Adaptive introgression as a driver of local adaptation to climate in European white oaks

Thibault Leroy\(^1,2\), Jean-Marc Louvet\(^1\), Céline Lalanne\(^1\), Grégoire Le Provost\(^1\), Karine Labadie\(^3\) and Antoine Kremer\(^1\)

\(^1\)BIOGECO, INRA, Université de Bordeaux, 69 Route d’Arcachon, 33612, Cestas, France; \(^2\)ISEM, CNRS, IRD, EPHE, University of Montpellier, Place Eugène Bataillon, 34095, Montpellier, France; \(^3\)Genoscope, Institut de biologie François-Jacob, Commissariat à l’Energie Atomique (CEA), Université Paris-Saclay, Evry, France

Introduction

Evolutionary biologists are becoming increasingly fascinated by the tracking of adaptive genetic changes, as our understanding of paleoecology and genomics (Shapiro & Hofreiter, 2014) and climate reconstructions at various spatial and temporal scales (Mauri et al., 2015) improve. Ultimately, assembling data relating to historical and genetic changes will increase our understanding of how, when, and how rapidly evolutionary shifts have enhanced adaptation. The recent report that introgression with Neanderthals or Denisovans increased the adaptive potential of modern Eurasian humans provides an emblematic example of a major evolutionary shift supported by historical and genomic evidence (Dannemann & Racimo, 2018). Adaptive shifts have also been predicted in nonmodel plants and animals that have repeatedly witnessed large-scale environmental changes over longer temporal scales due to Quaternary climatic oscillations (Dynesius & Jansson, 2000). Forest trees are relevant target species for explorations of evolutionary changes, as their past history and distribution can be reconstructed easily, owing to the availability of large quantities of fossil remains (Brewer et al., 2017; Wagner et al., 2018). Furthermore, understanding how trees adapt to changing environments has become a major topic of interest in theoretical and applied ecological genetics and genomics. On the one hand, there are concerns that trees may not cope with the velocity of ongoing climatic change; on the other, land managers and foresters are seeking methods of adaptive management (Lindner et al., 2010). Adaptation can be triggered by new alleles originating from new mutations, neutral standing genetic variation, or adaptive introgression. Adaptation can be investigated empirically by assessing adaptive divergence in common garden experiments, or by tracking genetic changes over successive generations. The common garden approach has been widely

Summary

- Latitudinal and elevational gradients provide valuable experimental settings for studies of the potential impact of global warming on forest tree species. The availability of long-term phenological surveys in common garden experiments for traits associated with climate, such as bud flushing for sessile oaks (Quercus petraea), provide an ideal opportunity to investigate this impact.
- We sequenced 18 sessile oak populations and used available sequencing data for three other closely related European white oak species (Quercus pyrenaica, Quercus pubescens, and Quercus robur) to explore the evolutionary processes responsible for shaping the genetic variation across latitudinal and elevational gradients in extant sessile oaks. We used phenotypic and genomic evidence (Dannemann & Racimo, 2018) to explore the evolutionary processes responsible for shaping the genetic variation across latitudinal and elevational gradients in extant sessile oaks. We used phenotypic surveys in common garden experiments and climatic data for the population of origin to perform genome-wide scans for population differentiation and genotype–environment and genotype–phenotype associations.
- The inferred historical relationships between Q. petraea populations suggest that interspecific gene flow occurred between Q. robur and Q. petraea populations from cooler or wetter areas. A genome-wide scan of differentiation between Q. petraea populations identified single nucleotide polymorphisms (SNPs) displaying strong interspecific relative divergence between these two species. These SNPs followed genetic clines along climatic or phenotypic gradients, providing further support for the likely contribution of introgression to the adaptive divergence of Q. petraea populations.
- Overall, the results indicate that outliers and associated SNPs are Q. robur ancestry-informative. We discuss the results of this study in the framework of the postglacial colonization scenario, in which introgression and diversifying selection have been proposed as essential drivers of Q. petraea microevolution.
used for long-lived species, such as trees, and has revealed the existence of high levels of adaptive divergence between extant populations distributed along large geographic gradients (Alberto et al., 2013a). Moreover, adaptive divergence is maintained between tree populations, despite extensive gene flow potentially causing pollen swamping effects that might constrain adaptation (Savolainen et al., 2007). In many European tree species, pollen swamping results from intra- or interspecific gene flow, as many congeneric species live together in the same stands and interspecific hybridization is widespread. In contrast to the pollen swamping effect, it has also been suggested that gene flow may actually enhance adaptation by increasing the ability of populations to respond to natural selection (Kremer & Le Corre, 2012; Yeaman, 2015). It has also recently been suggested that interspecific gene flow may enhance adaptation by facilitating adaptive introgression (Suarez-Gonzalez et al., 2018). Here, we explore the genomic footprints of adaptation variation in sessile oak (Quercus petraea (Matt.) Liebl.), a widespread species in Europe, and address the potential contribution of interspecific gene flow to local adaptation.

