Extensive recent secondary contacts between four European white oak species

Thibault Leroy, Camille Roux, Laure Villate, Catherine Bodénès, Jonathan Romiguier, Jorge A. P. Paiva, Carole Dossat, Jean-Marc Aury, Christophe Plomion, Antoine Kremer

Published in final edited form as:
New Phytol. 2017 April ; 214(2): 865–878. doi:10.1111/nph.14413.

Extensive recent secondary contacts between four European white oak species

Thibault Leroy[1], **Camille Roux**[2], **Laure Villate**[1], **Catherine Bodénès**[1], and **Jonathan Romiguier**[2]
Jorge A. P. Paiva[3,4], **Carole Dossat**[5], **Jean-Marc Aury**[5], **Christophe Plomion**[1], and **Antoine Kremer**[1]

[1]BIOGECO, INRA, Univ. Bordeaux, 33610 Cestas, France [2]Department of Ecology and Evolution, University of Lausanne, Lausanne 1015, Switzerland [3]iBET, Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2780-901 Oeiras, Portugal [4]Institute of Plant Genetics, Polish Academy of Sciences, 34 Strzeszynska street, PL-60-479 Poznań, Poland [5]Commissariat à l'Energie Atomique (CEA), Institut de Genomique (IG), Genoscope, Evry 91057, France

Summary

- Historical trajectories of tree species during the late quaternary have been well reconstructed through genetic and paleobotanical studies. However many congeneric tree species are interfertile, and the timing and contribution of introgression to species divergence during their evolutionary history remains largely unknown. .
- We quantified past and current gene flow events between four morphologically divergent oak species (*Quercus petraea, Q. robur, Q. pyrenaica, Q. pubescens*), by two independent inference methods: diffusion approximation to the joint frequency spectrum (∂a∂i) and approximate Bayesian computation (ABC). For each pair of species, alternative scenarios of speciation allowing gene flow over different timescales were evaluated.
- Analyses of 3,524 SNPs randomly distributed in the genome, showed that these species evolved in complete isolation for most of their history, but recently came into secondary contact, probably facilitated by the most recent period of postglacial warming.
- We demonstrated that i) there was sufficient genetic differentiation before secondary contact for the accumulation of barriers to gene flow, and ii) current European white oak genomes are a mosaic of genes that have crossed species boundaries and genes impermeable to gene flow.

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 **Twitter accounts:** lab (@Ubiogeco), individual (@TiBoLeroyFr).

Author Contributions

T.L. & C.R. designed and performed the research, analyzed the data and drafted the manuscript. L.V.; C.B. contributed to data analysis and interpretation. J.R. carried out the phylogenetic analyses. J.A.P.P. performed the *Q. suber* sequencing and edited the manuscript. C.D. & J-M.A. reconstructed the reference haploid assembly. C.P. & A.K contributed to the design of the research, interpretation and drafted the manuscript. All authors read and approved the manuscript.

Europe PMC Funders Author Manuscripts

Keywords

Speciation; *Quercus*; secondary contact; interglacial recolonization; reproductive isolation

Europe PMC Funders Author Manuscripts

Introduction

The historical trajectories of tree species during the late quaternary have been well reconstructed through genetic and paleobotanical studies over the last decade (Cheddadi *et al.,* 2005; Hewitt, 2000; Hu *et al.,* 2009). These investigations have retraced the major continental migration pathways occurring during the most recent glacial-postglacial periods. These pathways have been particularly well described for widely distributed European forest tree species, such as Norway spruce (Tollefsrud *et al.,* 2015), Scots pine (Cheddadi *et al.,* 2006), beech (Magri *et al.,* 2006), sessile and pedunculate oaks (Petit *et al.,* 2002a; Brewer *et al.,* 2002) and larch (Wagner *et al.,* 2015). One of the key features that emerged from these studies was the recurrent retraction/expansion of populations during the glacial/interglacial periods that recurrently occurred during the Pleistocene and that molded the extant genetic differentiation. During glacial periods, allopatric divergence between populations was reinforced by isolation in different refugial areas (Iberian Peninsula, Italy and Balkans), whereas recolonization during interglacial periods restored genetic contacts among the expanding populations from the refugia. While this general scenario holds for many tree species, it can also be extended at a larger scale, for related sympatric and hybridizing species. Understanding how related species have maintained their species integrity in the context of the glacial/interglacial remains a challenging and pending issue.

As mentioned, many tree genera (*Abies, Pinus, Fraxinus, Quercus* …) comprise groups of interfertile species that share today the same or proximal habitats. Their ability to intercross with closely phylogenetically related species indicates a wide semi-permeability of boundaries to gene flow (Mallet, 2005). This semi-permeability may result from physical linkage to species barrier loci, reducing introgression rates at some loci. However, it may also result from adaptive introgression, because hybridization and subsequent introgression may contribute to rapid evolutionary change by introducing advantageous mutations into another gene pool (*e.g.* Martin *et al.,* 2006; Whitney *et al.,* 2006; Castric *et al.,* 2008, Lexer & Widmer, 2008; Whitney *et al.,* 2010). A well-known example of species introgression and genome permeability to gene flow in forest trees is oaks. Temperate white oaks form a limited group of species widespread throughout Europe and are known to be interfertile (Lepais *et al.,* 2013). The four oak species, *Quercus petraea, Q. robur, Q. pubescens* and *Q. pyrenaica*, have partially overlapping distributions in Europe. *Q. petraea* and *Q. robur* are found throughout the continent, whereas *Q. pubescens* is mostly located in southern Europe, and *Q. pyrenaica* in the extreme western part of the continent. These species have at least three features making them ideal case studies for retracing historical pathways of species divergence:

i) The recent post-glacial history of temperate European white oak species has been reconstructed from both genetic and historical evidence (Petit *et al.,* 2002a, 2002b and 2002c). The locations of the principal refugia of oak species in Europe (Iberian Peninsula, Italy, the Balkans) have been identified, together with

their postglacial recolonization routes. However, the initial links between postglacial lineages and oak species remain unclear (Petit *et al.,* 2002a and 2002b). Little is known about the sequence of speciation events within the European white oak complex, except the start of the radiation (1-5 million years ago, Hubert *et al.,* 2014).

ii) Interspecific hybridization between these four species has been investigated in detail in mixed stands and it has been suggested that they form a species complex (Lepais *et al.,* 2009; Lepais & Gerber, 2011). These studies also showed that species boundaries were maintained largely through postmating prezygotic barriers, especially between *Q. petraea* and *Q. robur* (Abadie *et al.,* 2012, Bodénès *et al.,* 2016).

iii) The *Q. robur* (pedunculate oak) genome sequence is now available (Plomion *et al.,* 2016). Many other genomic resources are also available for European white oaks: a transcript atlas (Lesur *et al.,* 2011; Lesur *et al.,* 2015), expression profiles (Le Provost *et al.,* 2012; Ueno *et al.,* 2013; Le Provost *et al.,* 2016), a SNP array (Lepoittevin *et al.,* 2015) and genetic linkage maps (Bodénès *et al.,* 2012; Bodénès *et al.,* 2016). Bodénès and coworkers (2012; 2016) observed a remarkably high degree of collinearity between pedunculated and sessile oak genomes.

Although contemporary interspecific gene flow between European white oaks has received considerable attention (Lepais *et al.,* 2009; Lepais & Gerber, 2011, Gailing & Curtu, 2014), the historical occurrence and evolutionary imprint of gene flow remain largely unexplored. It is unknown whether species contacts were disrupted, maintained or reinforced during the Pleistocene, especially during the glacial and interglacial periods. Hence its imprint on extant species divergence counterbalancing divergent selection is still an open debate. By using two different independent inference methods based on a large genomic SNP survey, we addressed these pending issues by raising the following questions: (i) did divergence within the white oak complex occur in sympatry or include periods of strict allopatry? (ii) Were periods of interspecific contacts associated to the climatic history, *i.e.* to glacial/interglacial sequences? (iii) Are current genomic landscapes of differentiation explained solely by demographic factors or also shaped in part by natural selection? Our investigations were conducted in the four white oak species (*Q. petraea, Q. robur, Q. pyrenaica, Q. pubescens*), and we explored the species contacts by considering separately each pair of species. In a more global approach, we completed the pairwise historical perspective by reconstructing the phylogenetic relationships among the four species.

Materials and Methods

Sampling

We used samples from a mixed stand of four white oak species (Q. petraea, *Q. robur, Q. pubescens* and *Q. pyrenaica*) located near Briouant (latitude: 43.2166, longitude: 0.8667, elevation 200 to 400 meters) in South-West France. We sampled eight trees for *Q. petraea,* 66 for *Q. pubescens,* 73 for *Q. robur* and 65 for *Q. pyrenaica*. Additional *Q. petraea* genotypes were also sampled from two mixed oak stands (also containing *Q. robur*) located

Europe PMC Funders Author Manuscripts

less than 15 km apart on the northern slopes of the foothills of the French Pyrenees: 49 trees in Ade (latitude 43.2617, longitude: -0.0098 elevation: 370 m) and 22 trees in lbos (latitude: 43.1489, longitude: -0.0134 elevation: 425 m). All these stands were of natural origin and have been described in previous studies (Lepais & Gerber, 2011).

SNP genotyping and mapping

We genotyped 7913 SNPs in 283 individuals with an Illumina® Infinium Iselect Custom Genotyping Array (Illumina Inc., San Diego, CA, USA). The development of the high-density SNP array and the SNP genotyping methods have been described elsewhere (Lepoittevin *et al.,* 2015). Briefly, SNP array was derived from two different SNP calling projects, both using sequences from *Q. robur* and *Q. petraea* samples. The first one is a set of 1,371 SNPs derived from Sanger sequences across 709 gene fragments. The second set is composed of 6,542 SNPs derived from 454 reads at 5, 112 unigenes (Ueno *et al.,* 2010). Given the relatively high frequency of sequencing errors due to the 454 technology, uncommon alleles (MAF>0.2) were discarded (for details, see Lepoittevin *et al.* 2015). Despite this screening procedure during the SNP discovery step, MAF followed a clear L shape distribution in natural populations, with as expected more lower MAF values in *Q. pyrenaica* and *Q. pubescens* (Lepoittevin *et al.* 2015). Distributions of the SNPs on a high-density composite linkage map (Bodénès et al., 2016) are displayed on Figs Sl-S6.

