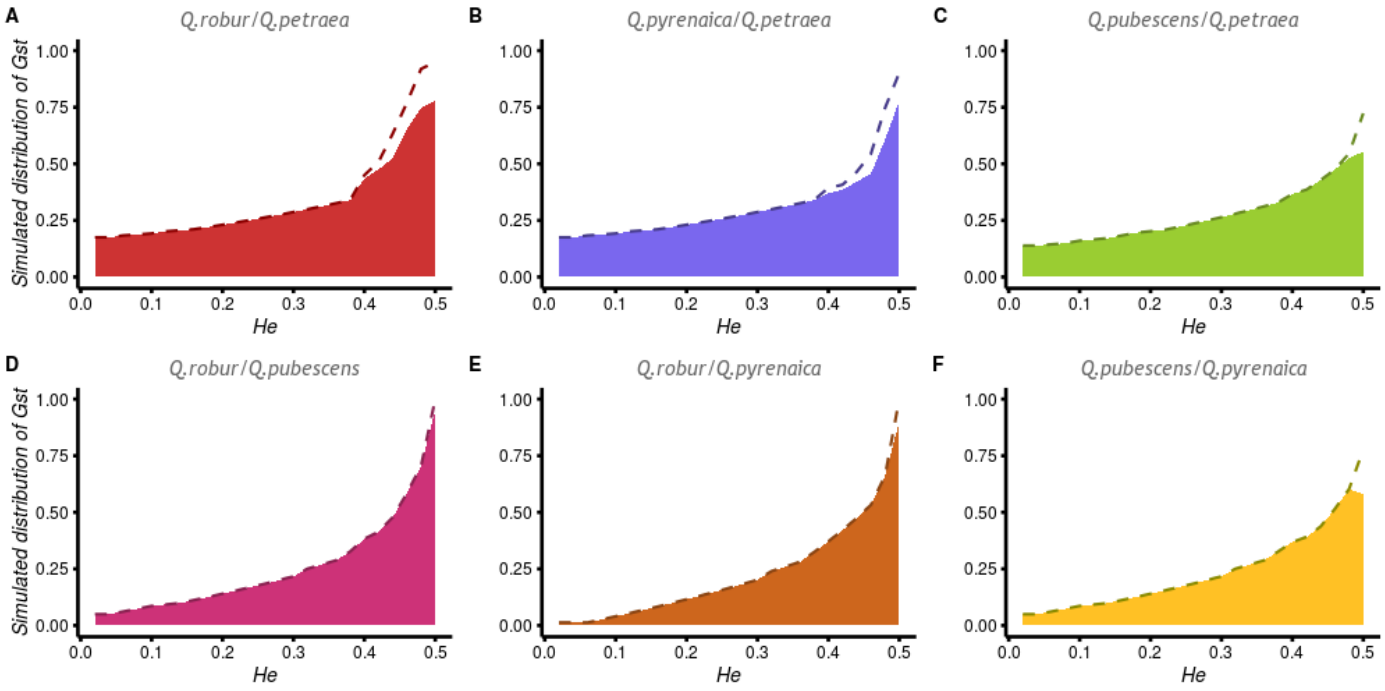


Fig. 4: Empirical null distributions of G_{ST} *values resulting from the backward simulation of recent secondary contact for Q. robur/Q. petraea (A), Q. pyrenaica/Q. petraea (B), Q. pubescens/Q. petraea (C), Q. robur/Q. pyrenaica (D), Q. robur/Q.p yrenaica (E) and Q. pubescens/Q. pyrenaica (F). For each pair of species, 5 million independent SNPs were simulated under the recent SC demographic model, assuming homogeneity of gene flow (i.e. no barriers) and the estimated 95 confidence intervals for estimates of the timing of recent SC (Table S2). Loci above the 0.9999 quantile (outside the colored area) were considered to be outliers. Maximum* G_{ST} *values are also reported (dashed lines).*

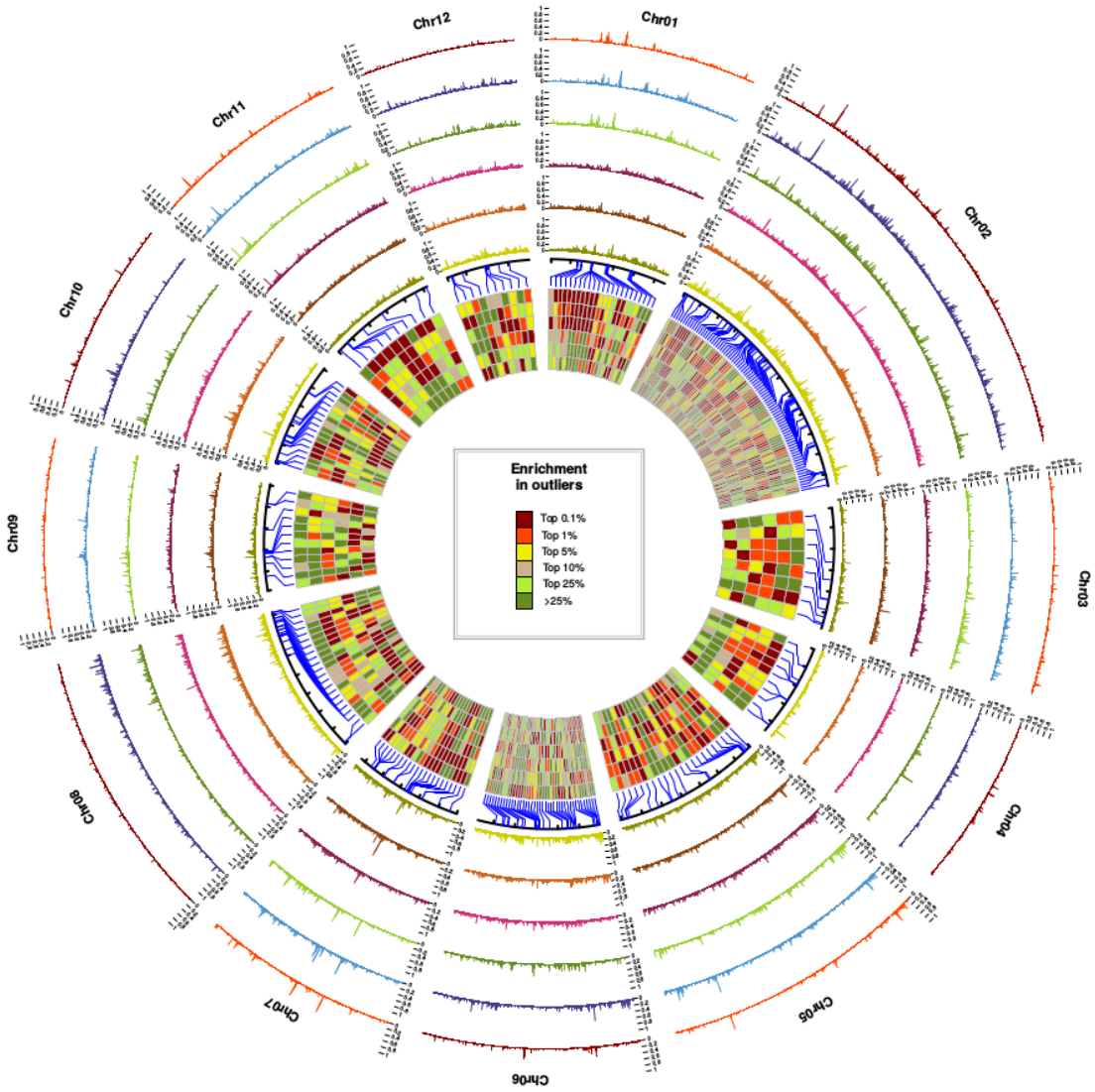


Fig. 5: *Local density in outliers per non-overlapping 10 kb sliding window. From outside to inside, the species pairs are Q. robur/Q. petraea, Q. pyrenaica/Q. petraea, Q. pubescens/Q. petraea, Q. robur/Q. pubescens, Q. robur/Q. pyrenaica and Q. pubescens/Q. pyrenaica. Detailed patterns are accessible from: https://github.com/ThibaultLeroyFr/GenomeScansByABC/. Each rectangle in the inner circle represents the level of enrichment in outliers with He ≤ 0.2 for each pair of species at a given position in the genome, assuming the same order of pairs. These rectangles correspond to the 281 most outlier-enriched windows found in at least one of the six pairs (top 0.1%).*

730 Table 1: Genomic positions, gene names, annotations for 32 candidate genes. All other annotations are available in Table S3 (see FileS1 for details, including