The geographic distribution of sessile oak extends from Spain to southern Scandinavian, and this species mostly grows in forests also containing the closely related pedunculate oak (Quercus robur), another white oak species with a range that extends farther north and east, to the Ural Mountains (Leroy et al., 2020). At more southern latitudes, sessile oak also shares habitats with pubescent (Quercus pubescens) and Pyrenean (Quercus pyrenaica) oaks. Hybridization between these four European white oak species has been reported in sympatric stands (Curtu et al., 2007; Lepais & Gerber, 2011). Furthermore, hybridization is thought to have played a major role in the expansion of Q. petraea populations during postglacial recolonization, as suggested by the widespread sharing of chloroplast genomes between Q. petraea and Q. robur in sympatric stands (Petit et al., 2002). Interestingly, demographic inferences based on approximate Bayesian computation simulations support scenarios in which these four European white oak species came into secondary contact, at the onset of the last glacial period (Leroy et al., 2017, 2020). A number of questions relating to this scenario remain unanswered. How much did introgression actually contribute to today’s distribution of Q. petraea? Did introgression contribute to the adaptation of Q. petraea or was it a purely neutral process? Which adaptive alleles were introgressed, and from which other white oak species? We tackle these questions here by exploring imprints of historical population splits and admixture events through studies of whole-genome sequences from Q. petraea populations. We also performed genome scans to search for genomic footprints of divergence between Q. petraea populations, with either phenotypic data from common garden experiments or climatic data for the populations of origin. This experimental setup permitted us to test our earlier suggested hypothesis that introgression facilitated the postglacial expansion of Q. petraea following the tracks of Q. robur, a more pioneer species (Petit et al., 2004).

Materials and Methods

Sampling and sequencing

We sampled eight Q. petraea populations (with up to 20 individuals per pool; Table 1) from lowlands to middle elevations in the Pyrenees (up to 1600 m) in southwest France. These eight populations are distributed along two neighboring elevational transects (the Ossau and Luz valleys) and are stands of natural origin (elevation of 100 m to 1600 m). We also collected data for 10 Q. petraea populations (with up to 25 individuals per pool; Table 1) growing in a large common garden experiment in Silléngny, eastern France, corresponding to a total of 116 sessile oak provenances in Europe (Saenz-Romero et al., 2017). The populations selected were sampled along a latitudinal gradient extending from southern France to northern Germany (latitude 44°N to 53°N).

We also included a reference population of the four main white oak species in Europe (Q. petraea, Q. robur, Q. pubescens, Q. pyrenaica) corresponding to four oak stands located in southwest France analyzed in a companion paper (Leroy et al., 2020). Samples of reference species populations came from the same geographic region. Such a sampling strategy does not bias species comparison, as shown by an earlier methodological study showing that species differentiation is only moderately impacted by the geographic origin of populations (Bodénès et al., 1997). The Q. petraea reference population of Leroy et al. (2020) corresponded to a pool of 13 individuals from the same low-elevation population of the Luz valley (L1: Laveyron in Table 1). We used the two populations as pseudoreplicates, to check the accuracy of our pool-sequencing strategy.

DNA was extracted from individual trees with a cetyltrimethyl ammonium bromide DNA extraction protocol (Doyle & Doyle, 1987) (latitudinal gradient) or with the Invisorb Spin Plant Mini Kit (Startec Molecular; elevational gradient) according to the manufacturer’s instructions (Startec Molecular GmbH, Berlin, Germany). DNA yields were evaluated with a NanoDrop 1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) (elevational gradient) or with an Infinite F200 (Tecan Group Ltd, Männedorf, Switzerland), and DNA samples were mixed in equimolar amounts to obtain a single pool for each population. After trimming, the number of 2 × 100 bp paired-end reads retained for analyses ranged from 668 million to 998 million reads per pool, corresponding to a rough estimate of sequencing coverage of 180–270×, assuming an oak genome size of 740 Mb. The raw data have been deposited in the Sequence Read Archive: PRJEB32209.

Climatic and phenological data

Monthly mean climate data were obtained from WorldClim (Hijmans et al., 2005) for the period 1950–2000. We downscaled these values to a spatial resolution of 100 m using locally weighted regressions and a finer resolution digital elevation model (100 m) before point overlay and accounted for topography as previously described (Zimmermann et al., 2009; Dullinger et al., 2012). We
then calculated yearly average temperature and annual precipitation sums (Supporting Information Fig. S1; Table 1).

The date of leaf unfolding was recorded separately in two different common garden experiments. Saplings of populations sampled along the elevational gradient were transplanted to a common garden experiment in Toulenne in southwest France in spring 2007, and phenological observations were conducted over seven successive years (2009–2015) – see Firmat et al. (2017) for further details. Similarly, saplings of populations sampled along the latitudinal gradient were installed in a common garden located in the northeast of France in Sillegny in 1989 and 1993, and phenological observations were conducted in 2015 – see Firmat et al. (2017) and Torres-Ruiz et al. (2019) for further details. Population means were calculated in each common garden, and standardized values for each common garden were used to study genotype–phenotype associations across the two common gardens (Table 1). Earlier comparisons across different common garden experiments showed, indeed, that the date of leaf unfolding exhibited strong genetic stability; for example, high genetic correlations across sites (Detory et al., 2010).