Outgroup genome sequencing

A *Quercus suber* (cork oak) genotype from Herdade dos Leitões, Montargil, Portugal was used as the outgroup. *Q. suber* belongs to a different, more ancestral group of oaks than the white oaks (Hubert *et al.,* 2014). This genotype was sequenced and orthologous regions to the studied loci were identified. DNA extraction, library preparation and Illumina sequencing are detailed in Methods S1. In total, 96 million Illumina paired-end reads were mapped onto a haploid version of the recently released reference *Q. robur* genome sequence (Plomion *et al.,* 2016). We used the Haplomerger vl (Huang *et al.,* 2012) pipeline to reconstruct allelic relationships in the released diploid assembly and to reconstruct a reference haploid assembly. The diploid genome was first soft-masked with trf (to mask tandem repeats) (Benson, 1999), RepeatMasker (to mask simple repeats, low complexity and Viridiplantae-specific transposable elements) (Smit *et al.,* 2013–2015), DUST (to mask low complexity sequences) (Margulis *et al.,* 2006) and RepeatScout (to mask unknown transposable elements) (Price *et al.,* 2005). We then inferred a scoring matrix specific to the oak genome sequence, using 5% of the diploid assembly. The haploid genome was obtained from the soft-masked assembly and the specific scoring matrix, with the default parameters of Haplomerger. Our read mapping pipeline includes bowtie2 v.2. 1.0 for mapping (Langmead & Salzberg 2012), Picard v.1.88 for removing duplicates (http://picard.sourceforge.net) and SAMtools 1.1 for variant calling (Li *et al.,* 2009). Given the relatively low sequencing depth (12.4x per base, on average, considering bases covered at least once, see also Fig. S7), varFilter was used to discard sites with an extremely high coverage (100x), which are likely to correspond to repetitive regions of the genome.

We checked the quality of this *Q. suber* genotyping, by individually blasting each of the primers described by Lepoittevin *et al.* (2015) against our reference haploid genome. We

Europe PMC Funders Author Manuscripts

retained only SNPs directly flanked by the two flanking primer sequences (no neighboring gaps or SNPs). Only 1,549 SNPs successfully fulfilled this criterion and were used for phylogenetic analysis.

Population genetics analysis

Principal component analysis (PCA) was performed with the FactoMineR package (Lê *et al.,* 2008) implemented in R (R Core Team 2016) to visualize population structure. For admixture analysis, we ran fastSTRUCTURE v1.0 (Raj *et al.,* 2014) with different values for the number of ancestral populations (K), ranging from 1 to 12, with 10 replicates per K value. We used two different strategies to determine the best value of K: the automatic heuristic method implemented in fastSTRUCTURE and the lowest cross-validation error. In all clustering algorithms, the best value of K is the optimal number of groups partitioning the dataset under a given model. This best value of K must be interpreted with caution, because either the sampling ascertainment scheme or violations of the hypothesis to the underlying model can lead to confounding results or erroneous interpretations (for limitations see Schwartz & McKelvey 2009; Kalinowski 2011; Raj *et al.,* 2014). We excluded first-generation hybrids, by retaining only individuals with mean cluster membership (Qvalues) values above 0.9 for all subsequent analyses.

We also calculated summary statistics for diversity and genetic divergence between pairs of species with GENEPOP (Raymond & Rousset 1995). Allele frequencies, and observed (Ho) and expected (He) heterozygosities were estimated for the four white oak species. We used Genetix v.4.05 (Belkhir *et al.,* 1996) to estimate both multilocus and per-locus Weir & Cockerham's F_{ST} values for each pair of species (Weir & Cockerham 1984).

Phylogenetic analysis

Concatenated alignments of all sites were used to infer the maximum likelihood tree with RAxML v8.1.5 (GTR+CAT model, 100 bootstrap replicates) (Stamatakis 2014). The tree was rooted with 1,549 SNPs from the *Quercus suber* outgroup. Tree representations were generated with online Interactive Tree Of Life software (iTOL; Letunic & Bork 2011).

ABC analysis

Simulated models—We used an ABC approach to investigate the demographic history of the four European white oak species (Beaumont *et al.,* 2002). We carried out a statistical evaluation of four models of speciation, for the six possible pairs of species (Fig. 1). All these scenarios included the subdivision of an ancestral panmictic population into two daughter populations (pop1 and pop2) at time T_{SPLIT}. The models assumed that the three diploid populations had independent sizes that remained constant over time (N_{anc}, N_{pop1}, N_{pop2}). Three of the scenarios assumed substantial gene flow since T_{SPLIT}: ancient migration (AM), continuous migration (IM) and secondary contacts (SC). The remaining model assumed strict isolation (SI). In the AM model, migration occurred after T_{SPLIT} but stopped before the present (at time T_{AM}). In the IM model, migration was assumed to have occurred without interruption since T_{SPLIT}. In the SC model, the daughter populations were assumed to have started to evolve in complete isolation early in speciation, with secondary gene flow beginning to occur at time T_{SC}.

Coalescent simulations—For each model, we simulated 10 million data sets with msnsam, a modified version of Hudson's ms program allowing variability of sample size across loci (Hudson 2002; Ross-Ibarra *et al.,* 2008). Simulations were performed with random prior draws from a modified version of priorgen (Ross-Ibarra *et al.,* 2008). For all simulations, summary statistics were calculated with mscalc (Ross-Ibarra *et al.,* 2008; 2009; Roux *et al.,* 2011). All these programs and all input files used are available for download at a GitHub repository (https://github.com/ThibaultLeroyFr/WhiteOaksABC). Large uniform prior distributions were used for all parameters common to all models. The relevance of the chosen prior-model combinations was first evaluated by generating a subset of summary statistics similar to that observed, as proposed by Cornuet *et al.,* (2010). For each model, effective population sizes N_{pop1}, N_{pop2} and N_{anc} were sampled from a uniform distribution running from 0 to 10,000,000. For speciation models assuming homogeneous migration of loci, all loci moving in a given direction had the same effective migration rate, which was independent of that for migration in the opposite direction. M_{pop1} (= $4.N_{pop1}.m_{pop2->pop1}$) and M_{pop2} (= $4.N_{pop2}.m_{pop1->pop2}$) were sampled from uniform distributions (0-100), where $m_{pop2->pop1}$ and $m_{pop1>pop2}$ are the proportions of migrants from population pop2 in population pop1 and of migrants from population pop2 in population pop1, respectively. In our models, special attention was paid to two important genomic features known to bias demographic inferences: genomic heterogeneities in effective migration rates and effective population sizes. For models assuming genomic heterogeneity in effective migration rates, locus-specific effective migration rates were randomly sampled from a beta distribution shaped by parameters “a” and “b” (“a” randomly chosen from 0-100 and “b” from 0-500; Roux *et al.,* 2013). Our SI model assumed genomic variation of effective population size, to take into account the confounding effects of linked selection (Charlesworth *et al.,* 1997; Charlesworth 2009; Cruickshank and Hahn 2014). For this model, a number of loci unlinked to barriers are randomly sampled from a uniform distribution between 0 and the total number of polymorphic loci for the considered pair. For these loci, effective population sizes were sampled from the same uniform distribution than for models assuming genomic homogeneity (0-10,000,000). For the remaining loci, locus-specific effective population sizes were randomly sampled from the beta distribution shaped by parameters “a” and “b”.

Model selection—Posterior probabilities for each of the four models (SI, IM, AM, SC) were estimated with a feedforward neural network by nonlinear multivariate regression, with the model itself considered as an additional parameter to be inferred under the ABC framework, in the R package “abc” (Csillery *et al.,* 2012). The 10,000 replicate simulations (0.025% of the total number) closest to the observed values for the summary statistics were selected and weighted with an Epanechnikov kernel. Computations were performed with 20 trained neural networks and eight hidden networks in the regression.

Finally, we estimated the probability of correctly supporting the best model, as described by Fagunges *et al.,* (2007), by simulating 1,000 pseudo-observed datasets (PODS) under each of the four models compared, with parameters sampled from the same prior distributions. We then used the model selection procedure described above to estimate the relative posterior probabilities of each model for each PODS. The robustness of the demographic

Europe PMC Funders Author Manuscripts

inference for each pair of species was then estimated from distributions of posterior probabilities over PODS.

Parameter estimation—Posterior distribution of the parameters was estimated for the best models using nonlinear regressions. We first used a logit transformation of the parameters on the 10,000 best replicate simulations providing the smallest Euclidian distance (Csillery *et al.,* 2012). Posterior probability were then estimated using the neural network procedure and obtained by means of weighted nonlinear multivariate regressions of the parameters on the summary statistics using 25 trained neural networks and 10 hidden networks.

∂a∂i analysis

Eleven demographic scenarios were compared in a modified version of ∂a∂i, a diffusion approximation method for demographic inference (Gutenkunst *et al.,* 2009; Tine *et al.,* 2014; Le Moan *et al.,* 2016). These models were derived from the four scenarios (SI, AM, IM and SC) compared in the ABC analysis. We evaluated scenarios including genomic homogeneity and heterogeneity in introgression rates between loci and in effective population size. SI, AM, IM and SC describe the same models as investigated by ABC, with genomes displaying similar migration rates and effective population sizes. SI2N describes the SI model with two categories of loci, with reduced or non-reduced effective population sizes, at proportions alpha and 1-alpha in genomes. AM2M, IM2M and SC2M describe the AM, IM and SC models but with two categories of loci in each case: loci linked to species barriers (a proportion P of loci) and loci unlinked to species barriers (a proportion 1-P of loci with $M_A>0$ and $M_B>0$). For these three models, barriers were assumed to be symmetric: $M_{pop1} = M_{pop2} = 0$ for P loci. AM2M2P, IM2M2P and SC2M2P describe the AM2M, IM2M and SC2M models, but with asymmetric barriers. Two independent values of P were used, one for each migration direction (see Methods S2 for details). ∂a∂i estimates the maximum composite likelihood for each model. Two simulated annealing procedures, one cold and the other hot, were successively included before quasi-Newtonian optimization (Tine *et al.,* 2014; Le Moan *et al.,* 2016). For each model, we report the maximum likelihood after 150 independent runs. We then compared the 11 best diffusion fits to the observed joint spectrum of allele frequencies, using the Akaike information criterion (AIC) to take into account heterogeneity in the number of parameters between the 11 models tested (Tine *et al.,* 2014; Le Moan *et al.,* 2016).

Results

Genetic diversity

We genotyped 283 individuals of four different white oak species: *Q. petraea* (n=79), *Q. robur* (n=73), *Q. pubescens* (n=66) and *Q. pyrenaica* (n=65) at 7,913 SNPs (Lepoittevin *et al.,* 2015). Overall, 4,211 SNPs from the initial set of 7,913 SNPs satisfied Illumina quality control criteria, displayed unambiguous patterns on genotyping and were polymorphic in at least one species. No SNP was found to be differentially fixed in any of the six pairs of species. In total, 3,232 SNPs (76.75%) were common to all four species, even if sequences from only two species (*Q. robur* & *Q. petraea*) were used to design the SNP array,

suggesting a substantial fraction of shared polymorphisms between the four species. The rest of the dataset consisted of 167 polymorphisms specific to one species (3.97%), 279 to two species (6.62%) and 533 to three species (12.66%). Mean genomic differentiation, as assessed by F_{ST}, varied from 0.068 (*Q. pubescens* - *Q. pyrenaica*) to 0.131 (*Q. pubescens* - *Q. robur*). For all pairs, standard deviations of F_{ST} values between SNPs were relatively high, ranging from 0.109 to 0.162, indicating high levels of genetic heterogeneity and differentiation within genomes (Figs. S1-S6; S8). Our SNP data was based on SNP detection in a limited number of coding regions (Lepoittevin et al., 2015), and several SNPs were identified in some transcripts. We ensured that inferences about population structure and demographics remained unbiased, by reducing the degree of physical linkage disequilibrium between markers through the random sampling of a single SNP for each of the 3,524 transcripts.