731 references). Species patterns were determined from the analysis of the most outlier-enriched windows between pairs, as detailed in Table S4.

chr.	*positions (regions)*	*Candidate gene*	*Gene name*	*Gene functions*	*#Paralogs*	*Pattern*
Intrinsic barriers						
Flowering						
Chr08	62493250-62513250	Qrob_P0412860.2	Two-component response regulator-like PRR73	photoperiodic flowering response, circadian clock	1	Complex
Chr08	62673250-62693250	Qrob_P0749650.2	Floral homeotic protein APETALA 2 | Transcr. factor RAP2-7	Delay transition to flowering time, biotic/abiotic stresses	0	Complex
Chr07	44592072-44632072	Qrob_P0088630.2	Transcr. activator DEMETER (DME) / protein ROS1-like	Transcriptional activator required for floral development	1	Complex
Pollen development (Cycloartenol synthases)						
Chr06	28660694-28680694	Qrob_P0684330.2	Cycloartenol synthase 2	Sterol or triterpenoid synthesis, pollen development	13	Complex
Chr06	28690694-28710694	Qrob_P0684400.2	Cycloartenol synthase 2	Sterol or triterpenoid synthesis, pollen development	0	Shared
Chr06	28770694-28810694	Qrob_P0684360.2	Cycloartenol synthase 2	Sterol or triterpenoid synthesis, pollen development	0	Shared
Pollen recognition & seed germination (G-type lectin S-receptor-like Serine/threonine kinase genes)						
Chr03	26118565-26138565	Qrob_P0538230.2	G-type lectin S-receptor-like STK At2g19130	Putatively involved in recognition of pollen	0	Complex
Chr05	4001785-4021785	Qrob_P0641650.2	G-type lectin S-receptor-like STK At1g11300	Putatively involved in recognition of pollen	190	Complex
Chr06	1453464-1473464	Qrob_P0430150.2	G-type lectin S-receptor-like STK LECRK	Regulates expression of immunity genes & seed germination	36	**Q. pyrenaica-specific**
Chr06	1513464-1563464	Qrob_P0430190.2	G-type lectin S-receptor-like STK LECRK	Regulates expression of immunity genes & seed germination	36	**Q. pyrenaica-specific**
Chr08	1749735-1769735	Qrob_P0138480.2	G-type lectin S-receptor-like STK LECRK	Regulates expression of immunity genes & seed germination	36	**Q. pyrenaica-specific**
Embryo development & organogenesis						
Chr12	30514806-30534806	Qrob_P0436820.2	Transcriptional corepressor LEUNIG-like isoform	Leaf, flower (gynoecium) and embryo development	5	**All except. Q. pub/Q.py**
Chr02	31611918-31651918	Qrob_P0297580.2	CHD3-type chromatin-remodeling factor PICKLE	Repressor of LEC1, activator of embryo development	0	Complex
Chr06	43006752-43026752	Qrob_P0309300.2	Probable N-acetyltransferase HLS1-like	Auxin-responsive gene expression, shoot organogenesis	1	**All except. Q. rob/Q.pe**
Chr08	49805801-49825801	Qrob_P0248780.2	Receptor-like protein 12 (RLP12)	Meristem maintenance control, organogenesis	0	Complex
Photoreceptor & UV-B tolerance						
Chr02	37788250-37808250	Qrob_P0338870.2	Ultraviolet-B receptor UVR8	Photoreceptor/response to UV, circadian clock, stomata	2	Shared
Chr03	34667252-34687252	Qrob_P0500290.2	DNA mismatch repair protein MSH2	UV-B-induced DNA damage response pathway	0	Shared
Chr06	11400887-11420887	Qrob_P0577750.2	Transcription factor MYB12	Positive regulator of flavonoid biosynthesis, UV-B tolerance	0	**All except. Q. rob/Q.pe**
Ecological barriers - abiotic stresses						
Nramp metal transporters						
Chr09	7739510-7759510	Qrob_P0191830.2	Metal transporter Nramp5	Manganese and cadmium uptake	8	**Q. pubescens-specific**
Chr06	38764748-38784748	Qrob_P0097150.2	Metal transporter Nramp6	Involved in iron ion homeostasis	1	**All except. Q. rob/Q.pe**
Dehydration/lateral root growth		-				

Chr08	68418235-68443235	Qrob_P0457680.2	Protein DEHYDRATION-INDUCED 19-like	Stress-induced sensor, interacting with CPK11 & S-Rnase	0	**All except. Q. rob/Q.pe**
Chr02	27061274-27081274	Qrob_P0304800.2	Transcription factor WER	Controls cell fate specification, e.g. hairy roots or stomata	0	**All except. Q. rob/Q.pe**
Chr02	28240136-28260136	Qrob_P0299670.2	Root Primordium Defective 1 (RPD1)	Lateral root morphogenesis; active cell proligeration	0	Complex
Chr02	36392234-36442234	Qrob_P0422470.2	Alkaline/neutral invertase CINV2	Regulator of root growth, sucrose catabolism	7	Shared
Chr09	19506093-19526093	Qrob_P0418880.2	1-aminocyclopropane-1-carboxylate synthase	Ethylene biosynthetic process, lateral root formation	4	**All except. Q. rob/Q.pe**
Chr04	19182244-19222244	Qrob_P0652510.2	Phosphatidylinositol-3-phosphatase myotubularin-1 (or 2)	Role in soil-water-deficit stress	2	Shared
Chr02	96008743-96028743	Qrob_P0387640.2	Protein WVD2-like 6	Organ stockiness (periph. root cap, trichomes & leafs)	0	**All except. Q. rob/Q.pe**
Freezing/cold adaptation						
Chr09	18536093-18556093	Qrob_P0768740.2	Dehydration-responsive element-binding protein 1	Key role in freezing tolerance and cold acclimation	6	Complex
Chr02	112153196-112173196	Qrob_P0339800.2	B3 domain-containing transcription factor VRN1	Vernalization responsiveness, repressor of FLC	3	**All except. Q. pub/Q.py**
Ecological barriers - biotic stresses						
Pathogen resistance/mycorrhization						
Chr10	12310682-12340682	Qrob_P0070130.2	Transportin MOS14	Plant immunity (splicing of resistant genes)	0	**Q. pyrenaica-specific**
Chr01	26763631-26783631	Qrob_P0648170.2	Ubiquitin carboxyl-terminal hydrolase 12-like	Protein deubiquitination, regulator of disease resistance	3	Complex
Chr05	3053435-3073435	Qrob_P0622110.2	Nodulation-signaling pathway 1 (NSP1)	Nodulation & mycorrhization	0	Complex

732

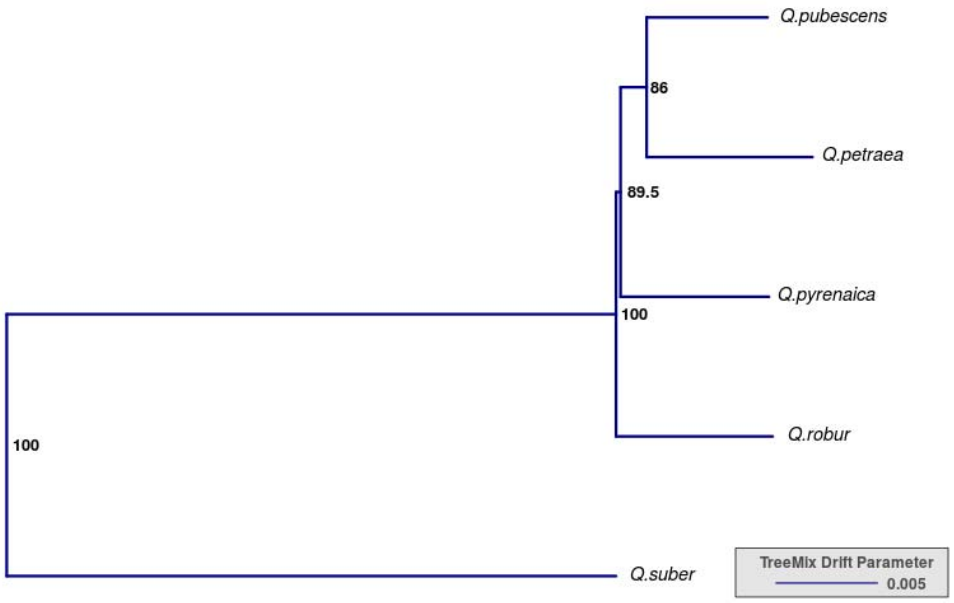


Fig. S1: Best maximum-likelihood phylogram for white oaks estimated under Treemix. *The analysis was performed on a dataset of 9,084,835 SNPs. The support of each node was determined from 200 replicate analyses, as described by Michael G. Harvey (http://www.mharvey.org/softwaredata/).*