### Mapping and single nucleotide polymorphism calling

We used the pipeline described by Leroy et al. (2020); see also https://github.com/ThibaultLeroyFr/GenomeScansByABC/tree/master/SNP_calling_filtering to identify reliable single nucleotide polymorphisms (SNPs). In brief, we used BOWTIE2 v.2.1.0 to map sequencing data onto the v.2.3 oak haplome assembly (Plomion et al., 2018), removed duplicates with Snp-frequency-diff.pl from the POPULATION2 suite (Koffler et al., 2011) to select biallelic SNPs with at least 10 alternative alleles. Positions with a mean coverage of $< 50 \times$ for any of the 18 populations and sites in the top 2% in terms of coverage per population were ignored. SNPs with a minimum allele frequency $< 0.02$ were filtered out to exclude most Illumina sequencing errors. A total of 37 062 111 SNPs was identified (Table S1).

### Population splits and gene flow inferences

We used genome-wide allele frequency data derived from allele counts to build a population tree with TREE MIX v.1.12 (Pickrell & Pritchard, 2012) and to test for the presence of gene flow between populations. Given that Q. petraea populations are known to be frequently connected by gene flow with closely related species (Lepais & Gerber, 2011; Leroy et al., 2017), we also evaluated the influence of interspecific gene flow on tree topology. We performed an analysis with a set of 1757 476 infragenic SNPs commonly detected in the 18 previously described sessile oak populations, and in the reference populations of four white oak species (including Q. petraea) described by Leroy et al. (2020) were included in the analysis. Variable numbers of migration nodes $m$, ranging from 0 to 13, were evaluated. For each fixed value of $m$, we performed 1000 replicated TREE MIX analyses with a python script from Michael G. Harvey (https://github.com/mgharvey/misc_python/blob/master/bin/TreeMix/treemix_tree_with_bootstrap.py). We then used SumTrees.py from the DENDROPY suite to generate a consensus tree with all the bootstrap values (Sukumaran & Holder, 2010). We used the total variation explained for each replicate in the R companion script of TREE MIX (plotting_funcs.R) as a judgment criterion for evaluation of the best number of migration nodes. We also evaluated the robustness of

### Geographic and climatic data for the Quercus petraea populations studied.

<table>
<thead>
<tr>
<th>Code</th>
<th>Location</th>
<th>Elevation (m)</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Temperature</th>
<th>Precipitation (mm yr$^{-1}$)</th>
<th>Leaf unfolding</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>Laveyron, Luz Valley, France</td>
<td>131</td>
<td>43.75</td>
<td>–0.22</td>
<td>12.33</td>
<td>901</td>
<td>–1.333</td>
<td>20</td>
</tr>
<tr>
<td>L8</td>
<td>Chêze, Luz Valley, France</td>
<td>803</td>
<td>42.92</td>
<td>–0.03</td>
<td>9.27</td>
<td>914</td>
<td>0.817</td>
<td>20</td>
</tr>
<tr>
<td>L12</td>
<td>Gédre, Luz Valley, France</td>
<td>1235</td>
<td>42.78</td>
<td>0.02</td>
<td>7.12</td>
<td>1016</td>
<td>1.011</td>
<td>20</td>
</tr>
<tr>
<td>L16</td>
<td>Péguières, Luz Valley, France</td>
<td>1630</td>
<td>42.87</td>
<td>–0.12</td>
<td>6.58</td>
<td>982</td>
<td>1.724</td>
<td>18</td>
</tr>
<tr>
<td>O1</td>
<td>Josbaig, Ossau Valley, France</td>
<td>259</td>
<td>43.22</td>
<td>–0.73</td>
<td>12.24</td>
<td>979</td>
<td>–1.309</td>
<td>20</td>
</tr>
<tr>
<td>O8</td>
<td>Le Hourcq, Ossau Valley, France</td>
<td>841</td>
<td>42.90</td>
<td>–0.43</td>
<td>9.16</td>
<td>933</td>
<td>–0.324</td>
<td>20</td>
</tr>
<tr>
<td>O12</td>
<td>Gabas, Ossau Valley, France</td>
<td>1194</td>
<td>42.88</td>
<td>–0.42</td>
<td>7.35</td>
<td>1031</td>
<td>0.036</td>
<td>20</td>
</tr>
<tr>
<td>O16</td>
<td>Arloute, Ossau Valley, France</td>
<td>1614</td>
<td>42.88</td>
<td>–0.40</td>
<td>5.16</td>
<td>1164</td>
<td>0.427</td>
<td>10</td>
</tr>
</tbody>
</table>

Date of leaf unfolding expressed as standardized values for common gardens (see the Materials and Methods section). Negative values indicate early flushing, and positive values indicate late flushing.