Strong interspecific structure

We investigated population structure, by principal component analysis (PCA) and inference-based methods. In PCA (Fig. 2-A and 2-B), the first component accounted for 11.6% of total inertia and clearly separated *Q. robur* individuals from those of the *Q. petraea, Q. pubescens, Q. pyrenaica* gene pools (Fig. 2-A and 2-B). The second component (5.87%) isolated *Q. petraea* samples from *Q. pubescens* and *Q. pyrenaica*. The third component (3.02%) separated *Q. pubescens* samples from those corresponding to the other three species. The three first components were, therefore, sufficient to detect four distinct genetic clusters, consistent with both the morphological identification of the four species (Viscosi *et al.,* 2009) and a previous species assignment performed with 807 individuals from the mixed stand at Briouant genotyped for 10 SSR loci (Lepais *et al.,* 2009). However, the species assignments for some individuals were different from those previously reported (e.g. gray arrows - Fig. 2-A and 2-B), suggesting that a small number of markers can be insufficient to ensure an accurate species assignment (Neophytou 2014).

An individual-based clustering analysis was carried out in fastSTRUCTURE to infer population structure from the data. The automatic heuristic method implemented in the software suggested that the best partitioning of the data set was achieved with K=5 or K=6. Cross-validation error was observed smallest for K=5, indicating that statistical support was strongest for a model with five clusters. The bar plot obtained for K=5 (Fig. 2-C; Fig. S9) indicated that the individuals belonged to four main clusters (average membership proportions of 0.28, 0.26, 0.24, 0.22 for the four genotypes), with low average membership rates for the fifth cluster ($<1 \times 10^{-4}$), to which none of the individuals was mainly assigned. Clustering patterns at K = 5 revealed strong associations between our four main clusters and prior species assignments based on the genotyping of 10 SSR loci (Fig. 2, Lepais *et al.,* 2009). For some individuals, species assignment was discordant with the results obtained in the previous study. These individuals were reassigned to the genetic clusters identified in this study (K=5, Fig. 2-C). In total, 256 of the 283 individuals had Q-values (estimates of membership coefficient) exceeding 0.9 for membership of one of the four main clusters: 66 *Q. robur*, 70 *Q. petraea*, 55 *Q. pubescens* and 65 *Q. pyrenaica* individuals. The remaining 27 individuals were considered to be admixed. For all subsequent analyses, only the 256 individuals clearly assigned to one of the four main genetic clusters were retained.

Phylogeny

We then investigated the phylogenetic relationships between the four main clusters (Fig. 3). A maximum likelihood tree based on 1,549 SNPs and rooted on a *Q. suber* genotype (see Materials & Methods for details) suggested that individuals of the *Q. mbur* cluster first split from the individuals of the other three genetic clusters. These three clusters then split into two subclades, one corresponding to *Q. pyrenaica* and the other to *Q. petraea* and Q. *pubescens*. The most recent split likely occurred between *Q. petraea* and *Q. pubescens*. With the notable exception of *Q. pubescens* (Fig. S9), the phylogenetic relationships between species were consistent with the population split observed with the increasing number of axes or clusters in our population structure analyses (Fig. 2), suggesting that both population genetic structure and phylogeny analyses yielded reliable conclusions about the sequence of speciation events.

Strong statistical support for secondary contact

We investigated the importance of gene flow during the European white oak radiation, by comparing four models of divergence (Fig. 1) in two independent approaches: ABC (Beaumont 2002) and ∂a∂i (Gutenkunst *et al.,* 2009). Two of the models tested assumed that the sister species were currently isolated (SI and AM models), whereas the other two models assumed ongoing gene flow (IM and SC models).

ABC analyses of the six possible pairs of species systematically provided unambiguous statistical support for the SC model (Table 1). The posterior probabilities of this model always exceeded 98.5%, outperforming the IM (from 0.1 % to 0.7%), AM (from 0.4% to 0.7%) and SI (from 0.1% to 0.3%) models. Using leave-one-out cross validations based on PODS simulated 1,000 times for each model, we estimated a robustness of 1in support of SC in all comparisons (Fig. S10-A and S10-B) meaning that the probability to wrongly accept SC was close to zero.

We then investigated the effects on genomic introgression of putative barriers accumulated during the isolation period preceding secondary contact. We used ABC to compare alternative SC scenarios: the SC-homo scenario assumes genomic homogeneity in introgression rates (Table 2), whereas the SC-hetero scenario assumes variation in introgression rates among loci. For all pairs of species, ABC provided strong statistical support for the SC-hetero scenario, with relative posterior probabilities of 97.1% to 98% (Table 2). Robustness, estimated with 1,000 PODS simulated under the SC-homo and SC-hetero models, exceeded 0.98 for the six pairs studied (Fig. S10-C and S10-D).

In parallel to the ABC analysis, we also investigated demographic history through a diffusion-based approach implemented in ∂a∂i (Gutenkunst *et al.,* 2009; Tine *et al.,* 2014). The maximum composite likelihood was estimated for 11 alternative models. These models included the four main scenarios compared by ABC (Fig. 1; Methods S2) and their derivatives. Some assumed heterogeneity in between-locus variation in introgression rates (AM2M, IM2M, SC2M, AM2M2P, IM2M2P and SC2M2P; Methods S2), and one also assumed variation in genetic drift (SI2N; Methods S2). As in the ABC analysis, the SC model outperformed the other models for all six pairs of species ($\Delta AIC \leq 4$; Table 3). For

each pair of species, the maximum-likelihood models obtained with ∂a∂i were consistent with the observed joint allele frequency spectrum (Fig. S11-S16), confirming that the best-fit models were able to reproduce observations. However, unlike ABC, ∂a∂i did not distinguish between heterogeneous and homogeneous models of introgression (Table 3). This may be due to the different ways in which these two methods deal with heterogeneity. In ABC, we simulated heterogeneities by assuming that introgression rates followed a beta distribution across loci. However, such continuous variation across loci is not compatible with ∂a∂i, for which we simulated heterogeneities by assuming two discrete categories of loci: a proportion *P* of loci with a migration rate equal to 0, and a proportion 1-*P* with a migration rate greater than 0.

Estimated parameters of the SC model

For each pair of species, we examined parameter estimates under the best-fit SC model. Posterior distributions of parameters were built from 10,000 accepted simulations and were found to be poorly differentiated from their prior distributions for some parameters. This was the case for ancestral population size, migration rates and T_{SPLIT} (Table 4 and Fig. S17-S22), suggesting that the data provided little information about these parameters. Conversely, for each pair of species, the posterior probabilities of the timing of secondary contact (T_{SC} and its derivative ratio T_{SC}/T_{SPLIT}) were very different from the corresponding prior distributions. Our estimates thus suggest that all secondary contacts occurred recently [T_{SC}/T_{SPLIT} varying between 0.0027 (95% CI: 0 – 0.0148) for *Q. robur* – *Q. pubescens* to 0.0082 (95% CI: 0 – 0.0561) for *Q. pubescens* – *Q. petraea*].

Discussion

We used a large population genetic survey to investigate the sequence of speciation events in the European temperate white oak complex and the history of genetic contacts between these species. First we provided fine phylogenetic resolution among the four species, enlightening the sequence of speciation events (Fig. 4). Second our data suggested that there had been a long period of strict isolation followed by a recent period of massive secondary contact during the last 1% of the divergence time (0.27% to 0.82%; Fig. 4). Third introgression resulted in a very heterogeneous distribution of migration rates throughout the genome.

It is tempting to insert these results in a time frame for the discussion, given the existing knowledge of the timing of speciation events in the genus *Quercus*. Estimates of divergence ages within the European white oaks are still imprecise and vary between 1 to 5 million years according to the most refined dated phylogeny available today (Hubert *et al.,* 2014). We will consider this time period as the lag within which the earliest split (*Q. robur* vs. the ancestor of the 3 other species) and the most recent split (*Q. pubescens* vs. *Q. petraea*) among the 4 species occurred. Starting from this range of time for the divergence and based on mean inferred values of the ratio T_{SC}/T_{SPLIT} (Table 4), our results suggest that the massive secondary contacts occurred between 2,700 and 13,500 years for the most recent contacts (between *Q.robur* and *Q. pubescens*) and 8,200 to 41,000 for the most ancient contacts between *Q. pubescens* and *Q. petraea*). Taking into account the confidence interval of the relative time of the secondary contacts (Table 4) the figures become 0 to 74,000 for

Q.robur – *Q. pubescens*, and 0 to 280,500 for *Q. pubescens* – *Q. petraea*. While these time frames may seem rather vague, they can still be interpreted in light of the timing of climatic oscillations that occurred during the Pleistocene. It is well know that climatic oscillations were with periods lasting 41 to 100-kyr (Dynesius and Jansson, 2000). These figures suggest that the secondary contacts between our white oaks species most likely occurred either during the ongoing or last (Eemian) interglacial period or the two last glacial periods (Riss and Wurm).

Phylogenetic tree topology in European white oaks

Using a phylogenetic framework, we determined the chronology of speciation events within the European white oak complex. This chronology had long remained unclear, because previous oak phylogenetic investigations were unable to distinguish between these species with a high level of confidence (poor support and short branches, Hubert *et al.,* 2014). We found support for an initial splitting off of the *Q. robur* cluster from the three other clusters (*Q. petraea, Q. pubescens* and *Q. pyrenaica*). This split was also supported by the significant contribution of *Q. robur* to the variance explained by the first axis of the PCA (Fig. 2). Our phylogeny also suggested that *Q. pyrenaica* subsequently split off from the other clusters, with *Q. petraea* and *Q. pubescens* diverging more recently. Earlier phylogenetic reconstruction considering temperate white oaks were unable to resolve the relationship between our 4 species (Hubert *et al.,* 2014; Denk and Grimm, 2010; Bellarosa *et al.,* 2005). However recent genetic survey conducted with genetic markers highlighted closer genetic relationships between *Quercus pubescens* and *Q. petraea* (Rellstab *et al.,* 2016a; Neophytou *et al.* 2015; Curtu *et al.,* 2007; Lepais *et al.,* 2009). These results were also confirmed by phenetic analysis based on leaf morphological features showing that *Q. pubescens* and *Q. petraea* share strong leaf morphological similarities in contrast to *Q. pyrenaica* and *Q. robur* (Viscosi et al., 2009).