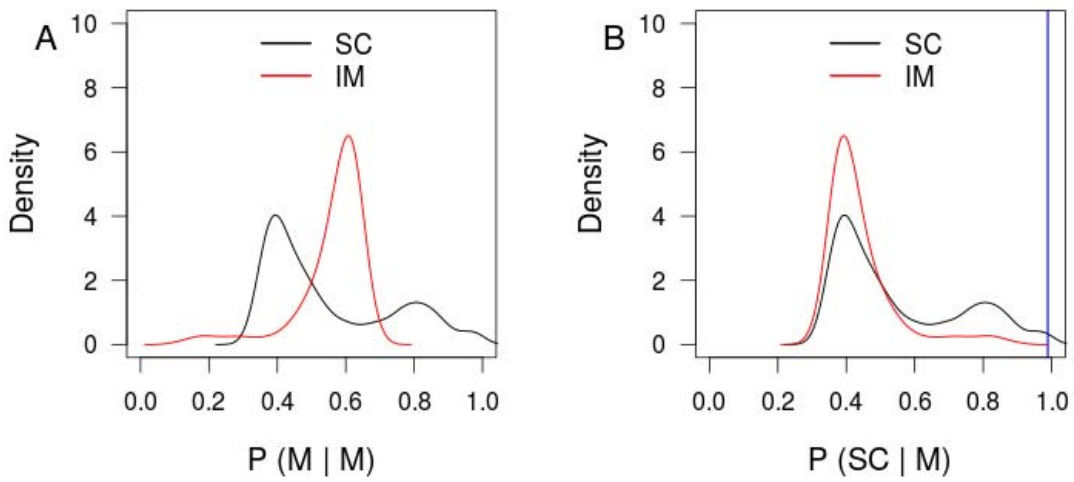


Fig. S2: ABC model checking. *Distributions of posterior probabilities across 1000 PODS that each simulated model is retrieved [P(M|M)] (A) and supporting secondary contact (SC) when isolation-with-migration (IM) and SC are simulated [P(SC|M)] (B). Posterior probability of SC for the* Q. robur/ Q. petraea *pair (real dataset) is shown in blue. The estimated robustness of the demographic inference in support of SC is 1.*

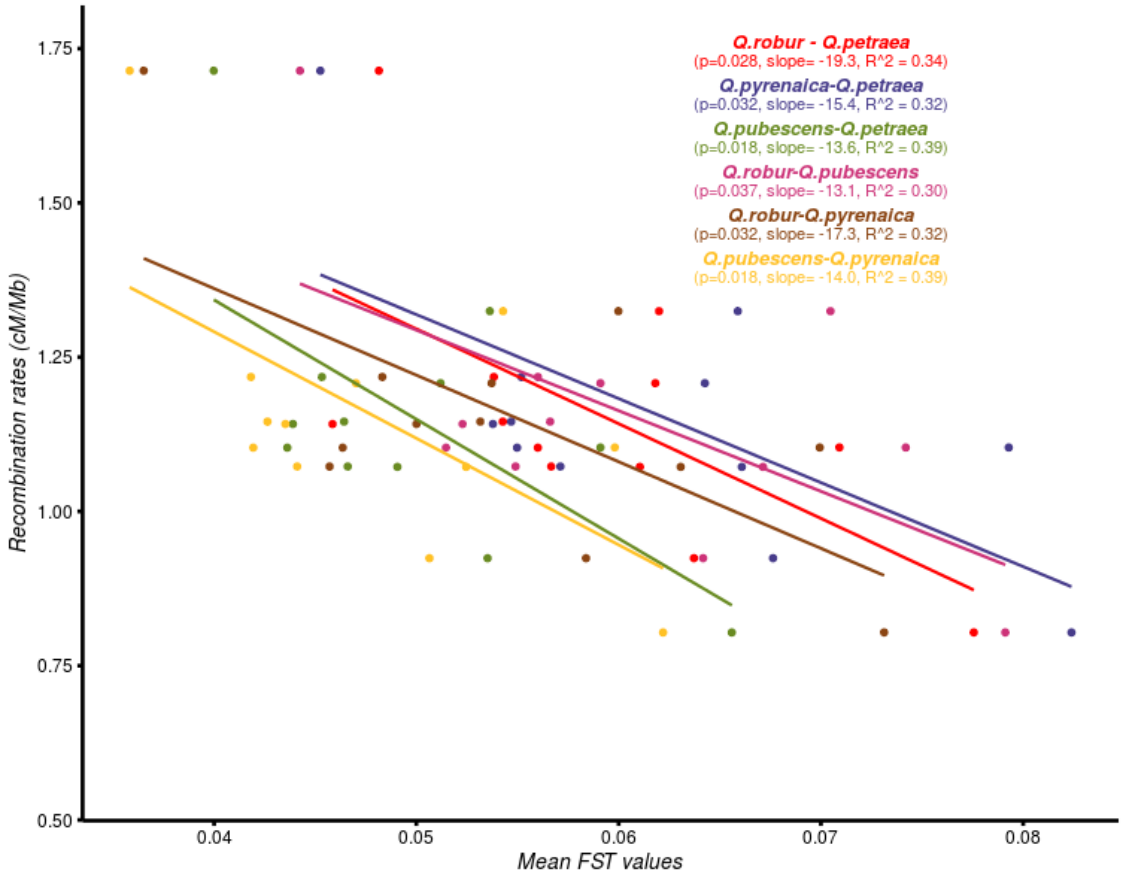


Fig. S3: Relationship between recombination rates (estimated at the whole chromosomal level) and F_{ST} values