---

**Table 1** Geographic and climatic data for the *Quercus petraea* populations studied.

<table>
<thead>
<tr>
<th>Code</th>
<th>Location</th>
<th>Elevation (m)</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Temperature</th>
<th>Precipitation (mm yr$^{-1}$)</th>
<th>Leaf unfolding</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Saint Sauvant, France</td>
<td>155</td>
<td>46.38</td>
<td>0.12</td>
<td>11.78</td>
<td>786</td>
<td>–0.166</td>
<td>25</td>
</tr>
<tr>
<td>97</td>
<td>Grésigne, France</td>
<td>310</td>
<td>44.04</td>
<td>1.75</td>
<td>12.05</td>
<td>791</td>
<td>–1.139</td>
<td>25</td>
</tr>
<tr>
<td>124</td>
<td>Killarney, Ireland</td>
<td>50</td>
<td>52.01</td>
<td>–9.50</td>
<td>9.96</td>
<td>1362</td>
<td>4.084</td>
<td>25</td>
</tr>
<tr>
<td>204</td>
<td>Bézanges, France</td>
<td>275</td>
<td>48.76</td>
<td>6.49</td>
<td>9.50</td>
<td>751</td>
<td>0.371</td>
<td>25</td>
</tr>
<tr>
<td>217</td>
<td>Berce, France</td>
<td>165</td>
<td>47.81</td>
<td>0.39</td>
<td>10.65</td>
<td>698</td>
<td>0.434</td>
<td>25</td>
</tr>
<tr>
<td>218</td>
<td>Longchamps, France</td>
<td>235</td>
<td>47.26</td>
<td>5.31</td>
<td>10.59</td>
<td>801</td>
<td>–0.920</td>
<td>22</td>
</tr>
<tr>
<td>219</td>
<td>Tronas, France</td>
<td>245</td>
<td>46.68</td>
<td>2.83</td>
<td>10.63</td>
<td>742</td>
<td>1.350</td>
<td>25</td>
</tr>
<tr>
<td>233</td>
<td>Vacheres, France</td>
<td>650</td>
<td>43.90</td>
<td>5.63</td>
<td>10.22</td>
<td>797</td>
<td>–1.532</td>
<td>25</td>
</tr>
<tr>
<td>253</td>
<td>Gohrd, Germany</td>
<td>85</td>
<td>53.10</td>
<td>10.86</td>
<td>8.30</td>
<td>635</td>
<td>0.983</td>
<td>25</td>
</tr>
<tr>
<td>256</td>
<td>Lappwald, Germany</td>
<td>180</td>
<td>52.26</td>
<td>10.99</td>
<td>8.50</td>
<td>597</td>
<td>0.650</td>
<td>25</td>
</tr>
</tbody>
</table>
the migration nodes, by counting the number of replicated analyses supporting the migration node.

We used the $f_3$ statistic calculated with the THREPOP v.0.1 from the TREEmix suite (~k 1000) to determine whether a focal population $X$ was the product of admixture between two populations $Y$ and $Z$. Significant negative values of $f_3$ ($X;Y,Z$) were considered to indicate admixture — for details, see Reich et al. (2009) and Schaeffer et al. (2016). The $f_3$ statistics were calculated for all possible three-population combinations, but we report only $f_3$ ($X;Y,Q.robur$) values, given the known particular contribution of $Q.robur$ to the evolution of the populations investigated (see the Results section). No significant negative values were obtained for either $f_3$ ($X;Y,Q.pyrenaica$) or $f_3$ ($X;Y,Q.pubescens$).

**Genome scans** Genome-wide scans were performed with the BayPASS v.2.1 software package (Gautier, 2015). BayPASS takes confounding demographic effects into account by estimating the covariance matrix of allele frequency between populations. The core model reports locus $XtX$, which is analogous to $F_{ST}$ but explicitly corrected for this covariance matrix, thus accounting for neutral correlations of allelic frequencies (Gunther & Coop, 2013). For each SNP, BayPASS can also compare models integrating population-specific covariates (here, temperature, precipitation, and date of leaf unfolding) by including a regression coefficient for the cline along the covariate gradient in the base model. All the covariates used here were scaled as described by Gautier (2015). Direct comparisons of the likelihood of models (Bayes factors, BFs) including and not including particular covariates were used to evaluate the genotype–environment association (GEA), defined as the association between changes in allele frequency and each population-specific climate variable (temperature, precipitation), and the genotype–phenotype association (GPA), defined as the association between changes in allele frequency and mean date of leaf unfolding. The identification of outliers ($XtX$ or BFs) was based on a calibration procedure using pseudo-observed datasets (PODs), as discussed by Gautier (2015). For this step, we used the BayPass UTILS.R script as a source for the simulation of 100 000 SNPs. We then ran BayPASS again, to generate quantile values for $XtX$ and BFs based on these PODs. Conservative and very conservative quantile values were used (‘minor outlier’ 0.999 for $XtX$ and 0.9999 for BF; ‘main outliers’ 0.99999 for both $XtX$ and BF) as thresholds to identify outliers. We then generated clusters of neighboring SNPs, by bulking SNPs separated by <10 kb. We investigated only clusters of associated SNPs containing at least two outliers, including a ‘main’ outlier, to exclude the random associations that would be expected to occur in such a large dataset. We also investigated clusters with no ‘main’ outlier, provided that the cluster concerned contained at least five ‘minor’ outliers.

**Manual gene annotations** We first performed manual gene annotations based on protein blast searches ($P<1 \times 10^{-5}$) against the Arabidopsis protome. Gene functions were identified by manual inspection of all the best hits per candidate gene per region. In this study, functional annotation was achieved by performing extensive manual literature searches for each gene, using the strategy described by Leroy et al. (2020) rather than automatic approaches based on Gene Ontology-oriented methods. This approach was preferred to ensure that we obtained accurate and detailed information about gene function, supported by checked and traceable references.