Unambiguous signature of recent secondary contacts

Our inferences about the evolutionary history of the European white oak complex revealed clear signals of very recent secondary contacts after a long period of isolation. Models assuming one secondary contact event clearly outperformed all alternative scenarios of speciation. This was true for all pairs of species and independently for both ∂a∂i and ABC-based inferences. Our time frame further suggests that the contacts occurred during the very last glacial or interglacial periods. We contend that contacts were facilitated during interglacial periods, when species migrated northwards from different geographically separated refugial zones, as was clearly shown by phylogeographic studies in oaks during the onset to the current interglacial period (Petit *et al.,* 2002a). But was it only during interglacial periods? During glacial periods, were species genetically isolated while present in the same refugial areas, or because they were restricted to different refugial areas. There is no paleobotanical clue to answer this question, as no clear interspecific differences in pollen morphology allow species identification within the deciduous oaks. *Q. petraea, Q. robur and Q. pubescens* are today present in the three major refugial areas (Iberian Peninsula, Italy and the Balkan Peninsula), while *Q. pyrenaica* is only spread today in the Iberian Peninsula. Hence at least for the three former species, one cannot exclude that they were present in the three refugial areas. However, although paleo oak distribution is limited by the scarcity of

pollen sequences, there is evidence that deciduous oaks were restricted in mountains at mid-altitude sites during glacial periods (Brewer et al., 2002), thus favoring genetic isolation between species (Moracho et al., 2016). One may therefore conclude that secondary contacts were less likely to occur during glacial than during interglacial periods. Finally it is worthwhile recalling that despite our uncertainties about the timing of recent contacts, our results are the first to suggest that the four white oak species were genetically isolated during a rather long period before getting into contact (Fig. 4).

How did secondary contact shape the current pattern of nuclear and cytoplasmic divergence? Assuming that species were separated during glacial periods (either within or between refugial areas), isolation lasted more than 100,000 years thus reinforcing interspecific divergence at both the cytoplasmic and nuclear levels (Kremer *et al.,* 2010). Isolation probably resulted in associations between nuclear and cytoplasmic genomes, as still found in some rare cases in extant populations (Kremer *et al.,* 2002; Petit *et al.,* 2002a). Evidence for genetic divergence between refugia during glacial periods is provided by the strong associations between chloroplast haplotypes and routes of postglacial recolonization (Petit *et al.,* 1997; Petit *et al.,* 2002a). In Europe, postglacial migration was initiated from at least three independent oak refugia located in the Iberian, Italian and Balkan peninsulas in southern Europe (Willis, 1996; Tzedakis *et al.,* 2002). These areas are still characterized by different chloroplast haplotypes. When contacts were restored during postglacial expansion, introgression resulted in cytoplasmic capture of chloroplast genomes and genetic homogeneization among interfertile species, as observed today (Dumolin-Lapègue *et al.,* 1998; Petit *et al.,* 1993; Petit *et al.,* 2004). However introgression was incomplete at the nuclear level, due to mating barriers, and this has resulted in the maintenance of nuclear genetic divergence. As also shown by our results, even a handful of DNA markers captured most of the variation contributing to species divergence today (Bodénès *et al.,* 1997; Lepais *et al.,* 2009; Lepais & Gerber, 2011).

Genomic heterogeneity of species divergence: gene flow vs divergent selection

Our results underpin the imprint of secondary contacts on the extant level of species divergence in European white oaks. However they also suggest that natural selection has molded the genomic landscape of species divergence. Demography alone cannot explain the genomic heterogeneity in introgression rates as we observed. Indeed, using our ABC framework, models incorporating unequal introgression rates among loci due to semipermeable barriers to introgression strongly outperformed models assuming equal levels of gene flow (Table 2). This rejection of the hypothesis of genomic homogeneity is, however, not significantly supported by five out of six ∂a∂i simulations (Table 4). As emphasized by Roux et al. (2014), the higher power of an ABC framework to infer genomic heterogeneity in introgression rates as compared to methods assuming discrete groups of loci can account for this difference. By investigating models with variable introgression rates among loci, we demonstrated that intrinsic and/or ecological selection also play a key role in shaping the genomic landscape of species divergence (see also Rellstab *et al.,* 2016b). Previous study based on controlled pollination experiments reported evidences for both pre- and postzygotic reproductive barriers to gene flow in European white oaks (Abadie *et al.,* 2012; Lepais *et al.,* 2013).

Finally our results suggest that the exploration of demographic events and impacts is a prerequisite step before investigating footprints of divergent selection. For example coalescent population genetics approaches are used to determine the mechanisms of speciation, including the maintenance or absence of interspecific gene flow. Methods assessing the maintenance of gene flow have been little used to date (but see Wang *et al.,* 2014 and Christe *et al.,* 2016 for examples in *Populus*), and divergence scenarios involving secondary contact have, thus, been largely neglected. Pathogenic species provide an emblematic example, in which ecological hypotheses of divergence with gene flow continue to predominate (Leroy *et al.,* 2014), despite clear evidence for secondary contact from studies explicitly testing different demographic scenarios (Gladieux *et al.,* 2011; Lemaire *et al.,* 2016; Leroy *et al.,* 2016). Deciphering the evolutionary history of species is of utmost importance and studies of this kind should be carried out before attempting to evaluate the genomic imprint of natural selection. Indeed, detection methods of genomic imprints due to natural selection based on genome scans try to disentangle the genomic footprints due to selection and neutral processes. Even for unrealistically simple demographic scenarios, the efficiency with which these methods detect loci under selection differs considerably between scenarios (Excoffier *et al.,* 2009; De Mita *et al.,* 2013; Fourcade *et al.,* 2013). This is particularly true for secondary contact, because violations of model assumptions result in a very high rate of false-positives (Lotterhos & Whitlock 2014). One of the key challenges we currently face is, thus, characterization of the genomic footprints predicted for realistic scenarios (Hermisson *et al.,* 2009), in the most favorable way, under the best inferred demographic scenario (Le Moan *et al.,* 2016).

Limitations of our study and future prospects

Although this study outlines a credible scenario for a very recent and extensive secondary contact between European white oak species, our study carries also some limitations. First, while our SNP dataset is appropriate for inferring likely scenarios of divergence, it is not suitable to estimates the dates of different evolutionary events. Indeed, large SNP datasets are powerful enough to decipher simple scenarios of divergence, but present limitations to yield reliable estimates of parameters, at least in the absence of any knowledge of the mutation rate for the species under consideration. Second, recurrent cycles of isolation and contacts occurring during the quaternary climate oscillations cannot be ruled out (Fig. 4). Third, our models assuming that effective population sizes remain constant through time. However, it is not unlikely that climate oscillations may have led to quite large temporal variation in population sizes, a source of variation which is not taken into account in this study and which could have a significant effect on the estimation of parameters. Finally, results of this study are based on pairwise comparisons of species. Simultaneous contacts with the other species- or with an unsampled species are not taken into account. However the congruence of the results across the different pairwise analyses suggests that multiple contacts, if they occurred, were as well recent. To tackle these limitations, more complex demographic scenarios involving more than two populations, multiple cycles of contacts and non-constant population sizes need to be tested. This will be easier to achieve using DNA sequences rather than SNPs, because inferences based on sequences avoid ascertainment bias and are expected to lead to more accurate estimations of parameters such as the dating of the different periods of gene flow or the estimation of the population sizes. Historical

inferences can further be improved by using ancient DNA retrieved from fossil remains. Such allochronic investigations of “ancient” genetic parameters are currently underway and their results will later be compared with our data obtained on extant populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research described here was funded by the European Research Council under the European Union’s Seventh Framework Programme (FP/2014-2019; ERC Grant Agreement no. 339728, TREEPEACE project). The genotyping and the *Q. suber* sequencing were financially supported by the European Union (EVOLTREE Network of Excellence, EU FP7 Agreement no. 016322) and by Fundação para a Ciência e Tecnologia (FCT, Portugal) (PTDC/AGR-GPL/101785/2008), respectively. T.L. and L.V. were also supported by the ANR GENOAK Project (11-BSV6-009-01). J.A.P.P. acknowledges the European Union’s Seventh Framework Programme for research, technological development and demonstration (EU FP7 Agreement no. 621321) and the Polish financial sources for education (2015-2019) allocated to project np (W26/7.PR/2015). We thank Fundação João Lopes Fernandes (Portugal) for providing the *Q. suber* plant material. We also thank the Genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul), the TGCC/CCRT (CEA), France Génomique (ANR-10-INBS-09-08) and the Biogenouest BiRD core facility (Université de Nantes) for providing computing and storage resources.

References

Abadie P, Roussel G, Dencausse B, Bonnet C, Bertocchi E, Louvet J-M, Kremer A, Garnier-Géré P. Strength, diversity and plasticity of postmating reproductive barriers between two hybridizing oak species (*Quercus robur* L. and *Quercus petraea* (Matt) Liebl.). Journal of Evolutionary Biology. 2012; 25:157–173. [PubMed: 22092648]

Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. Genetics. 2002; 162:2025–2035. [PubMed: 12524368]

Belkhir, K., Borsa, P., Chikh, L., Raufaste, N., Bonhomme, F. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171. Université de Montpellier II; Montpellier (France): 1996. Available from: http://www.genetix.univ-montp2.fr/genetix/intro.htm

Bellarosa R, Simeone M, Papini A, Schirone B. Utility of ITS sequence data for phylogenetic reconstruction of Italian *Quercus* spp. Molecular Phylogenetics and Evolution. 2005; 34:355–370. [PubMed: 15619447]

Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research. 1999; 27:573–580. [PubMed: 9862982]

Bodénès C, Chancerel E, Ehrenmann F, Kremer A, Plomion C. High-density linkage mapping and distribution of segregation distortion regions in the oak genome. DNA Research. 2016; 23:115–124. [PubMed: 27013549]

Bodénès C, Chancerel E, Gailing O, Vendramin GG, Bagnoli F, Durand J, Goicoechea PG, Soliani C, Villani F, Mattioni C, et al. Comparative mapping in the Fagaceae and beyond with EST-SSRs. BMC Plant Biology. 2012; 12:153. [PubMed: 22931513]

Bodenes C, Joandet S, Laigret F, Kremer A. Detection of genomic regions differentiating two closely related oak species *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. Heredity. 1997; 78:433–444.

Brewer S, Cheddadi R, de Beaulieu JL, Reille M. The spread of deciduous *Quercus* throughout Europe since the last glacial period. Forest Ecology and Management. 2002; 156:27–48.

Castric V, Bechsgaard J, Schierup MH, Vekemans X. Repeated adaptive introgression at a gene under multiallelic balancing selection. PLoS Genetics. 2008; 4:e1000168. [PubMed: 18769722]

Charlesworth B. Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics. 2009; 10:195–205.