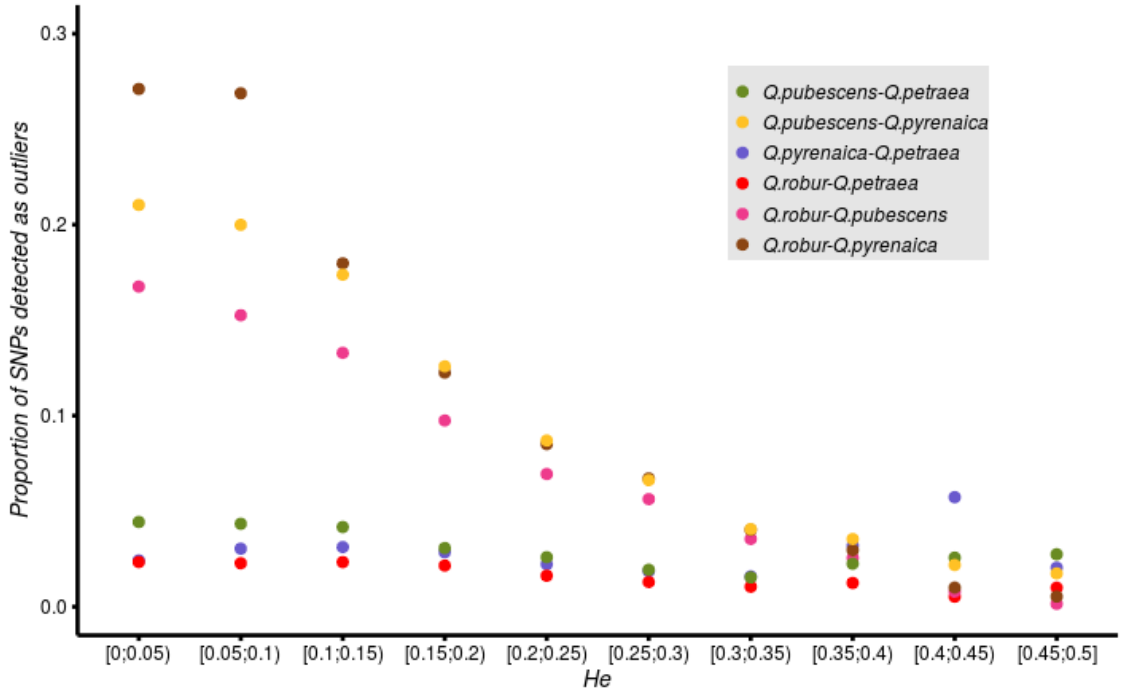
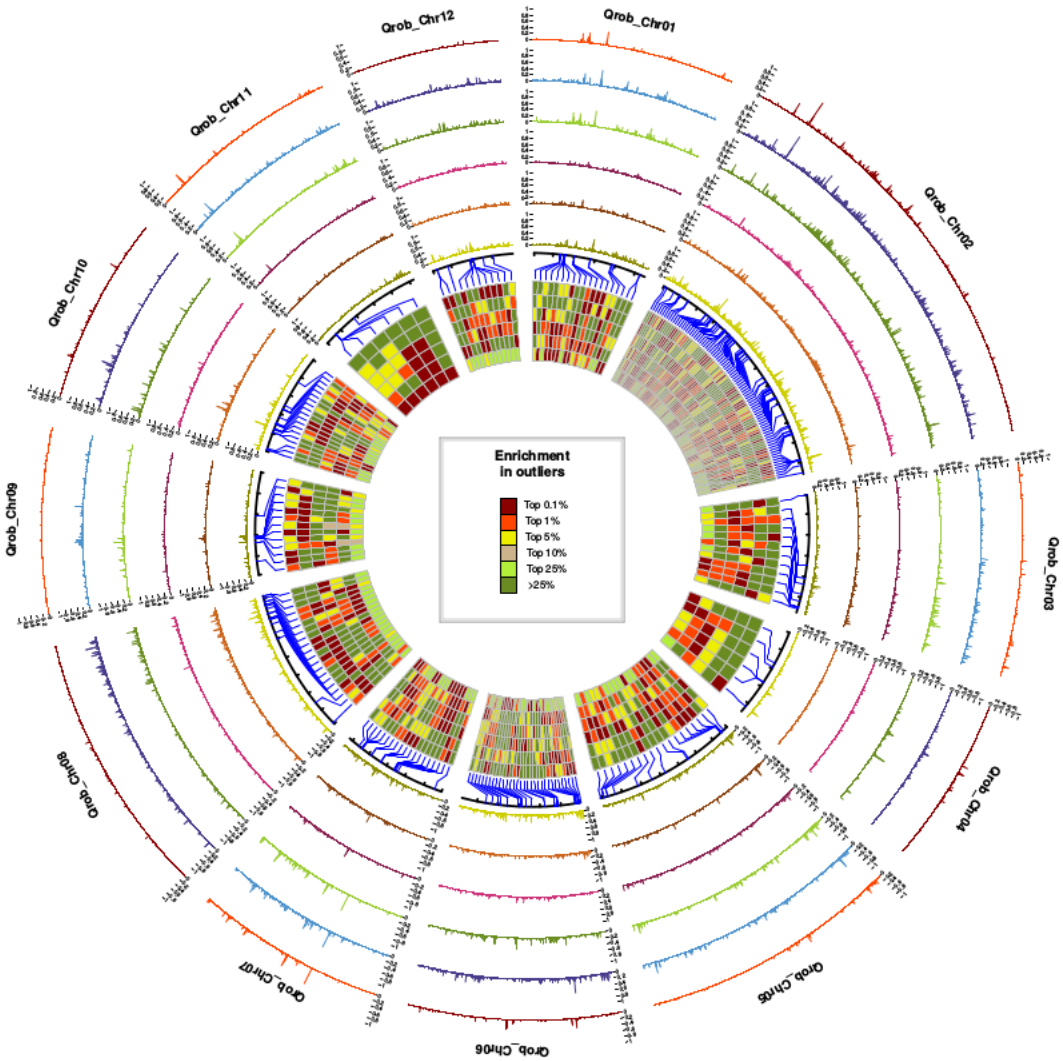


Fig. S4: Proportion of SNPs detected as outliers within a range of heterozygosity. *Outlier loci are loci with* G_{ST} *values above the 0.9999 quantile based on neutral simulations (colored area, Fig. 4).*



774

775 ***Fig. S5:*** *Local density in outliers per non-overlapping 10 kbsliding window. From outside*
776 *to inside, the species pairs are Q. robur/Q. petraea, Q. pyrenaica/Q. petraea, Q.*
777 *pubescens/Q. petraea, Q. robur/Q. pubescens, Q. robur/Q. pyrenaica and Q. pubescens/Q.*
778 *pyrenaica. Each rectangle in the inner circle represents the level of enrichment in*
779 *outliers with He ≤ 0.3 for each pair of species at a given position in the genome,*
780 *assuming the same order of pairs. These regions correspond to the 281 most-enriched*
781 *windows found in at least one of the six pairs (top 0.1%).*

782

783

784 Table S1: Relative posterior probabilities of the IM (isolation-with-migration) and SC (secondary contact) scenarios averaged over 20 ABC analyses. The best

785 supported model is shown in bold. Standard deviations are indicated in brackets.

	IM	**SC**
Q. robur – Q. petraea	0.01117 (±0.01089)	**0.98883** (±0.01089)
Q. robur – Q. pyrenaica	0.00751 (±0.01152)	**0.99249** (±0.01152)
Q. robur – Q. pubescens	0.01281 (±0.01342)	**0.98719** (±0.01342)
Q. pubescens – Q. petraea	0.01451 (±0.02231)	**0.98549** (±0.02231)
Q. pubescens – Q. pyrenaica	0.00913 (±0.01152)	**0.99087** (±0.01152)
Q. pyrenaica – Q. petraea	0.00622 (±0.00697)	**0.99378 (±0.00697)**

786

787

788

789

790

Table S2: Median and 95% confidence intervals of parameter estimates for both the SC (first round; italics, lower row) and the recent SC scenario of speciation (second round; bold, upper row). Npop1, Npop2 and Nanc correspond to the posteriors of the effective population sizes in species 1, species 2 and their ancestor, respectively. Mpop1 and Mpop2 correspond to the effective migration rates (Mpop1=4.Npop1.$m_{pop2>pop1}$ and Mpop2=4.Npop2.$m_{pop1>pop2}$). T_{SPLIT} and T_{SC} correspond to the unscaled times of divergence and secondary contact, respectively).

	Npop1	Npop2	Nanc	Mpop1	Mpop2	Tsplit	Tsc	Tsc/Tsplit
Q. robur – Q. petraea	**4.161×10^{6}**	**5.862×10^{6}**	**5.505×10^{6}**	**48.422**	**45.020**	**61.462**	**79.186×10^{-3}**	**1.487×10^{-3}**
	[1.99-7.79] x 10^{6}	**[1.81-9.43] x 10^{6}**	**[0.89-8.64] x 10^{6}**	**[7.54-94.80]**	**[7.22-45.02]**	**[31.92-86.39]**	**[24.64-258.30] x 10^{-3}**	**[0.43-4.05] x 10^{-3}**
	5.391×10^{6}	*5.591×10^{6}*	*5.287×10^{6}*	*31.327*	*50.847*	*34.030*	*126.478×10^{-3}*	*4.363×10^{-3}*
	[1.07-9.40] x 10^{6}	*[1.38-9.52] x 10^{6}*	*[0.31-9.85] x 10^{6}*	*[1.45-89.87]*	*[23.39-78.54]*	*[9.40-73.33]*	*[29.17-404.21] x 10^{-3}*	*[0.98-11.97] x 10^{-3}*
Q. robur – Q. pyrenaica	**4.615×10^{6}**	**3.193×10^{6}**	**5.267×10^{6}**	**34.332**	**33.795**	**69.181**	**77.638×10^{-3}**	**1.197×10^{-3}**
	[1.59-8.50} x 10^{6}	**[0.60-9.02} x 10^{6}**	**[0.53-9.40} x 10^{6}**	**[4.62-91.17}**	**[2.82-95.66}**	**[35.59-92.19}**	**[21.45-316.52} x 10^{-3}**	**[0.31-4.95} x 10^{-3}**
	3.618×10^{6}	*7.398×10^{6}*	*3.024×10^{6}*	*41.395*	*30.673*	*34.529*	*65.995×10^{-3}*	*2.227×10^{-3}*
	[0.71-8.51]x 10^{6}	*[2.12-9.69]x 10^{6}*	*[0.60-8.23]x 10^{6}*	*[10.11-82.76]*	*[2.09-93.74]*	*[11.63-68.29]*	*[19.70-188.15] x 10-*	*[0.62-5.47] x 10-*

