Molecular Ecology (2009) **18**, 2228–2242 doi: 10.1111/j.1365-294X.2009.04137.x

Species relative abundance and direction of introgression in oaks

O. LEPAIS,*†‡ R. J. PETIT,*† E. GUICHOUX,*† J. E. LAVABRE,*†§ F. ALBERTO,*† A. KREMER*† and S. GERBER*†

*INRA, UMR 1202 BIOGECO, 69 route d'Arcachon, F-33612 Cestas cedex, France, †Université de Bordeaux, UMR 1202 BIOGECO, 69 route d'Arcachon, F-33612 Cestas cedex, France

Abstract

Successful hybridisation and subsequent introgression lead to the transfer of genetic material across species boundaries. In this process, species relative abundance can play a significant role. If one species is less abundant than the other, its females will receive many heterospecific gametes, increasing mate-recognition errors and thus hybridisation rate. Moreover, first-generation hybrids will also more likely mate with the more abundant species, leading to asymmetric introgression. These predictions have important fundamental consequences, especially during biological invasions or when a rare species threatened by extinction is surrounded by individuals from a related species. However, experimental tests in nature of the importance of the relative abundance of each species on hybridisation dynamics remain scarce. We assess here the impact of species relative abundance on hybridisation dynamics among four species from the European white oak species complex. A total of 2107 oak trees were genotyped at 10 microsatellite markers and Bayesian clustering methods were used to identify reference trees of each species. We then used these reference trees to simulate purebred and hybrid genotypes to determine optimal threshold for genetic assignment. With this approach, we found widespread evidence of hybridisation between all studied oak species, with high occurrence of hybrids, varying from 11% to 31% according to stand and sampling strategies. This finding suggests that hybridisation is a common phenomenon that plays a significant role in evolution of this oak species complex. In addition, we demonstrate a strong impact of species abundance on both hybridisation rate and introgression directionality.

Keywords: frequency-dependent process, genetic assignment, hybridisation, microsatellites, *Quercus*, species delimitation

Received 10 November 2008; revision received 16 January 2009; accepted 26 January 2009

Introduction

Interspecific mating associates heterogeneous genomes, giving rise to new allelic combinations (Rieseberg & Carney 1998). When hybridisation is successful, first-generation hybrids may mate with parental species, producing back-crossed individuals. This leads to gene introgression with transfer of genetic material across species boundaries (Anderson 1949; Martinsen *et al*. 2001; Kim *et al*. 2008). Hybridisation and introgression imply some contact between species so that mating can occur. It has long been argued that local species abundance will impact hybridisation dynamics (Hubbs 1955; Mayr 1963). The rationale is that in species where females exert male choice through prezygotic isolation, hybridisation rate will increase when species relative abundances become sharply unbalanced, because the females belonging to the rare species then receive too many heterospecific gametes and are more likely to make mate-recognition errors (Wirtz 1999; Chan *et al*. 2006). Such a mechanism, sometimes called Hubbs' principle, has been hypothesised in animals (reviewed by Rhymer & Simberloff 1996; Wirtz 1999) and in plants (reviewed by Rieseberg 1997). Differences

Correspondence: Gerber Sophie, UMR BIOGECO, Genetic Team, 69 route d'Arcachon, F-33612 Cestas, France. Fax: +33(0)557122881; E-mail: sophie@pierroton.inra.fr

‡Present address: School of Biological and Environmental Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK.

§Present address: Estacion Biologica de Donana, CSIC, Integrative Ecology Group, Pabellon del Peru, Avda. M. Luisa S/N, E-41013 Sevilla, Spain.

© 2009 Blackwell Publishing Ltd

in species proportion could have consequences beyond the first hybrid generation. This is because first generation hybrids (F_1) will also be more likely to mate with the more abundant species, producing backcrossed individuals that will be more similar to the common species (Anderson & Hubricht 1938; Rieseberg 1997). The validity of Hubbs' prediction is interesting to check because it has important practical and fundamental consequences. For instance, if the minority species is represented by only few individuals that produce a high proportion of hybrids, the species might become locally extinct, by pollen swamping and dilution of the genome of the rare species, although its genes will persist at least temporarily in hybrid individuals (Levin *et al*. 1996; Rhymer & Simberloff 1996). Another situation where species proportion can be highly unbalanced is when a colonising species spreads in an area already occupied by a related species. In this case, the invading species is initially rare, and matings with the local species are likely. Genetic material of the local species incorporated into the invading species can then reach high frequency as the invading population experiences rapid demographic growth, resulting in asymmetric introgression of neutral genes (Currat *et al*. 2008). Clearly, species relative abundance can have important consequences on hybridisation dynamics, affecting both hybridisation rates and the direction of introgression. Although some researchers have acknowledged the fact that species proportion can play an important role in introgression dynamics, only few have experimentally demonstrated its reality in nature (e.g. Buggs 2007; but see Burgess *et al*. 2005; Prentis *et al*. 2007; Field *et al*. 2008; Zhou *et al*. 2008). Additional empirical surveys addressing this issue with different organisms are therefore needed.