Europe PMC Funders Author Manuscripts

Charlesworth B, Nordborg M, Charlesworth D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetics Research. 1997; 70:155–174.

Cheddadi R, de Beaulieu J-L, Jouzel J, Andrieu-Ponel V, Laurent J-M, Reille M, Raynaud D, Bar-Hen A. Similarity of vegetation dynamics during interglacial periods. Proceedings of the National Academy of Sciences, USA. 2005; 102:13939–13943.

Cheddadi R, Vendramin GG, Litt T, François L, Kageyama M, Lorentz S, Laurent J-M, De Beaulieu J-L, Sadori L, Jost A, et al. Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. Global Ecology and Biogeography. 2006; 15:271–282.

Christe C, Stölting KN, Paris M, Fraïsse C, Bierne N, Lexer C. Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. Molecular Ecology. 2016; in press. doi: 10.1111/mec.13765

Cornuet J-M, Ravigné V, Estoup A. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). BMC Bioinformatics. 2010; 11:401–401. [PubMed: 20667077]

Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology. 2014; 23:3133–3157. [PubMed: 24845075]

Csilléry K, François O, Blum MGB. abc: an R package for approximate Bayesian computation (ABC). Methods in Ecology and Evolution. 2012; 3:475–479.

Curtu AL, Gialing O, Finkeldey R. Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. BMC Evolutionary Biology. 2007; 7:218. [PubMed: 17996115]

De Mita S, Thuillet A-C, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Molecular Ecology. 2013; 22:1383–1399. [PubMed: 23294205]

Denk K, Grimm G. The oaks of western Eurasia: traditional classifications and evidence from two nuclear markers. Taxon. 2010; 59:351–366.

Dumolin-Lapègue S, Pemonge MH, Petit RJ. Association between chloroplast and *mitochondrial lineages in oaks*. Molecular Biology and Evolution. 1998; 15:1321–1331. [PubMed: 9867518]

Dynesius M, Jansson R. Evolutionary consequences of changes in species geographical distributions driven by Milankovitch climate oscillations. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97:9115–9120. [PubMed: 10922067]

Excoffier L, Hofer T, Foll M. Detecting loci under selection in a hierarchically structured population. Heredity. 2009; 103:285–298. [PubMed: 19623208]

Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. Statistical evaluation of alternative models of human evolution. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104:17614–17619. [PubMed: 17978179]

Fourcade Y, Chaput-Bardy A, Secondi J, Fleurant C, Lemaire C. Is local selection so widespread in river organisms? Fractal geometry of river networks leads to high bias in outlier detection. Molecular Ecology. 2013; 22:2065–2073. [PubMed: 23506373]

Gailing O, Curtu AL. Interspecific gene flow and maintenance of species integrity in oaks. Annals of Forest Research. 2014; 57:5–18.

Gladieux P, Vercken E, Fontaine MC, Hood ME, Jonot O, Couloux A, Giraud T. Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. Molecular Biology and Evolution. 2011; 28:459–471. [PubMed: 20837605]

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genetics. 2009; 5:e1000695. [PubMed: 19851460]

Hermisson J. Who believes in whole-genome scans for selection? Heredity. 2009; 103:283–284. [PubMed: 19654610]

Hewitt G. The genetic legacy of the Quaternary ice ages. Nature. 2000; 405:907–913. [PubMed: 10879524]

Hu FS, Hampe A, Petit RJ. Paleoecology meets genetics: deciphering past vegetational dynamics. Frontiers in Ecology and the Environment. 2009; 7:371–379.

Huang S, Chen Z, Huang G, Yu T, Yang P, Li J, Fu Y, Yuan S, Chen S, Xu A. HaploMerger: Reconstructing allelic relationships for polymorphic diploid genome assemblies. Genome Research. 2012; 22:1581–1588. [PubMed: 22555592]

Hubert F, Grimm GW, Jousselin E, Berry V, Franc A, Kremer A. Multiple nuclear genes stabilize the phylogenetic backbone of the genus *Quercus*. Systematics and Biodiversity. 2014; 12:405–423.

Hudson RR. Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics. 2002; 18:337–338. [PubMed: 11847089]

Kalinowski ST. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity. 2011; 106:625–632. [PubMed: 20683484]

Kremer A, Kleinschmit J, Cottrell J, Cundall EP, Deans JD, Ducousso A, König AO, Lowe AJ, Munro RC, Petit RJ, et al. Is there a correlation between chloroplastic and nuclear divergence, or what are the roles of history and selection on genetic diversity in European oaks? Forest Ecology and Management. 2002; 156:75–87.

Kremer, A., Le Corre, RJ., Petit, A., Ducousso, A. Historical and contemporary dynamics of adaptive differentiation in European oaks. Molecular approaches in natural resource Conservation. DeWoodyBickham, J.Michler, C.Nichols, K.Rhodes, G., Woeste, K., editors. Cambridge University Press; New-York: 2010. p. 101-116.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012; 9:357–359. [PubMed: 22388286]

Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software. 2008; 25:1–18.

Le Moan A, Gagnaire P-A, Bonhomme F. Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. Molecular Ecology. 2016; 25:3187–3202. [PubMed: 27027737]

Le Provost G, Lesur I, Lalanne C, Da Silva C, Labadie K, Aury JM, Leple JC, Plomion C. Implication of the suberin pathway in adaptation to waterlogging and hypertrophied lenticels formation in pedunculate oak (*Quercus robur* L.). Tree Physiology. 2016; in press. doi: 10.1093/treephys/tpw056

Le Provost G, Sulmon C, Frigerio JM, Bodénès C, Kremer A, Plomion C. Role of waterlogging-responsive genes in shaping interspecific differentiation between two sympatric oak species. Tree Physiology. 2012; 32:119–134. [PubMed: 22170438]

Lemaire C, De Gracia M, Leroy T, Michalecka M, Lindhard-Pedersen H, Guerin F, Gladieux P, Le Cam B. Emergence of new virulent populations of apple scab from nonagricultural disease reservoirs. New Phytologist. 2016; 209:1220–1229. [PubMed: 26428268]

Lepais O, Gerber S. Reproductive patterns shape introgression dynamics and species succession within the European white oak species complex. Evolution. 2011; 65:156–170. [PubMed: 20722727]

Lepais O, Petit RJ, Guichoux E, Lavabre JE, Alberto F, Kremer A, Gerber S. Species relative abundance and direction of introgression in oaks. Molecular Ecology. 2009; 18:2228–2242. [PubMed: 19302359]

Lepais O, Roussel G, Hubert F, Kremer A, Gerber S. Strength and variability of postmating reproductive isolating barriers between four European white oak species. Tree Genetics & Genomes. 2013; 9:841–853.

Lepoittevin C, Bodénès C, Chancerel E, Villate L, Lang T, Lesur I, Boury C, Ehrenmann F, Zelenica D, Boland A, et al. Single-nucleotide polymorphism discovery and validation in high-density SNP array for genetic analysis in European white oaks. Molecular Ecology Resources. 2015; 15:1446–1459. [PubMed: 25818027]

Leroy T, Caffier V, Celton J-M, Anger N, Durel C-E, Lemaire C, Le Cam B. When virulence originates from nonagricultural hosts: evolutionary and epidemiological consequences of introgressions following secondary contacts in *Venturia inaequalis*. New Phytologist. 2016; 210:1443–1452. [PubMed: 26853715]

Leroy T, Le Cam B, Lemaire C. When virulence originates from non-agricultural hosts: new insights into plant breeding. Infection, Genetics and Evolution. 2014; 27:521–529.

Lesur I, Durand J, Sebastiani F, Gyllenstrand N, Bodénès C, Lascoux M, Kremer A, Vendramin GG, Plomion C. A sample view of the pedunculate oak (*Quercus robur*) genome from the sequencing of hypomethylated and random genomic libraries. Tree Genetics & Genomes. 2011; 7:1277–1285.

Lesur I, Le Provost G, Bento P, Da Silva C, Leplé J-C, Murat F, Ueno S, Bartholomé J, Lalanne C, Ehrenmann F, et al. The oak gene expression atlas: insights into Fagaceae genome evolution and the discovery of genes regulated during bud dormancy release. BMC Genomics. 2015; 16:1–23. [PubMed: 25553907]

Letunic I, Bork P. Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Research. 2011; 39:475–478. [PubMed: 20852262]

Lexer C, Widmer A. The genic view of plant speciation: recent progress and emerging questions. Philosophical Transactions Of The Royal Society B. 2008; 363:3023–3036.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]

Lotterhos KE, Whitlock MC. Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. Molecular Ecology. 2014; 23:2178–2192. [PubMed: 24655127]

Magri D, Vendramin GG, Comps B, Dupanloup I, Geburek T, Gömöry D, Latałowa M, Litt T, Paule L, Roure JM, et al. A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. New Phytologist. 2006; 171:199–221. [PubMed: 16771995]

Mallet J. Hybridization as an invasion of the genome. Trends in Ecology & Evolution. 2005; 20:229–237. [PubMed: 16701374]

Martin NH, Bouck AC, Arnold ML. Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. Genetics. 2006; 172:2481–2489. [PubMed: 16415358]

Moracho E, Moreno G, Jordano P, Hampe A. Unusually limited pollen dispersal and connectivity of Pedunculate oak (Quercus robur) refugial populations at the species' southern range margin. Molecular Ecology. 2016; 25:3319–3331. [PubMed: 27146553]

Morgulis A, Gertz EM, Schäffer AA, Agarwala R. A Fast and Symmetric DUST Implementation to Mask Low-Complexity DNA Sequences. Journal of Computational Biology. 2006; 13:1028–1040. [PubMed: 16796549]

Neophytou C, Gärtner SM, Vargas-Gaete R, Micheils HG. Genetic variation of Centra European oaks: shaped by evolutionary factors and human intervention? Tree Genetics and genomes. 2015; 11:79.

Neophytou C. Bayesian clustering analyses for genetic assignment and study of hybridization in oaks: effects of asymmetric phylogenies and asymmetric sampling schemes. Tree Genetics & Genomes. 2014; 10:273–285.

Petit RJ, Bodénès C, Ducousso A, Roussel G, Kremer A. Hybridization as a mechanism of invasion in oaks. New Phytologist. 2004; 161:151–164.

Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, van Dam B, Deans JD, et al. Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest Ecology and Management. 2002a; 156:49–74.

Petit RJ, Csaikl UM, Bordács S, Burg K, Coart E, Cottrell J, van Dam B, Deans JD, Dumolin-Lapègue S, Fineschi S, et al. Chloroplast DNA variation in European white oaks: Phylogeography and patterns of diversity based on data from over 2600 populations. Forest Ecology and Management. 2002b; 156:5–26.