							3	
Q. robur – Q. pubescens	**3.479 x 10^6** **[1.05-8.89} x 10^6**	**6.448 x 10^6** **[1.67-9.59} x 10^6**	**3.222 x 10^6** **[0.56-9.38} x 10^6**	**79.897** **[7.25-95.94}**	**77.358** **[20.87-97.14}**	**53.156** **[14.45-83.61}**	**11.823 x 10^{-3}** **[5.07-22.12} x 10^{-3}**	**0.245 x 10^{-3}** **[0.09-0.72} x 10^{-3}**
	3.027 x 10^6 *[0.97-9.37]x 10^6*	*5.532 x 10^6* *[1.17-9.41]x 10^6*	*4.719 x 10^6* *[0.57-9.29] x 10^6*	*32.648* *[1.78-91.62]*	*35.432* *[3.08-92.91]*	*23.619* *[3.07-92.91]*	*93.266 x 10^{-3}* *[11.81-517.76] x 10-3*	*4.682 x 10^{-3}* *[0.65-16.66 x] 10-3*
Q. pubescens – Q. petraea	**5.959 x 10^6** **[1.49-9.77]x 10^6**	**5.565 x 10^6** **[1.23-9.66]x 10^6**	**6.786 x 10^6** **[1.51-9.56] x 10^6**	**78.811** **[23.54-98.32]**	**27.534** **[5.62-73.63]**	**46.641** **[19.63-78.39]**	**98.043 x 10^{-3}** **[36.68-284.71] x 10-3**	**2.176 x 10^{-3}** **[0.77-6.24] x 10-3**
	6.185 x 10^6 *[1.42-9.62]x 10^6*	*5.951 x 10^6* *[1.77-9.44]x 10^6*	*2.910 x 10^6* *[0.27-8.61] x 10^6*	*37.906* *[4.30-87.71]*	*40.956* *[3.50-94.01]*	*36.234* *[9.01-73.38]*	*96.771 x 10^{-3}* *[2.48-313.90] x 10-3*	*3.061 x 10^{-3}* *[0.77-9.88] x 10-3*
Q. pubescens – Q. pyrenaica	**5.819 x 10^6** **[1.89-9.57]x 10^6**	**5.408 x 10^6** **[1.67-9.24]x 10^6**	**4.931 x 10^6** **[0.95-9.45] x 10^6**	**45.753** **[7.30-90.25]**	**96.076** **[73.56-99.74]**	**46.869** **[18.40-78.34]**	**37.309 x 10^{-3}** **[16.10-91.09] x 10-3**	**0.865 x 10^{-3}** **[0.32-2.16] x 10-3**
	5.417 x 10^6 *[1.49-9.48]x 10^6*	*5.666 x 10^6* *[1.16-9.50]x 10^6*	*4.558 x 10^6* *[0.42-9.23] x 10^6*	*45.895* *[8.61-91.91]*	*30.793* *[2.66-92.66]*	*28.160* *[7.85-61.71]*	*103.036 x 10^{-3}* *[10.69-437.49] x 10-*	*3.952 x 10^{-3}* *[0.60-15.55] x 10-3*

							3	
Q. pyrenaica – Q. petraea	**7.169 x 10^{6}**	**8.890 x 10^{6}**	**5.073 x 10^{6}**	**97.785**	**41.107**	**66.037**	**22.327 x 10^{-3}**	**0.383 x 10^{-3}**
	[3.13-9.68]x 10^{6}	**[6.21-9.89]x 10^{6}**	**[0.82-9.89] x 10^{6}**	**[91.70-99.58]**	**[7.02-88.93]**	**[37.93-86.20]**	**[6.12-97.76] x 10-3**	**[0.10-1.14] x 10-3**
	5.604 x 10^{6}	*5.782 x 10^{6}*	*5.23 x4 10^{6}*	*69.106*	*80.362*	*30.015*	*56.592 x 10^{-3}*	*2.116 x 10^{-3}*
	[1.47-9.46]x 10^{6}	*[1.45-9.65]x 10^{6}*	*[0.50-9.62] x 10^{6}*	*[7.02-98.44]*	*[15.52-99.00]*	*[8.67-68.66]*	*[9.62-28.09] x 10-3*	*[0.39-8.80] x 10-3*

Table S3: Genomic positions, gene names and annotations for the best candidate of each of the 215 detected regions. References are shown in File S1.

Chr.	Region-Pos-Start	Chr-Pos-End	Homol. CpDNA	#Genes	Best candidate	Other genes	Gene name Best Candidate
Chr01	13033024	13053024	N	1	Qrob_P0300140.2	--	LURP-one-related 17
Chr01	13523024	13543024	N	2	Qrob_P0300370.2	Qrob_P0300380.2	SEC5A-like
Chr01	14906647	14921647	N	1	Qrob_P0442350.2	--	cytochrome P450 CYP736A12-like
Chr01	15281647	15301647	N	1	Qrob_P0503340.2	--	MED25
Chr01	15341647	15361647	N	1	Qrob_P0503330.2	--	uncharacterized protein LOC100242396 iso. X2
Chr01	19962849	20002849	**Y**	7	*cpDNA*	Qrob_P0270310.2 to Qrob_P0270370.2	--
Chr01	20032849	20092849	**Y**	18	*cpDNA*	Qrob_P0269980.2 to Qrob_P0270190.2	--
Chr01	20172849	20212849	N	1	Qrob_P0738230.2	--	uncharacterized protein LOC107407642
Chr01	26763631	26783631	N	1	Qrob_P0648170.2	--	ubiquitin carboxyl-terminal hydrolase 12-like
Chr01	27363730	27393730	N	1	Qrob_P0605010.2	--	aluminum-activated malate transporter 2-like
Chr01	27829550	27859550	N	1	Qrob_P0595490.2	--	protein FAF-like
Chr01	38136926	38156926	N	0	--	--	--
Chr01	38566926	38586926	N	1	Qrob_P0726670.2	--	kinesin light chain 3
Chr01	39066926	39086926	N	1	Qrob_P0701720.2	--	oxidoreductase At4g09670-like
Chr01	39456926	39476926	N	0	--	--	--
Chr02	1645714	1675714	N	0	--	--	--
Chr02	2481169	2501169	N	1	Qrob_P0029210.2	--	GDSL esterase/lipase At1g28610-like
Chr02	2721169	2741169	N	0	--	--	--
Chr02	6582759	6602759	N	0	--	--	--
Chr02	9512582	9532582	**Y**	4	*cpDNA*	Qrob_P0433280.2 to Qrob_P0433310.2	--
Chr02	14453089	14483089	N	6	*mtDNA gene(s)*	Qrob_P0764110.2 to Qrob_P0764160.2	NADH dehydrogenase (mitochondrion)
Chr02	14533089	14563089	**Y**	2	*cpDNA*	Qrob_P0028300.2 to Qrob_P0028320.2	--
Chr02	15893089	15913089	N	0	--	--	--
Chr02	22439402	22459402	N	2	Qrob_P0204840.2	Qrob_P0204860.2	Lysine histidine transporter-like 8
Chr02	25082336	25107336	N	1	*mtDNA gene(s)*	Qrob_P0359400.2	Mitochondrial import receptor subunit TOM5
Chr02	27061274	27081274	N	1	Qrob_P0304800.2	--	Transcription factor WER
Chr02	27980136	28000136	N	2	Qrob_P0299860.2	Qrob_P0299850.2	Glutamate receptor 2.9
Chr02	28240136	28260136	N	2	Qrob_P0299670.2	Qrob_P0299680.2	root primordium defective 1 (RPD1)
Chr02	28300136	28320136	N	0	--	--	--
Chr02	28340136	28360136	N	0	--	--	--
Chr02	31611918	31651918	N	2	Qrob_P0297580.2	Qrob_P0297570.2	CHD3-type chromatin-remodeling factor PICKLE
Chr02	32892045	32912045	N	0	--	--	--
Chr02	36392234	36442234	N	3	Qrob_P0422470.2	Qrob_P0422480.2; Qrob_P0422490.2	Alkaline/neutral invertase CINV2
Chr02	37091551	37121551	N	1	Qrob_P0685110.2	--	unknown
Chr02	37281551	37301551	N	0	--	--	--
Chr02	37788250	37808250	N	1	Qrob_P0338870.2	--	Ultraviolet-B receptor UVR8
Chr02	39068973	39088973	N	1	Qrob_P0228250.2		threonylcarbamoyladenosine tRNA methylthiotransferase
Chr02	39118973	39138973	N	0	--	--	--
Chr02	39938973	39958973	N	0	--	--	--
Chr02	41110823	41140823	N	0	--	--	--
Chr02	41270823	41290823	N	0	--	--	--
Chr02	41330823	41350823	N	0	--	--	--
Chr02	41880823	41900823	N	0	--	--	--
Chr02	42838387	42868387	N	1	Qrob_P0685680.2	--	uncharacterized protein LOC109005465