**Results**

**Population history and splits** We used TREEmix to infer the evolutionary history of European white oaks, focusing, in particular, on the most likely sequence of population splits. This software infers the relationships between populations as a bifurcating tree. Interestingly, a strict drift model can explain most of the variance in relatedness between populations (mean over 1000 replicates: 0.926; median: 0.950; Fig. 1a). Looking at the $Q.petreae$ populations (Figs 1b, S2), the consensus tree over the 1000 replicates supports basal splits and longer branches for populations at higher elevations than for lowland populations. The only exception was population #124 (Killarney, Ireland), which was at the base of all $Q.petreae$ populations. In comparisons with all other $Q.petreae$ populations and the $Q.robur$ reference population, population #124 also presented significant negative values of the $f_3$ statistic (Fig. S3), suggesting that this population is probably an admixture between $Q.petreae$ and $Q.robur$. Several significant negative $f_3$ values were also obtained for the L16 population (Pégues, France), for comparisons with several lowland populations (#9: Saint Sauvant, France; #97: Gréssigne, France; and #233: Vachères, France; Fig. S3).

We then sequentially added migration events to the tree (Fig. S2). As expected, the variance in relatedness between populations was better explained as the number of migration events increased. At $m=1$, the proportion of the variance explained increased (mean: 0.955; Fig. 1a) in two-thirds of replicated simulations due to a migration node between $Q.pubescens$ and the L8 population (Chêze, France; 661/1000). Otherwise, TREEmix inferred (219/1000) a migration event between $Q.robur$ and population #124. The inferred tree remained inconsistent with the expected species tree (Leroy et al., 2017, 2020), but this first node helped to explain most of variance generated by the relative positions of $Q.pubescens$ and $Q.robur$. At $m=4$ (Fig. 1c), the proportion of the variance explained again increased substantially (mean: 0.976) and the bootstrap replicate maximizing the likelihood was, for the first time, consistent with the expected species tree. Again, TREEmix inferred a migration event between $Q.robur$ and the Irish population (#124). It also provided support for candidate introgression events from $Q.robur$ for populations at the highest elevation (i.e. from $Q.robur$ to O16 (Artouste, France), and from $Q.robur$ to the ancestral population of the modern L12 (Gèdre, France), L16 (Pégues, France), and #124 (Killarney, Ireland) populations; Fig. 1c).
Genome scans

For the identification of SNPs potentially subject to selection, and therefore displaying higher levels of differentiation than would be expected under a hypothesis of neutrality, we calculated the $X_tX$ statistic, an $F_{ST}$-like statistic explicitly accounting for population history (Gunther & Coop, 2013). We identified 761,554 SNPs deviating from neutral expectations (2.05% of all SNPs investigated, presenting $X_tX$ values above the 0.999 quantile threshold and referred to hereafter as ‘minor outliers’), including 107,764 for which the evidence was strong (0.29%, with $X_tX$ values above the 0.99999 quantile threshold referred to hereafter as ‘main outliers’; Table S1). These SNPs are distributed over all the chromosomes (outer circle, Fig. 2). Interestingly, $X_tX$ outliers were found to be strongly enriched in SNPs that were highly differentiated between species, particularly between $Q. robur$ and $Q. petraea$ (Fig. 3). More broadly, we observed a strong correlation between the intraspecific $X_tX$ value estimated for the 18 populations and the interspecific $F_{ST}$ between $Q. robur$ and $Q. petraea$ ($P<2.2 \times 10^{-16}$, $R^2=0.247$; Fig. S4).

Among the outlying SNPs detected with the $X_tX$ statistic, we identified GEA- and GPA-associated SNPs covarying with mean annual temperature, precipitation, or of the date of leaf unfolding recorded in common gardens (Figs 2, S5–S16). In total, we identified 5682 SNPs associated with at least one covariate. More precisely, we identified 1331 SNPs as outliers (including 216 main outliers) associated with temperature, 2932 as outliers (277 main outliers) associated with precipitation, and 1572 as outliers (125 main outliers) associated with leaf unfolding (Table S1). For the 153 SNPs involved in two different associations with the three covariates, the largest proportion of SNPs (143/153, 93%) was significantly associated with both temperature and leaf unfolding. The remaining set of common SNPs was found to be associated with both temperature and precipitation (6/153, 4%) or with both precipitation and leaf unfolding (4/153, 3%).

The significantly associated SNPs were highly enriched in SNPs strongly differentiated between $Q. petraea$ and $Q. robur$, especially leaf unfolding- and temperature-associated SNPs (Fig. 4). We applied a binning procedure to group covariable-associated SNPs in close vicinity within the genome (<10 kb apart). The SNPs were located in 780, 1617 and 1033 independent genomic regions for temperature, precipitation, and leaf unfolding, respectively (Table S1). No genomic regions matching the two GEA and GPA associations were detected, as expected given the near independence of the two climate covariates (Fig. S1C). In the vast majority of cases (87.1%, 76.0%, and 87.0% for temperature, precipitation and leaf unfolding, respectively), the genomic region consisted of a single associated SNP. Manual gene annotations were performed only for genes located in the 201 genomic regions containing at least two associated SNPs (see the Materials and Methods section for details). These regions accounted for a total of 2016 associated SNPs (mean: 10.2 associated SNPs/region; range: 2–137). These 201 regions included 13 that partially overlapped for leaf unfolding and temperature (Fig. 2, on chromosomes 1, 8, and 10 (see also Figs S5, S12, S14), and on scaffold Sc0000849), three that partially overlapped for precipitation and temperature (Fig. 2, on chromosomes 6, 9 and 10; see also Figs S10, S13, S14), and two that...
partially overlapped for precipitation and leaf unfolding (Fig. 2, on chromosomes 3 and 6; see also Figs S7, S10).