Petit RJ, Kremer A, Wagner DR. Geographic structure of chloroplast DNA polymorphisms in European oaks. Theoretical and Applied Genetics. 1993; 87:122–128. [PubMed: 24190203]

Petit RJ, Latouche-Hallé C, Pemonge M-H, Kremer A. Chloroplast DNA variation of oaks in France and the influence of forest fragmentation on genetic diversity. Forest Ecology and Management. 2002c; 156:115–129.

Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A. Chloroplast DNA footprints of postglacial recolonization by oaks. Proceedings of the National Academy of Sciences, USA. 1997; 94:9996–10001.

Plomion C, Aury J-M, Amselem J, Alaeitabar T, Barbe V, Belser C, Bergès H, Bodénès C, Boudet N, Boury C, et al. Decoding the oak genome: public release of sequence data, assembly, annotation and publication strategies. Molecular Ecology Resources. 2016; 16:254–265. [PubMed: 25944057]

Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005; 21:i351–i358. [PubMed: 15961478]

R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing; 2016. available from: https://www.R-project.org

Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics. 2014; 197:573–589. [PubMed: 24700103]

Raymond M, Rousset F. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. Journal of Heredity. 1995; 86:248–249.

Rellstab C, Bühler A, Graf R, Folly C, Gugerli F. Using joint multivariate analyses of leaf morphology and molecular-genetic markers for taxon identification in three hybridizing European white oaks (*Quercus* spp.) species. Annals of Forest Science. 2016a; 73:669–679.

Rellstab C, Zoller S, Walthert L, Lesur I, Pluess AR, Graf R, Bodénès C, Sperisen C, Kremer A, Gugerli F. Signatures of local adaptation in candidate genes of oaks (*Quercus* spp.) with respect to present and future climatic conditions. Molecular Ecology. 2016b; in press. doi: 10.1111/mec.13889

Ross-Ibarra J, Tenaillon M, Gaut BS. Historical divergence and gene flow in the genus *Zea*. Genetics. 2009; 181:1399–1413. [PubMed: 19153259]

Ross-Ibarra J, Wright SI, Foxe JP, Kawabe A, DeRose-Wilson L, Gos G, Charlesworth D, Gaut BS. Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. PLoS ONE. 2008; 3:e2411. [PubMed: 18545707]

Roux C, Castric V, Pauwels M, Wright SI, Saumitou-Laprade P, Vekemans X. Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation? PLoS ONE. 2011; 6:e26872. [PubMed: 22069475]

Roux C, Tsagkogeorga G, Bierne N, Galtier N. Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. Molecular Biology and Evolution. 2013; 30:1574–1587. [PubMed: 23564941]

Roux C, Fraïsse C, Castric V, Vekemans X, Pogson GH, Bierne N. Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a Mytilus hybrid zone. Journal of Evolutionary Biology. 27:1662–1675.

Schwartz MK, McKelvey KS. Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. Conservation Genetics. 2008; 10:441–452.

Smit, AF., Hubley, R., Green, P. RepeatMasker open-4.0. 2013. available from: http://www.repeatmasker.org

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30:1312–1313. [PubMed: 24451623]

Tine M, Kuhl H, Gagnaire P-A, Louro B, Desmarais E, Martins RST, Hecht J, Knaust F, Belkhir K, Klages S, et al. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nature Communications. 2014; 5:5770.

Tollefsrud MM, Latatowa M, van der Knaap WO, Brochmann C, Sperisen C. Late Quaternary history of North Eurasian Norway spruce (*Picea abies*) and Siberian spruce (*Picea obovata*) inferred from macrofossils, pollen and cytoplasmic DNA variation. Journal of Biogeography. 2015; 42:1431–1442.

Tzedakis PC, Lawson IT, Frogley MR, Hewitt GM, Preece RC. Buffered tree population changes in a Quaternary refugium: Evolutionary Implications. Science. 2002; 297:2044–2047. [PubMed: 12242441]

Ueno S, Klopp C, Leplé JC, Derory J, Noirot C, Léger V, Prince E, Kremer A, Plomion C, Le Provost G. Transcriptional profiling of bud dormancy induction and release in oak by next-generation sequencing. BMC Genomics. 2013; 14:1–15. [PubMed: 23323973]

Ueno S, Le Provost G, Leger V, Klopp C, Noirot C, Frigerio JM, Salin F, Salse J, Abrouk M, Murat F, Brendel O, et al. Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. BMC Genomics. 2010; 11:650. [PubMed: 21092232]

Viscosi V, Lepais O, Gerber S, Fortini P. Leaf morphological analyses in four European oak species (*Quercus*) and their hybrids: A comparison of traditional and geometric morphometric methods. Plant Biosystems. 2009; 143:564–574.

Wagner S, Litt T, Sánchez-Goñi M-F, Petit RJ. History of *Larix decidua* Mill. (European larch) since 130 ka. Quaternary Science Reviews. 2015; 124:224–247.

Wang J, Kallman T, Liu J, Guo Q, Wu Y, Lin K, Lascoux M. Speciation of two desert poplar species triggered by Pleistocene climatic oscillations. Heredity. 2014; 112:156–164. [PubMed: 24065180]

Weir BS, Cockerham CC. Estimating F-Statistics for the analysis of population structure. Evolution. 1984; 38:1358–1370. [PubMed: 28563791]

Whitney KD, Randell RA, Rieseberg LH. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. The American Naturalist. 2006; 167:794–807.

Whitney KD, Randell RA, Rieseberg LH. Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. New Phytologist. 2010; 187:230–239. [PubMed: 20345635]

Willis KJ. Where did all the flowers go? The fate of temperate European flora during glacial periods. Endeavour. 1996; 20:110–114.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

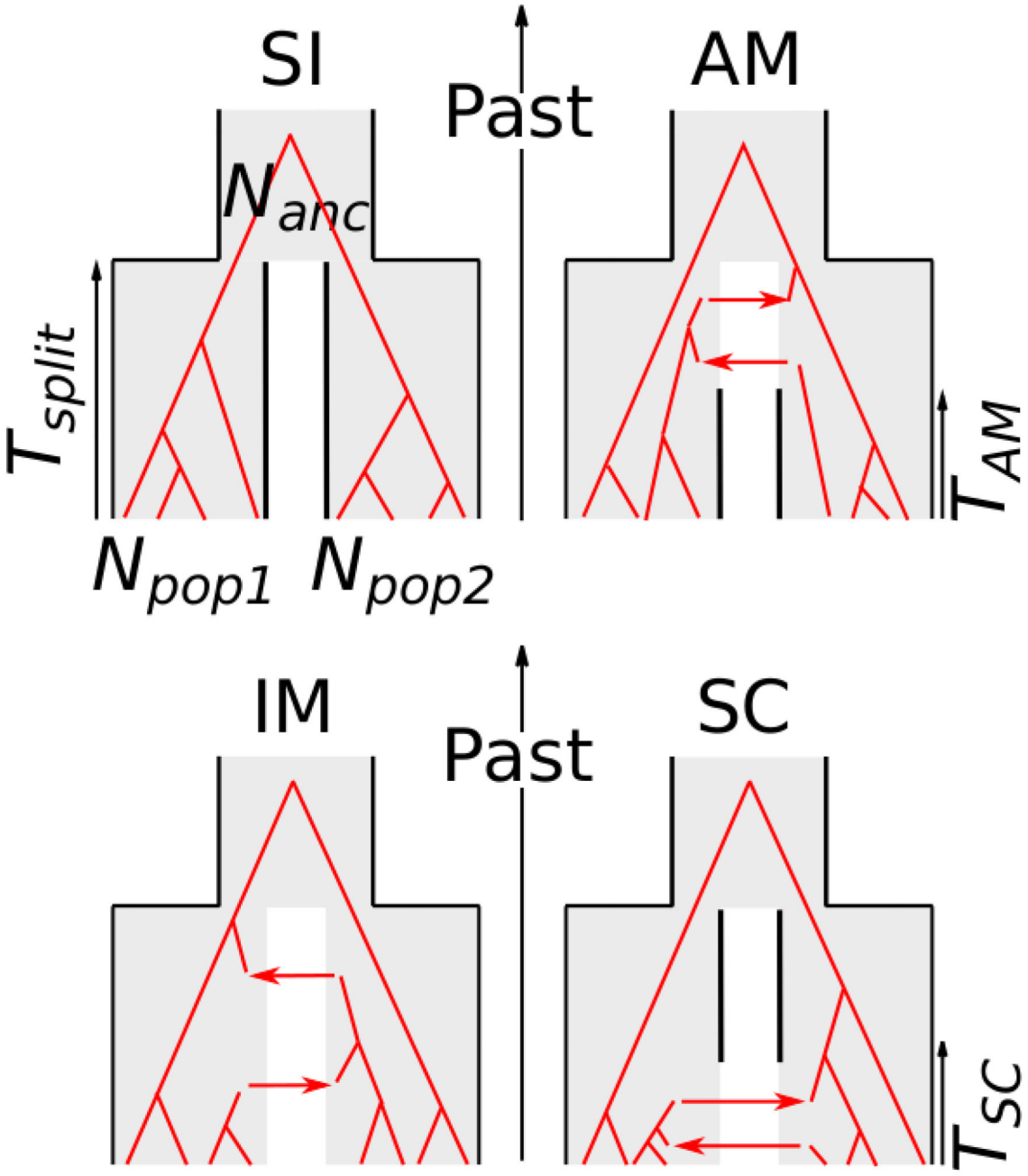

Fig. 1. The speciation scenarios compared

Four classes of models with different temporal patterns of migration were investigated in our ABC analyses: Strict Isolation (SI), Isolation Migration (IM), Ancient Migration (AM), and Secondary Contact (SC). T_{SPLIT} is the number of generations since the divergence time in all models. T_{AM} is the number of generations since the two species stopped exchanging migrants. T_{SC} is the number of generations since the two species began exchanging migrants. N_{anc}, N_{pop1} and N_{pop2} are the effective population sizes in the ancestral and in the two daughter species, respectively.