Chr02	44212076	44232076	N	1	Qrob_P0377200.2	--	uncharacterized protein LOC109010125
Chr02	46828931	46848931	N	2	Qrob_P0609970.2	Qrob_P0609980.2	global transcription factor group E8 (GTE8)
Chr02	48428931	48448931	N	2	Qrob_P0120600.2	Qrob_P0120590.2	Calcineurin B-like protein 7
Chr02	48608931	48628931	N	0	--	--	--
Chr02	49492919	49522919	N	0	--	--	--
Chr02	49612919	49632919	N	0	--	--	--
Chr02	51519012	51539012	N	0	--	--	--
Chr02	53139111	53159111	N	0	--	--	--
Chr02	53839111	53859111	N	0	--	--	--
Chr02	57904424	57924424	N	1	Qrob_P0343940.2	--	Receptor-like protein kinase HSL1
Chr02	68021459	68051459	N	1	Qrob_P0541680.2	--	Cytokinin hydroxylase
Chr02	68111459	68131459	N	0	--	--	--
Chr02	71082167	71102167	N	1	Qrob_P0459590.2	--	uncharacterized protein LOC109001885
Chr02	71864163	71894163	N	0	--	--	--
Chr02	72264163	72284163	N	1	Qrob_P0324500.2		nuclear transcription factor Y subunit A-3 (NFYA3)
Chr02	72424163	72474163	N	0	--	--	--
Chr02	73004163	73024163	N	1	Qrob_P0758700.2	--	glucan 1,3-alpha-glucosidase (PSL5)
Chr02	73044163	73074163	N	0	--	--	--
Chr02	73901260	73921260	N	1	Qrob_P0516630.2	--	Vinorine synthase
Chr02	76588861	76608861	N	1	Qrob_P0226470.2	--	Probable caffeoyl-CoA O-methyltransferase At4g26220
Chr02	77078861	77098861	N	1	Qrob_P0546070.2	--	Transportin MOS14
Chr02	77228861	77258861	N	0	--	--	--
Chr02	82229996	82249996	N	1	Qrob_P0298270.2	--	Putative glucose-6-phosphate 1-epimerase (AAPC)
Chr02	85660787	85680787	N	1	Qrob_P0365050.2	--	Zinc-finger homeodomain protein 9
Chr02	86250787	86270787	N	0	--	--	--
Chr02	94274060	94294060	N	0	--	--	--
Chr02	95548743	95568743	N	0	--	--	--
Chr02	96008743	96028743	N	2	Qrob_P0387640.2	Qrob_P0387630.2	protein WVD2-like 6
Chr02	96068743	96088743	N	0	--	--	--
Chr02	108643097	108663097	N	0	--	--	--
Chr02	108773097	108813097	N	0	--	--	--
Chr02	108913097	108943097	N	0	--	--	--
Chr02	112153196	112173196	N	2	Qrob_P0339800.2	Qrob_P0339790.2	B3 domain-containing transcription factor VRN1
Chr02	115017268	115037268	N	1	Qrob_P0110200.2	--	NAC domain-containing protein 35
Chr03	26118565	26138565	N	1	Qrob_P0538230.2	--	G-type lectin S-receptor-like serine/thr.-protein kinase At2g19130
Chr03	34667252	34687252	N	1	Qrob_P0500290.2	--	DNA mismatch repair protein MSH2
Chr03	43542295	43562295	N	1	Qrob_P0094300.2	--	Zinc finger CCCH domain-containing protein 51 (C3H51)
Chr03	43852295	43872295	N	1	Qrob_P0094440.2	--	Wall-associated receptor kinase-like 9 (or 23)
Chr03	48437182	48457182	N	1	Qrob_P0640360.2	--	Homeobox protein knotted-1-like 6
Chr03	51917862	51937862	N	0	--	--	--
Chr03	54841512	54861512	N	1	Qrob_P0202390.2	--	Cytochrome P450 CYP749A22
Chr03	55647135	55667135	N	1	Qrob_P0732610.2	--	unknown
Chr04	10306800	10326800	N	0	--	--	--
Chr04	17034485	17054485	N	0	--	--	--
Chr04	19182244	19222244	N	4	Qrob_P0652510.2	Qrob_P0652490.2 to Qrob_P0652520.2	phosphatidylinositol-3-phosphatase myotubularin-1 (or 2)
Chr04	27613262	27653262	N	3	Qrob_P0603460.2	Qrob_P0603450.2; Qrob_P0603480.2	uncharacterized protein LOC108988747
Chr04	41840838	41860838	N	1	Qrob_P0201440.2	--	ABC transporter G family member 15
Chr04	41890838	41910838	N	1	Qrob_P0201420.2	--	Proteasome subunit alpha type-1-A (ARS5 or PAF1)
Chr05	1619164	1639164	N	0	--	--	--

Chr05	3053435	3073435	N	1	Qrob_P0622110.2	--	nodulation-signaling pathway 1 (NSP1)
Chr05	4001785	4021785	N	1	Qrob_P0641650.2	--	G-type lectin S-receptor-like serine/thr.-protein kinase At1g11300
Chr05	4431785	4451785	N	0	--	--	--
Chr05	4631884	4661884	N	0	--	--	--
Chr05	4741884	4781884	N	1	Qrob_P0641900.2	--	Transcription factor WER
Chr05	4941884	4961884	N	1	Qrob_P0579770.2	--	WD repeat-containing protein 76
Chr05	24515628	24535628	N	1	Qrob_P0487290.2		N-terminal acetyltransferase A complex auxiliary subunit NAA15
Chr05	25145628	25165628	N	0	--	--	--
Chr05	37432464	37452464	N	0	--	--	--
Chr05	41298251	41318251	N	0	--	--	--
Chr05	44681390	44701390	N	1	Qrob_P0155740.2	--	Pentatricopeptide repeat-containing protein At5g16860
Chr05	45428867	45448867	N	3	Qrob_P0568000.2	Qrob_P0567980.2; Qrob_P0568020.2	Cationic amino acid transporter 1
Chr05	51849096	51869096	N	1	Qrob_P0358380.2	--	Probable WRKY transcription factor 14
Chr05	55666961	55686961	N	0	--	--	--
Chr05	65965694	65985694	N	2	Qrob_P0057770.2	Qrob_P0057780.2	L-type lectin-domain containing receptor kinase IX
Chr05	67075694	67135694	N	1	Qrob_P0058400.2	--	Anthocyanin regulatory C1 protein
Chr05	69018740	69038740	N	2	Qrob_P0540710.2	Qrob_P0540730.2	Aspartyl protease At5g10770 | aspartic protease in guard cell 1
Chr05	70624885	70644885	N	2	Qrob_P0457080.2	Qrob_P0457090.2	Beta-amyrin 28-oxidase
Chr06	375100	395100	N	0	--	--	--
Chr06	1183464	1203464	N	0	--	--	--
Chr06	1453464	1473464	N	1	Qrob_P0430150.2	--	G-type lectin S-receptor-like serine/threonine-protein kinase LECRK
Chr06	1513464	1563464	N	2	Qrob_P0430190.2	Qrob_P0430210.2	G-type lectin S-receptor-like serine/threonine-protein kinase LECRK
Chr06	2723464	2743464	N	1	Qrob_P0697210.2	--	Transcription factor MUTE
Chr06	3376543	3396543	N	0	--	--	--
Chr06	11400887	11420887	N	2	Qrob_P0577750.2	Qrob_P0577760.2	Transcription factor MYB12
Chr06	11910887	11930887	N	2	Qrob_P0631400.2	Qrob_P0631390.2	Nucleic acid-binding proteins superfamily isoform 1
Chr06	12930833	12950833	N	0	--	--	--
Chr06	13060833	13090833	N	3	Qrob_P0216660.2	Qrob_P0216650.2; Qrob_P0216670.2	DEAD-box ATP-dependent RNA helicase 40
Chr06	13850932	13870932	N	1	Qrob_P0036390.2	--	Phytosulfokine receptor 2
Chr06	15110932	15150932	N	2	Qrob_P0035650.2	Qrob_P0035660.2	Auxin-responsive protein IAA
Chr06	15816572	15831572	N	1	Qrob_P0091910.2	--	MATH domain-containing protein At5g43560
Chr06	17911572	17941572	N	1	Qrob_P0615970.2	--	unknown
Chr06	23826093	23846093	N	0	--	--	--
Chr06	26616505	26636505	N	3	Qrob_P0255820.2	Qrob_P0255830.2; Qrob_P0255840.2	multiple RNA-binding domain-containing protein 1
Chr06	28660694	28680694	N	1	Qrob_P0684330.2	--	Cycloartenol synthase 2
Chr06	28690694	28710694	N	2	Qrob_P0684400.2	Qrob_P0684390.2	Cycloartenol synthase 2
Chr06	28770694	28810694	N	1	Qrob_P0684360.2	--	Cycloartenol synthase 2
Chr06	29251119	29271119	N	1	Qrob_P0196720.2	--	Nucleolar complex protein 2
Chr06	30327075	30347075	N	0	--	--	--
Chr06	30497075	30517075	N	0	--	--	--
Chr06	30557075	30597075	N	0	--	--	--
Chr06	35573532	35593532	N	2	Qrob_P0200780.2	Qrob_P0200770.2	uncharacterized protein LOC108997646
Chr06	36671749	36691749	N	0	--	--	--
Chr06	38764748	38784748	N	1	Qrob_P0097150.2	--	Metal transporter Nramp6
Chr06	38874748	38894748	N	0	--	--	--
Chr06	43006752	43026752	N	1	Qrob_P0309300.2	--	Probable N-acetyltransferase HLS1-like
Chr06	43486851	43506851	N	1	Qrob_P0199740.2	--	transcription factor MYB44
Chr06	45124942	45144942	N	0	--	--	--
Chr06	45934942	45954942	N	1	Qrob_P0162820.2	--	Histone-lysine N-methyltransferase CLF