Gene annotations
Manual gene annotations were performed for the genes located within or close to these 201 regions (within 5 kb on either side) to exclude border effects. This led to the identification of 167 unique candidate genes. We found regulators of various growth and development processes in plants. As expected, we identified a set of genes involved in stomatal responses to water stress (e.g. ALY4, ALMT9, PBL10, PP2AA, RopGEF1, SUS3, SCD1). Concordantly, our genome scans also revealed some temperature-associated genes acting as regulators of the production of...
Gibberellins (e.g. GAMT2, ASPG1 or GA2ox). These genes also play important roles in various developmental processes, including seed dormancy (Shen et al., 2018).

We investigated the processes at work further, by generating ecological clines for each associated SNP (Notes S1). Most of the associated SNPs followed complex clines along the ecological gradients (i.e. nonlinear), highlighting the advantages of model implementation in BAYENV or BAYPASS rather than the use of linear or logistic model-derived methods (De Mita et al., 2013), but some SNPs displayed an almost linear change in allele frequency along the ecological gradient (Fig. 5). The clines were generally consistent with a continuum between the Q. petraea and Q. robur reference pools, with higher levels of Q. robur-like alleles in populations living in cooler and/or wetter environments, suggesting that interspecific introgression from Q. robur is an important source of adaptive variation for Q. petraea populations. Typical examples include the precipitation-associated SNPs and leaf-unfolding-associated SNPs located within two important genes.

Fig. 3 Interspecific $F_{ST}$ between the Quercus petraea reference pool and reference populations for three other oak species (for details, see Leroy et al., 2020) at XIX outlier loci. Tukey’s honestly significant difference criterion at a significance level of 0.05 is reported. (a, b, h) Quercus petraea–Quercus robur differentiation; (c, d, g) Q. petraea–Quercus pyrenaica differentiation; (e, f, i) Q. petraea–Quercus pubescens differentiation.

Fig. 4 Variation in interspecific $F_{ST}$ between Quercus robur and Quercus petraea for nonoutlier and outlier single nucleotide polymorphisms (SNPs), including sets of associated with polymorphisms genotype–environment association (GEA) and genotype-phenotype association (GPA). Tukey’s honestly significant difference criterion at a significance level of 0.05 is reported. SNPs for the GEA and GPA sets correspond to all associated SNPs (i.e. both the minor and main categories).
Discussion

We assembled data for phenotypic surveys in common garden experiments, climatic data for the origin of the population, and whole-genome sequences to assess the genetic divergence of the timing of leaf unfolding between extant populations of *Quercus petraea* at the genomic level. We drew inferences about population history by searching for footprints of population splits and admixture, and we conducted genome scans to explore the establishment of genetic divergence during the early Holocene in extant populations. We found that introgression from *Quercus robur* had made a major contribution to population divergence of some locally adapted *Quercus petraea* populations. This finding was supported by three major outcomes. First, we found that admixture events occurred mainly in *Quercus petraea* populations from cooler and wetter climates. Second, the SNPs displaying the highest level of genetic differentiation between *Quercus petraea* populations were also highly differentiated between *Quercus petraea* and *Quercus robur*. Third, some of the genes contributing to phenological divergence in *Quercus petraea* displayed clinal variation consistent with the geographic variation of introgression. These findings echo earlier results in a Californian oaks showing that introgression follows environmental gradients. For example, the level of admixture in *Quercus wislizenii* was correlated with climatic variables, with more introgressed alleles from *Quercus parvula* and *Quercus agrifolia* in areas where summer temperatures were lower (Dodd & Afzal-Rafii, 2004). Conversely, associations with moisture gradients were revealed in a Californian white oak (*Quercus engelmannii*), with less introgression at drought-associated genes (Oney-Birol et al., 2018).

Historical introgression from *Quercus robur* into *Quercus petraea*

The gene flow events inferred by *TreeMix* suggest that admixture events, mostly involving *Quercus robur*, underlie the genetic differentiation of current *Quercus petraea* populations over wide geographic gradients. Contemporary hybridization between other closely related white oaks and *Quercus petraea* has been reported in empirical contemporary gene flow studies (Curtat et al., 2007; Salvini et al., 2009; Lepais & Gerber, 2011), but we found evidence only for an evolutionary footprint of admixture with *Quercus robur* in our sampled populations of *Quercus petraea*. Today, *Quercus petraea* and *Quercus robur* are frequently found in sympatric stands across Europe, due to their common pattern of postglacial colonization dynamics. As suggested by the widespread sharing of chloroplast haplotypes between these two closely related species when present in the same stands (Petit et al., 2002), hybridization followed by recurrent backcrossing between *Quercus petraea* and *Quercus robur* has been an essential mechanism in *Quercus petraea* expansion. Similarly, hybridization and subsequent backcrossing between the pioneer (resident) species *Quercus robur* and the late successional (invading) species *Quercus petraea* made it possible for *Quercus petraea* to migrate with *Quercus robur* (Petit et al., 2004; Guichoux et al., 2013). Most of the admixture events inferred from the *TreeMix* model support directional introgression from *Quercus robur* to *Quercus petraea* (Fig. 1), as predicted by theory in the case of this invasion scenario (Currat et al., 2008).