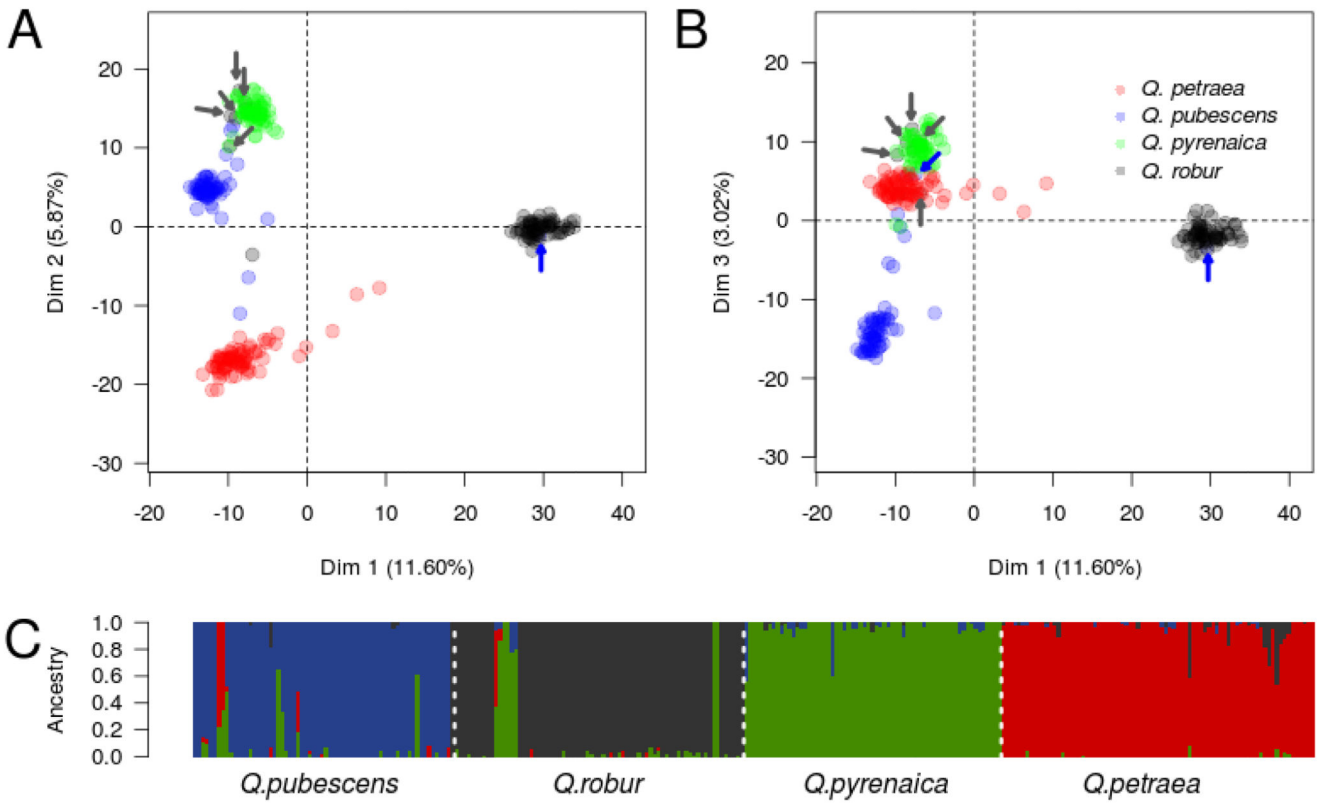

Fig. 2. Genetic structure of oak populations

(A) Principal component analysis (Dim 1: horizontal axis, Dim 2: vertical axis) of the 283 oak individuals represented by dots: *Q. petraea* (red), *Q. pubescens* (blue), *Q. pyrenaica* (green) and *Q. robur* (black). Arrows indicate individuals assigned to a group other than that expected on the basis of the original sampling.

(B) Principal component analysis (Dim 1: horizontal axis, Dim 3: vertical axis).

(C) Individual assignment to *K=5* genetic clusters by fastSTRUCTURE. Vertical lines represent individuals. Colors represent assignment to different genetic clusters. Species labels are based on prior knowledge of species delimitations from a previous population structure analysis (Lepais *et al.* 2009). Note that individuals show very restricted membership to the fifth group (see main text).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

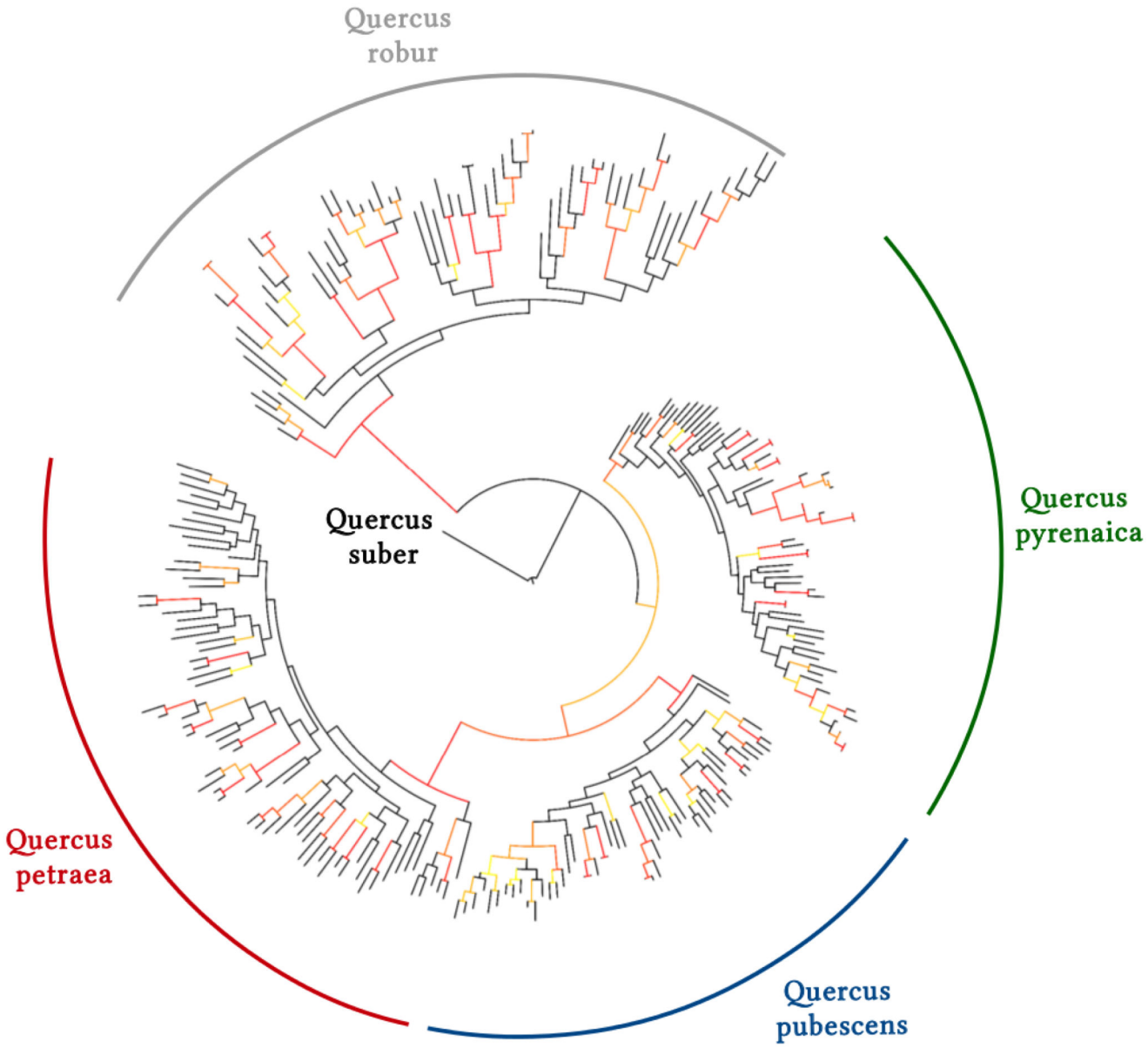

Fig. 3. Phylogenetic relationships of the 256 white oak genotypes rooted on *Q. suber* as the outgroup

Tree branches are colored according to the clade bootstrap value (from yellow (50%) to red (100%); black branches represent nodes with poor bootstrap support <50%). The 4 arc segments highlight the four main genetic clusters previously detected.

Europe PMC Funders Author Manuscripts

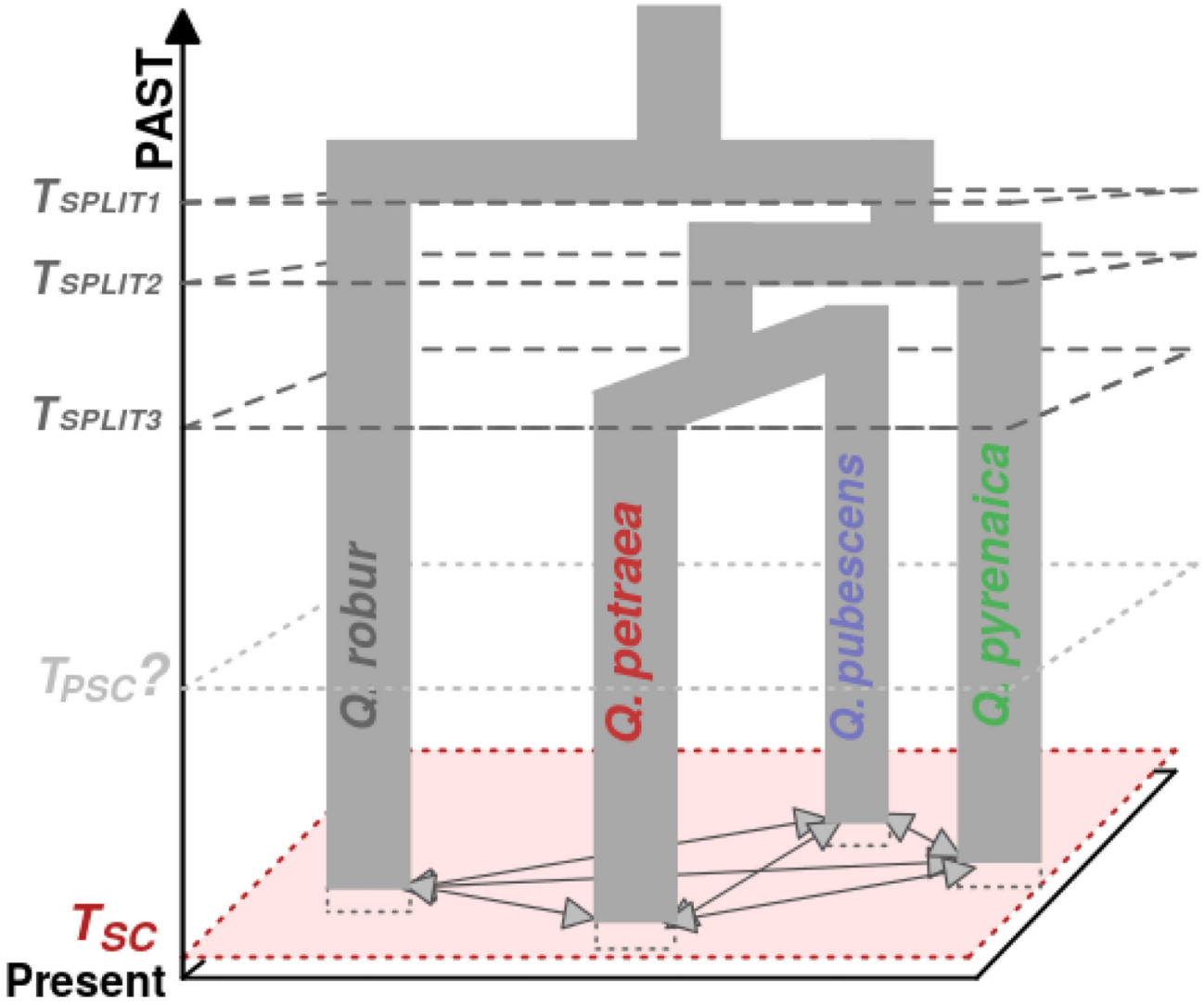

Fig. 4. Schematic drawing of the most likely scenario of divergence for European white oaks.