Chr06	49134942	49154942	N	0	--	--	--
Chr07	4437303	4457303	N	0	--	--	--
Chr07	10368555	10388555	N	0	--	--	--
Chr07	18388820	18418820	**Y**	6	*cpDNA*	Qrob_P0358910.2 to Qrob_P0358960.2	--
Chr07	24337285	24357285	N	1	Qrob_P0313140.2	--	Calcium-dependent protein kinase 1
Chr07	25403104	25433104	N	3	Qrob_P0222530.2	Qrob_P0222520.2; Qrob_P0222540.2	Ubiquitin carboxyl-terminal hydrolase 8
Chr07	26171992	26201992	N	0	--	--	--
Chr07	26231992	26251992	N	0	--	--	--
Chr07	26261992	26281992	N	1	Qrob_P0571810.2	--	UDP-glycosyltransferase 73C3
Chr07	27632353	27652353	N	1	Qrob_P0135450.2	--	NDR1/HIN1-like protein 10 / YLS9-like
Chr07	33563769	33593769	N	1	Qrob_P0692960.2	--	protein DETOXIFICATION 53
Chr07	39089942	39119942	N	1	Qrob_P0501800.2	--	Protein RRC1
Chr07	44392072	44412072	N	0	--	--	--
Chr07	44592072	44632072	N	2	Qrob_P0088630.2	Qrob_P0088620.2	Transcriptional activator DEMETER (DME) / protein ROS1-like
Chr07	44812072	44832072	N	1	Qrob_P0527880.2	--	Fructose-bisphosphate aldolase 6
Chr07	46455822	46475822	N	1	Qrob_P0407530.2	--	Phosphatidylinositol 4-phosphate 5-kinase 9
Chr08	1749735	1769735	N	2	Qrob_P0138480.2	Qrob_P0138490.2	G-type lectin S-receptor-like serine/threonine-protein kinase LECRK
Chr08	1769735	1789735	N	0	--	--	--
Chr08	49805801	49825801	N	2	Qrob_P0248780.2	Qrob_P0248770.2	Receptor-like protein 12 (RLP12)
Chr08	52916114	52936114	N	0	--	--	--
Chr08	56741354	56761354	N	2	Qrob_P0415280.2	Qrob_P0415270.2	Probable xyloglucan 6-xylosyltransferase 3
Chr08	59586085	59616085	N	1	Qrob_P0101640.2	--	Vacuolar protein-sorting-associated protein 33
Chr08	59816085	59836085	N	2	Qrob_P0101700.2	Qrob_P0101690.2	vacuolar amino acid transporter 1-like . ANT1
Chr08	60266085	60286085	N	2	Qrob_P0101980.2	Qrob_P0101970.2	Unknown
Chr08	61106085	61126085	N	1	Qrob_P0572080.2	--	Unknown
Chr08	62493250	62513250	N	2	Qrob_P0412860.2	Qrob_P0412870.2	Two-component response regulator-like PRR73
Chr08	62673250	62693250	N	1	Qrob_P0749650.2	--	floral homeotic protein APETALA 2 | Ethylene-responsive TF RAP2-7
Chr08	63973861	63993861	N	0	--	--	--
Chr08	64263157	64283157	N	0	--	--	--
Chr08	66229909	66249909	N	1	Qrob_P0303790.2	--	Transcription factor MYB1
Chr08	66989918	67019918	N	0	--	--	--
Chr08	68418235	68443235	N	2	Qrob_P0457680.2	Qrob_P0457670.2	protein DEHYDRATION-INDUCED 19-like
Chr08	70120166	70140166	N	1	Qrob_P0457330.2	--	Transcription factor RF2b
Chr08	71211039	71231039	N	1	Qrob_P0181360.2	--	Serotonin N-acetyltransferase 1
Chr08	71251039	71271039	N	0	--	--	--
Chr09	7739510	7759510	N	1	Qrob_P0191830.2	--	Metal transporter Nramp5
Chr09	13740400	13760400	N	1	Qrob_P0566960.2	--	Probable ubiquitin-like-specific protease 2B
Chr09	16741132	16771132	N	0	--	--	--
Chr09	18456093	18476093	N	0	--	--	--
Chr09	18536093	18556093	N	3	Qrob_P0768740.2	Qrob_P0768720.2; Qrob_P0768730.2	Dehydration-responsive element-binding protein 1
Chr09	19056093	19076093	N	1	Qrob_P0206840.2	--	uncharacterized protein LOC108998503
Chr09	19266093	19286093	N	0	--	--	--
Chr09	19506093	19526093	N	1	Qrob_P0418880.2	--	1-aminocyclopropane-1-carboxylate synthase
Chr09	29196103	29216103	N	0	--	--	--
Chr09	41460086	41490086	N	0	--	--	--
Chr10	1720404	1750404	N	1	Qrob_P0433610.2	--	Methyl-CpG-binding domain-containing protein 9
Chr10	2434008	2454008	N	4	Qrob_P0634350.2	Qrob_P0634320.2 to Qrob_P0634370.2	11-beta-hydroxysteroid dehydrogenase-like 4A
Chr10	6014763	6034763	N	0	--	--	--
Chr10	10504776	10534776	N	0	--	--	--

Chr10	11878969	11898969	N	1	Qrob_P0536640.2	--	Ubiquitin-conjugating enzyme E2
Chr10	12310682	12340682	N	3	Qrob_P0070130.2	Qrob_P0070120.2; Qrob_P0070160.2	Transportin MOS14
Chr10	14523130	14553130	N	1	Qrob_P0452190.2	--	Mini zinc finger protein 3
Chr10	14923130	14943130	N	1	Qrob_P0555320.2	--	NAC transcription factors
Chr10	21638000	21658000	N	0	--	--	--
Chr10	22573510	22593510	N	0	--	--	--
Chr10	24461929	24481929	N	0	--	--	--
Chr10	36958768	36978768	N	2	Qrob_P0282980.2	Qrob_P0282960.2	Translation factor GUF1 homolog
Chr10	37193301	37208301	N	1	Qrob_P0611980.2	--	uncharacterized protein LOC108983424
Chr11	2223424	2243424	N	0	--	--	--
Chr11	6766779	6806779	N	2	Qrob_P0084310.2	Qrob_P0084290.2	uncharacterized protein LOC103402562
Chr11	6826779	6856779	N	1	Qrob_P0084340.2	--	Cytochrome P450 CYP749A22
Chr11	16411596	16431596	N	0	--	--	--
Chr11	27662255	27682255	N	1	Qrob_P0246000.2	--	Serine carboxypeptidase-like
Chr11	45788493	45808493	N	0	--	--	--
Chr11	51542779	51562779	N	1	Qrob_P0099750.2	--	Brefeldin A-inhibited guanine nucleotide-exchange protein 1
Chr12	3028743	3058743	N	3	Qrob_P0112330.2	Qrob_P0112340.2; Qrob_P0112350.2	Methionyl-tRNA formyltransferase
Chr12	3558743	3578743	N	0	--	--	--
Chr12	8356108	8376108	N	1	Qrob_P0444560.2	--	1-phosphatidylinositol-3-phosphate 5-kinase FAB1B
Chr12	19009403	19029403	N	3	Qrob_P0695190.2	Qrob_P0695160.2; Qrob_P0695180.2	Olee1-like protein
Chr12	19419502	19449502	N	1	Qrob_P0328580.2	--	Endo-1,3;1,4-beta-D-glucanase
Chr12	20889502	20909502	N	0	--	--	--
Chr12	25542574	25562574	N	2	Qrob_P0115570.2	Qrob_P0115580.2	Probable cyclic nucleotide-gated ion channel 14
Chr12	30514806	30534806	N	3	Qrob_P0436820.2	Qrob_P0436800.2; Qrob_P0436810.2	transcriptional corepressor LEUNIG-like isoform

Table S4: Density in outlier SNPs in genomic regions of the 32 gene candidates. rob-pet corresponds to *Q.robur*/*Q.petraea*; pyr-pet= *Q.pyrenaica*/*Q.petraea*; pub-pet = *Q.pubescens*-*Q.petraea*; rob-pub = *Q.robur*-*Q.pubescens*; rob-pyr = *Q. robur*-*Q.pyrenaica*, pub-pyr= *Q.pubescens*-*Q.pyrenaica*. Enrichment*** <0.001 ** <0.01 * <0.05.