The ultimate outcome of hybridization followed by unidirectional backcrossing is the restoration of *Quercus petraea* within *Quercus robur* stands. As suggested by artificial backcross experiments performed with breeding populations of other closely related European white oak species (Diskin et al., 2006), the genome of the invading species (here *Quercus petraea*) should be regenerated in a limited number of generations, ultimately maintaining a limited imprint of the introgression from *Quercus robur*. However, our results show that genes introgressed from *Quercus robur* are maintained in

**Fig. 5** Near-linear genetic clines along ecological gradients. The clines shown are for the three single nucleotide polymorphisms (SNPs) detected by *BayPass*: (a) the precipitation-associated SNP Chr11:47276412, (b) the leaf-unfolding-associated SNP Chr07:51928762, and (c) the temperature-associated SNP Chr07:28987726. Best linear regressions for each associated SNP are shown in gray. Red, blue, green, and orange dotted lines correspond to allele frequencies estimated in the *Quercus petraea*, *Quercus pubescens*, *Quercus robur*, and *Quercus pyrenaica* reference populations, respectively. Clines for all other associated SNPs are shown in Supporting Information Notes S1.
Q. petraea populations, particularly those from northern latitudes or higher elevations. Previous studies based on Bayesian clustering methods have already highlighted the occurrence of more admixture at higher elevations (Alberto et al., 2010). Introgressed genes may still be present if admixture occurred recently, if there is a demographic imbalance between the invader and resident populations during initial contact (Currat et al., 2008), or if these populations are subject to directional selection. Studies of pollen remains have indicated that temperate deciduous oaks have been present in sampled areas since c. 10 000 yr BP in northern Germany (Alberto et al., 2010; Giesecke, 2016) and the Pyrénées (Jalut et al., 1992; Reille & Lowe, 1993) and since c. 8000 yr BP in Ireland (Kelleher et al., 2004). This timescale would be long enough for completion of the regeneration process and the eradication of introgression. We cannot rule out the possibility that demographic imbalance between these two species facilitated the introgression and maintenance of Q. robur alleles in Q. petraea, but the detection of introgression principally at higher elevations and latitudes suggests that some introgressed alleles were most probably maintained by selection, increasing the degree of differentiation between Q. petraea populations (Fig. 4). Interestingly, our results also show that genes highly differentiated between Q. petraea populations are also highly differentiated between Q. petraea and Q. robur (Figs 4, 5), suggesting their involvement in either adaptive divergence or reproductive barriers between the two species (Leroy et al., 2020). These two species display subtle differences in soil preferences when present in the same forest (Timbal & Aussenac, 1996; Eaton et al., 2016), but they also have different climate responses, as suggested by their allopatric (temperature- and precipitation-dependent) distributions at the edge of their ranges. Quercus robur extends farther north (up to Finland) and east (up to the Ural Mountains) – see Leroy et al. (2020) – and has a higher frequency in wetter climates (Eaton et al., 2016). We suspect that the introgression of Q. robur alleles for genes involved in these differential responses may have contributed to the expansion of Q. petraea populations to higher elevations in the Pyrenees and wetter climates in Ireland.

Genomic and genetic clines shaped by introgression

Association studies identified additional elements relating to the probable contribution of introgression to the adaptive divergence of Q. petraea populations. First, we recovered the clinal phenotypic gradient of leaf unfolding with temperature variations (Fig. S1B), as observed in previous common garden experiments (Ducouso et al., 1996; Vitasse et al., 2009; Alberto et al., 2011,2013b; Firmat et al., 2017), with populations from cooler climates (higher latitudes or elevations) flushing later than populations from warmer climates. Second, we found that the genes displaying clinal variations of allelic frequency along a temperature gradient were enriched in genes differentiated between the Q. petraea and Q. robur species. This raises questions about whether the genes introgressed from Q. robur also contributed to the later flushing of Q. petraea at higher elevations. No significant interspecific differences in leaf unfolding are observed between these two closely related species in common garden experiments or in situ (Kleinschmit, 1993; Jensen & Hansen, 2008; Wilkinson et al., 2017), although there is a slight trend towards earlier flushing in Q. petraea (Kleinschmit, 1993). However, Q. robur is well known to display ecotypic differentiation, with the recognition of late-flushing populations that have been attributed subspecies status as Q. robur var. tardiflora, mostly in eastern Europe, and flush almost a month later than Q. robur (Wesolowski & Rovinski, 2008; Utikina & Rubtsov, 2017). Similarly extreme late-flushing populations have also been reported in western Europe for Q. robur (Riedacker, 1968), but no such phenological differentiation has been reported for Q. petraea.