Red and light grey planes represent last secondary contact and putative older contact respectively. T_{SC}, T_{PSC}, T_{SPLIT1}, T_{SPLIT2} and T_{SPLIT3} represents times for the inferred secondary contact, a putative previous secondary contact, the split of *Q. robur* from the three other cluster, the split of *Q. pyrenaica* from the two last cluster and the split of *Q. petraea* from *Q. pubescens,* respectively. Time is not at scale.

Europe PMC Funders Author Manuscripts

Table 1

Relative posterior probabilities of the compared scenarios, as estimated by ABC.

For each scenario and each pair of species, posterior probabilities were averaged over 100 ABC analyses. Standard deviations are indicated in brackets. The model best supported by ABC is indicated in bold.

	Q. robur – Q. pyrenaica	*Q. robur – Q. pubescens*	*Q. robur – Q. petraea*	*Q. pubescens – Q. pyrenaica*	*Q. pubescens – Q. petraea*	*Q. pyrenaica – Q. petraea*
SI	0.0017 (±0.0035)	0.0031 (±0.0060)	0.0020 (±0.0037)	0.0005 (±0.0010)	0.0005 (±0.0010)	0.0008 (±0.0020)
IM	0.0008 (±0.0024)	0.0068 (±0.0132)	0.0005 (±0.0003)	0.0009 (±0.0016)	0.0051 (±0.0078)	0.0035 (±0.0088)
AM	0.0058 (±0.0108)	0.0054 (±0.0086)	0.0062 (±0.0104)	0.0040 (±0.0008)	0.0047 (±0.0096)	0.0069 (±0.0062)
SC	**0.9917 (±0.0150)**	**0.9847 (±0.0199)**	**0.9913 (±0.0141)**	**0.9946 (±0.0100)**	**0.9898 (±0.0132)**	**0.9889 (±0.0129)**

Europe PMC Funders Author Manuscripts

Table 2

Comparison of models including genomic homogeneity and heterogeneity in effective migration rates.

Mean posterior probabilities for 100 ABC analyses are presented. Standard deviations are indicated in brackets. The model best supported by ABC is indicated in bold.

SC	*Q. robur – Q. pyrenaica*	*Q. robur – Q. pubescens*	*Q. robur – Q. petraea*	*Q. pubescens – Q. pyrenaica*	*Q. pubescens – Q. petraea*	*Q. pyrenaica – Q. petraea*
homogeneity	0.0247 (± 0.0098)	0.0297 (± 0.0117)	0.0273 (± 0.0116)	0.0201 (± 0.0096)	0.0212 (± 0.0093)	0.0192 (± 0.0084)
heterogeneity	**0.9753 (± 0.0098)**	**0.9703 (± 0.0117)**	**0.9727 (± 0.0116)**	**0.9799 (± 0.0096)**	**0.9788 (± 0.0093)**	**0.9808 (± 0.0084)**

Europe PMC Funders Author Manuscripts

Table 3

Maximum likelihood of 11 alternative models of speciation, as estimated by ∂a∂i.

The reported likelihoods are the maximum estimated for 150 independent runs of ∂a∂i per model and per combination of species. Models in bold have the lowest Akaike information criterion (AIC) differences (Δi ≤ 4; see main text for details). SI, IM, AM and SC assume genomic homogeneity in effective population size and migration rates. SI2N assumes genomic heterogeneity in effective population size, with two categories of loci. 2M models assume genomic heterogeneity in introgression rates, with symmetric barriers to gene flow, and 2M2P models assume genomic heterogeneity in introgression rates, with asymmetric barriers to gene flow.

	Q. robur – Q. pyrenaica			***Q. robur – Q. pubescens***			***Q. robur – Q. petraea***			***Q. pub – Q. pyrenaica***			***Q. pubescens – Q. petraea***			***Q. pyreniaca – Q. petraea***		
	-log(L)	**AIC**	**Δ_i**	**-log(L)**	**AIC**	**Δ_i**	**-log(L)**	**AIC**	**Δ_i**	**-log(L)**	**AIC**	**Δ_i**	**-log(L)**	**AIC**	**Δ_i**	**-log(L)**	**AIC**	**Δ_i**
SI	-462.42	930.85	294.28	-455.61	917.22	268.02	-477.06	960.11	297.85	-332.23	670.45	78.11.	-366.41	738.81	128.53	-390.01	786.02	154.74
SI2N	-451.37	912.73	276.16	-439.96	889.91	240.71	-466.41	942.82	280.56	-319.62	649.24	56.90	-353.47	716.94	106.66	-367.09	744.18	112.90
AM	-323.19	658.39	21.82	-323.76	659.53	10.33	-337.14	686.28	24.02	-294.19	600.37	8.03	-305.61	623.21	13.93	-319.43	650.86	19.58
AM2M	-323.88	661.75	25.18	-325.36	664.72	15.52	-336.12	686.24	23.98	-323.17	660.34	68.00	-359.58	733.16	122.88	-313.74	641.49	10.21
AM2M2P	-330.16	676.32	39.75	-343.11	702.22	53.02	-387.37	790.75	128.49	-300.23	616.47	24.13	-366.41	748.81	138.53	-347.83	711.66	80.38
IM	-323.17	656.34	19.77	-323.67	657.35	8.15	-336.84	683.68	21.42	-294.19	598.37	6.03	-305.43	620.86	10.58	-319.42	648.83	17.55
IM2M	-325.45	662.89	26.32	-322.52	657.04	7.84	-335.77	683.53	21.27	-292.91	597.83	5.49	-305.62	623.24	12.96	-313.47	638.94	7.66
IM2M2P	-325.69	665.38	28.81	-397.24	808.24	159.28	-340.11	694.23	31.97	-293.58	601.16	8.82	-306.20	626.41	16.13	-319.08	652.15	20.87
SC	**-312.29**	**636.57**	**0**	**-318.60**	**649.20**	**0**	**-325.13**	**662.26**	**0**	**-290.17**	**592.34**	**0**	**-299.14**	**610.28**	**0**	-319.06	650.13	18.85
SC2M	**-311.82**	**637.65**	**1.08**	**-319.23**	**652.45**	**3.25**	-326.29	666.58	4.32	-292.43	598.85	6.51	-301.38	616.75	6.47	**-308.64**	**631.28**	**0**
SC2M2P	**-312.24**	**640.49**	**3.92**	**-317.88**	**651.76**	**2.56**	**-324.42**	**664.83**	**2.57**	-293.60	603.19	10.85	-299.59	615.18	4.90	-310.45	636.90	5.62

Europe PMC Funders Author Manuscripts

Table 4

ABC-based parameter estimates for the SC-heterogeneous model of speciation

The mean and 95% CI (round brackets) of each parameter are presented. N_{pop1} and N_{pop2} correspond to the posterior of the effective population sizes in species 1 and species 2 respectively, M_{pop1} and M_{pop2} corresponds to the effective migration rates ($M_{pop1} = 4.N_{pop1}.m_{pop2>pop1}$ and $M_{pop2}=4.N_{pop2}.m_{pop1>pop2}$). T_{SPLIT} and T_{SC} correspond to the unscaled times of divergence and secondary contact, respectively). The relative timing of secondary contact is indicated in bold.

	Sp1 : Q. robur ***Sp2 : Q. pyrenaica***	***Sp1 : Q. robur*** ***Sp2 : Q. pubescens***	***Sp1 : Q. robur*** ***Sp2 : Q. petraea***	***Sp1 : Q. pubescens*** ***Sp2 : Q. pyrenaica***	***Sp1 : Q. pubescens*** ***Sp2 : Q. petraea***	***Sp1 : Q. pyrenaica*** ***Sp2 : Q. petraea***
N_{pop1}	3.08×10^6 ($0.06 – 9.24$) $\times 10^6$	2.98×10^6 ($0.10 – 9.08$) $\times 10^6$	2.89×10^6 ($0.06 – 9.15$) $\times 10^6$	2.95×10^6 ($0.26 – 8.33$) $\times 10^6$	2.62×10^6 ($0.04 – 9.10$) $\times 10^6$	2.29×10^6 ($0.13 – 7.91$) $\times 10^6$
N_{pop2}	2.86×10^6 ($0.06 – 9.11$) $\times 10^6$	2.95×10^6 ($0.10 – 9.17$) $\times 10^6$	2.89×10^6 ($0.08 – 9.13$) $\times 10^6$	2.97×10^6 ($0.18 – 8.19$) $\times 10^6$	3.19×10^6 ($0.00 – 9.37$) $\times 10^6$	2.74×10^6 ($0.16 – 8.81$) $\times 10^6$
N_{anc}	5.44×10^6 ($0.38 – 9.79$) $\times 10^6$	5.41×10^6 ($0.36 – 9.78$) $\times 10^6$	5.43×10^6 ($0.38 – 9.80$) $\times 10^6$	5.67×10^6 ($0.42 – 9.81$) $\times 10^6$	5.57×10^6 ($0.38 – 9.81$) $\times 10^6$	5.45×10^6 ($0.41 – 9.80$) $\times 10^6$
M_{pop1}	37.01 (0.39 – 96.40)	37.00 (0.48 – 96.28)	36.03 (0.45 – 96)	37.27 (1.01 – 94.70)	35.17 (0.33 – 95.64)	38.64 (0.76 – 96.06)
M_{pop2}	37.26 (0.39 – 96.53)	37.08 (0.48 – 96.28)	36.08 (0.35 – 95.79)	36.98 (0.87 – 95.71)	40.44 (0.71 – 96.90)	37.28 (0.59 – 95.42)
T_{SPLIT}	42.95 (1.71 – 96.51)	43.85 (2.37 – 96.46)	43.56 (1.87 – 96.59)	39.13 (1.23 – 96.09)	40.99 (1.32 – 96.21)	40.86 (1.33 – 96.79)
T_{sc}	0.15 (0 – 0.96)	0.11 (0 – 0.67)	0.19 (0 – 1.14)	0.14 (0 – 0.71)	0.43 (0– 3.44)	0.14 (0 –0.86)
T_{sc} (%T_{SPLIT})	**0.0037 (0 – 0.0210)**	**0.0027 (0– 0.0148)**	**0.0045 (0 – 0.0255)**	**0.0044 (0 – 0.0199)**	**0.0082 (0 – 0.0561)**	**0.0039 (0 – 0.0205)**