			Proportion of outlier SNPs in the 10kb sliding window						
Chr.	**Position**	**Candidate gene**	**rob-pet**	**pyr-pet**	**pub-pet**	**rob-pub**	**rob-pyr**	**pub-pyr**	**Pattern**
Intrinsic barriers									
Flowering									
Chr08	62493250-62513250	Qrob_P0412860.2	0.0135	**0.2162*** **	0.0000	0.0000	**0.2387*** **	0.0000	complex
Chr08	62673250-62693250	Qrob_P0749650.2	0.0000	**0.2669*** **	0.012	**0.0319* **	**0.3108*** **	**0.4263*** **	complex
Chr07	44592072-44632072	Qrob_P0088630.2	0.0053	0.2027***	0.2773***	0.1013**	0.114*	0.016	complex
Cycloartenol synthase 2 genes									
Chr06	28660694-28680694	Qrob_P0684330.2	**0.0976** **	0.0041	**0.3699*** **	**0.0976** **	**0.1179** **	0.0122	complex
Chr06	28690694-28710694	Qrob_P0684400.2	**0.0296* **	**0.0574** **	**0.0500** **	**0.0407* **	**0.2241*** **	**0.0333* **	shared
Chr06	28770694-28810694	Qrob_P0684360.2	**0.1347** **	**0.0588** **	**0.0228* **	**0.0228* **	**0.2751*** **	**0.0531** **	shared
G-type lectin S-receptor-like serine/threonine-protein kinase genes									
Chr03	26118565-26138565	Qrob_P0538230.2	**0.1484** **	**0.2194*** **	0.0065	0.0129	0.0129	**0.2516*** **	complex
Chr05	4001785-4021785	Qrob_P0641650.2	0.0000	**0.0556** **	**0.2778*** **	**0.1111** **	0.0000	0.0000	complex
Chr06	1453464-1473464	Qrob_P0430150.2	0.0127	**0.0295* **	0.0000	0.0000	**0.2447*** **	**0.1814*** **	**Q. pyrenaica-specific**
Chr06	1513464-1563464	Qrob_P0430190.2	0.0000	**0.1596** **	0.0047	0.0047	**0.0751** **	**0.2676*** **	**Q. pyrenaica-specific**
Chr08	1749735-1769735	Qrob_P0138480.2	0.0037	**0.2132*** **	0.0000	0.0074	**0.2132*** **	**0.2500*** **	**Q. pyrenaica-specific**
Leaf, flower, embryo developments									
Chr12	30514806-30534806	Qrob_P0436820.2	**0.1333** **	**0.0667** **	**0.0667** **	**0.2667*** **	**0.0667** **	0.0000	**All except. Q. pub/Q.pyr**
Chr02	31611918-31651918	Qrob_P0297580.2	0.0042	**0.3583*** **	**0.0167* **	0.0083	**0.0167* **	**0.0917** **	complex
Chr06	43006752-43026752	Qrob_P0309300.2	0.0000	**0.2551*** **	**0.0816** **	**0.0663** **	**0.0255* **	**0.0561** **	**All except. Q. rob/Q.pet**
Chr08	49805801-49825801	Qrob_P0248780.2	0.0000	0.0136	0.0000	0.0027	**0.2439*** **	**0.3008*** **	complex
UV-B tolerance									
Chr02	37788250-37808250	Qrob_P0338870.2	**0.0521** **	**0.0208* **	**0.0677** **	**0.2865*** **	**0.0938** **	**0.0156* **	shared
Chr03	34667252-34687252	Qrob_P0500290.2	**0.0272* **	**0.0195* **	**0.2724*** **	**0.0584** **	**0.035* **	**0.0233* **	shared
Chr06	11400887-11420887	Qrob_P0577750.2	0.0043	**0.5174*** **	**0.0217* **	**0.0391* **	**0.1652** **	**0.0435* **	**All except. Q. rob/Q.pet**
Ecological barriers - abiotic stresses									
Natural resistance-associated macrophage protein (Nramp) genes									
Chr09	7739510-7759510	Qrob_P0191830.2	0.0061	0.0031	**0.1590** **	**0.1896*** **	0.0122	**0.2141*** **	**Q. pubescens-specific**
Chr06	38764748-38784748	Qrob_P0097150.2	0.0041	**0.0497** **	**0.2120*** **	**0.2972*** **	**0.1531** **	**0.0690** **	**All except. Q. rob/Q.pet**
Dehydration/lateral root growth									
Chr08	68418235-68443235	Qrob_P0457680.2	0.0000	**0.392*** **	**0.1200** **	**0.2640*** **	**0.208*** **	**0.1600** **	**All except. Q. rob/Q.pet**
Chr02	27061274-27081274	Qrob_P0304800.2	0.0132	**0.2863*** **	**0.0749** **	**0.0617** **	**0.2467*** **	**0.3084*** **	**All except. Q. rob/Q.pet**
Chr02	28240136-28260136	Qrob_P0299670.2	**0.0243* **	**0.4854*** **	0.0097	0.0146	0.0000	**0.5874*** **	complex
Chr02	36392234-36442234	Qrob_P0422470.2	**0.1972*** **	**0.2202*** **	**0.1789*** **	**0.2248*** **	**0.0550** **	**0.1376** **	shared
Chr09	19506093-19526093	Qrob_P0418880.2	0.0000	**0.0606** **	**0.0152* **	**0.0364* **	**0.0273* **	**0.2545*** **	**All except. Q. rob/Q.pet**

Chr04	19182244-19222244	Qrob_P0652510.2	**0.2094*** **	**0.2308*** **	**0.2009*** **	**0.0598** **	**0.0214* **	**0.0256* **	shared
Chr02	96008743-96028743	Qrob_P0387640.2	0.0000	**0.2403*** **	**0.2209*** **	**0.0465* **	**0.0349* **	**0.0853** **	**All except. Q. rob/Q.pet**
Freezing/cold adaptation									
Chr09	18536093-18556093	Qrob_P0768740.2	0.0000	0.0043	**0.3120*** **	**0.0470* **	**0.1496** **	**0.0470* **	complex
Chr02	112153196-112173196	Qrob_P0339800.2	**0.0741** **	**0.0185* **	**0.1852*** **	**0.1852*** **	**0.0185* **	0.0000	**All except. Q. pub/Q.pyr**
Ecological barriers - biotic stresses									
Pathogen resistance/mycorrhization									
Chr10	12310682-12340682	Qrob_P0070130.2	0.0000	**0.4086*** **	0.0029	0.0114	**0.4029*** **	**0.4343*** **	**Q. pyrenaica-specific**
Chr01	26763631-26783631	Qrob_P0648170.2	0.0044	0.0088	**0.0175* **	**0.1798*** **	0.0000	0.0088	complex
Chr05	3053435-3073435	Qrob_P0622110.2	0.0000	0.0000	**0.2564*** **	**0.2821*** **	**0.0256* **	**0.0513** **	complex

825 Table S5: Sampling site locations and number of individuals pooled for each species

Species	Sampling sites (South-West France)	GPS coordinates	# Individuals
Q. petraea	Laveyron	43.97476667 ; -0.27972222	13
Q. pubescens	Branne	44.839961 ; -0.204878	12
	Blaignan	45.319174 ; -0.855917	6
		Total pooled	18
Q. robur	ISS Landes	44.226283131 ; -1.011284422	20
Q. pyrenaica	ISS Landes	44.270133145 ; -1.0696972	20

826