In conclusion, we found that most outliers and associated loci are Q. robur ancestry-informative, and that introgression from Q. robur to Q. petraea is correlated to the overall clinal variation for some key adaptive traits, including leaf-unfolding-related traits in particular. Our results are, however, only based on statistical associations and do not necessarily imply causality. Complex demographic scenarios, as well as linked selection, are known to produce footprints that mimic local selection (e.g. Fraisse et al., 2018; Simon & Duranton, 2018). Causal inference methods based on hybrids of backcrossed seedlings or gene modification experiments will be needed to discover the genes/polymorphisms that matter for adaptation. In addition to shedding light on the evolutionary processes at work, this study identified several key oak genes involved in the stomatal behavior as potentially involved in adaptive introgression, providing support for the general view that stomatal regulation makes a major contribution to the maintenance of homeohydr in land plants (Brodribb & McAdam, 2017). Other leaf-unfolding-associated genes would be expected to have key roles, with pleiotropic effects. A typical example is provided by Flowering Locus C (FLC). FLC antagonizes the gibberellin pathway and downregulates flowering (Deng et al., 2011; Li et al., 2015). The effects of FLC are known to be suppressed by vernalization, suggesting that FLC plays a major role in seasonal and developmental timing.

Acknowledgements

This research was funded by the European Research Council under the European Union’s Seventh Framework Programme (TREEPEACE project, FP/2014-2019: ERC Grant Agreement no. 339728) and by the French ANR (GENOAK project, 11-BSV6-009-021). We thank the Genotoul Bioinformatics Platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) and the Biogenouest BiRD core facility (Université de Nantes) for providing computing and storage resources. We thank the staff of the Experimental Units of INRA Nancy (UELF, Unité Expérimentale Forestière de Lorraine) and INRA Toulouse (UE 0393, INRA, Domaine des Jarres, 33210-Toulouse) for their contribution during field phenological assessments and sampling. We also thank Paolo Paiva for providing access to Q. suber data, Nick Zimmerman for providing climate data, Mathieu Gautier for providing advice on how to make the best use of BAYPASS, and Quentin Rougemont for fruitful discussions. TL also thanks the project coordinator Benoit Nabholz (BirdIslandGenomic project, ANR-14-CE02-0002) for his support and feedback. We
are grateful to Nicolas Bierne and to the three anonymous reviewers for their helpful comments.

Author contributions

TL designed and performed the research, analyzed the data, and drafted the manuscript. J-ML and SD were involved in the acquisition of the climate and common garden data. CL performed the DNA extractions and equimolarly pooled DNA. GLP, JML, and CP were involved in the sampling. KL and J-MA generated the sequencing data. CP and AK contributed to the design of the research, interpretation, and drafted the manuscript.

ORCID

Jean-Marc Aury https://orcid.org/0000-0003-1718-3010
Sylvain Delzon https://orcid.org/0000-0003-3442-1711
Karine Labadie https://orcid.org/0000-0001-7467-8509
Thibault Leroy https://orcid.org/0000-0003-2259-9723
Christophe Plomion https://orcid.org/0000-0002-3176-2767

References


Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

**Fig. S1** Climatic and phenological data for the *Quercus petraea* populations studied.

**Fig. S2** Best-fit Trelmix trees for various numbers of migration edges.

**Fig. S3** F3-statistics for admixture between *Q. petraea* and *Q. robur*.

**Fig. S4** Intra- and interspecific differentiation between *Q. robur* and *Q. petraea*.

**Fig. S5** Scans for divergence, GEA and GPA associations for SNPs on chromosome 1.

**Fig. S6** Scans for divergence, GEA and GPA associations for SNPs on chromosome 2.

**Fig. S7** Scans for divergence, GEA and GPA associations for SNPs on chromosome 3.

**Fig. S8** Scans for divergence, GEA and GPA associations for SNPs on chromosome 4.

**Fig. S9** Scans for divergence, GEA and GPA associations for SNPs on chromosome 5.
Fig. S10 Scans for divergence, GEA and GPA associations for SNPs on chromosome 6.

Fig. S11 Scans for divergence, GEA and GPA associations for SNPs on chromosome 7.

Fig. S12 Scans for divergence, GEA and GPA associations for SNPs on chromosome 8.

Fig. S13 Scans for divergence, GEA and GPA associations for SNPs on chromosome 9.

Fig. S14 Scans for divergence, GEA and GPA associations for SNPs on chromosome 10.

Fig. S15 Scans for divergence, GEA and GPA associations for SNPs on chromosome 11.

Fig. S16 Scans for divergence, GEA and GPA associations for SNPs on chromosome 12.

Notes S1 Genetic clines along ecological gradients for GEA and GPA-associated SNPs.

Table S1 Number of SNPs used or outliers identified in this study.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the New Phytologist Central Office.

---

About New Phytologist

- New Phytologist is an electronic (online-only) journal owned by the New Phytologist Trust, a not-for-profit organization dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews and Tansley insights.

- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via Early View – our average time to decision is <26 days. There are no page or colour charges and a PDF version will be provided for each article.

- The journal is available online at Wiley Online Library. Visit www.newphytologist.com to search the articles and register for table of contents email alerts.

- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)

- For submission instructions, subscription and all the latest information visit www.newphytologist.com

See also the Editorial on this article by Plomion & Martin, 226: 943–946.