Hybridisation has been intensively studied in the genus *Quercus* (Arnold 2006). In particular, hybridisation and introgression are suspected to play a role in postglacial recolonisation of Europe by oaks (Petit *et al*. 2003). Detailed studies of mating system of the two species involved (*Quercus robur* and *Quercus petraea*) in controlled crosses (Steinhoff 1993; Steinhoff 1998; Kleinschmit & Kleinschmit 2000) or in natural populations (Bacilieri *et al*. 1996; Streiff *et al*. 1999) have shown that prezygotic and postzygotic barriers exist, but few studies have focused on the consequences of species abundance on hybridisation dynamics within this species complex. In one recent study, hybridisation rate between two oak species (*Q. petraea* and *Q. pyrenaica*) seemed unrelated to species relative abundance, but the number of investigated stands was limited (Valbuena-Carabaña *et al*. 2007). While oak species are only weakly genetically differentiated, they present important morphological and ecological differences. In forests where several oak species are found in sympatry, species are often clustered according to their ecological requirements (Bacilieri *et al*. 1995). Thus, relative proportions of oak species are expected to vary between stands as a result of local ecological conditions as well as stand history (including forest management). These species represent therefore a good model to test the hypothesis that species proportion affects hybridisation and introgression.

In this study, we adopted a blind (i.e. no a priori classification) approach (Duminil *et al*. 2006) to assign oaks to species and identify hybrids using microsatellite markers and Bayesian clustering methods. We analysed several populations from the four most common species of the European white oak complex in France. We first applied a clustering analysis to all trees studied and then used the results to identify reference trees of each species. These were used to generate artificial genotypes of known ancestry (pure species, hybrids and backcrosses) to determine objective and optimal thresholds for genetic assignment. We analysed several populations and stands with different species composition. This allowed us to test whether relative species abundance influences hybridisation dynamics in this species complex. The specific aims of this paper are (i) identifying hybrid individuals, (ii) estimating the pattern of hybridisation across species and populations, and (iii) testing the effect of parental species proportions on hybridisation rate and introgression.

Materials and methods

Species description

Four oak species were included in this study: *Quercus robur* L. (pedunculate oak), *Q. petraea* (Matt.) Liebl. (sessile oak), *Q. pubescens* Willd. (pubescent or downy oak) and *Q. pyrenaica* Willd. (Pyrenean or rebollo oak). *Quercus robur* and *Q. petraea* are widely distributed in Europe. *Quercus pyrenaica* is found along the Atlantic coast from Morocco and northwestern Spain to western France. *Quercus pubescens* is localised around the Mediterranean Basin with a northern latitudinal limit up to 50 degrees. Distribution range and local species presence are governed by climatic and edaphic factors (Rameau *et al*. 1989). In brief, *Q. pubescens* grows on limestones and in thermophilous stations, whereas *Q. pyrenaica* prefers sandy acidic soils. *Q. robur* is found on rich and deep soils and can support flooding, unlike the other oak species, while *Q. petraea* is found on poorer and dryer soils. Whereas the other three oak species are postpioneer species capable of colonising open land, *Q. petraea* is a late-successional species that grows in stable and well-established forest environment. Thus in the Aurignac region, composed of small forests and woodlands (see below), *Q. petraea* is found in the centre of the stands (Gonzalez *et al*. 2008). The species are traditionally identified during the growing season by examining leaf morphology. *Quercus robur* leaves have short petioles, several secondary veins and their basal parts are typically lobated (Kremer *et al*. 2002). *Quercus petraea* leaves have a longer petiole, no secondary veins and a regular leaf shape. *Quercus*

© 2009 Blackwell Publishing Ltd

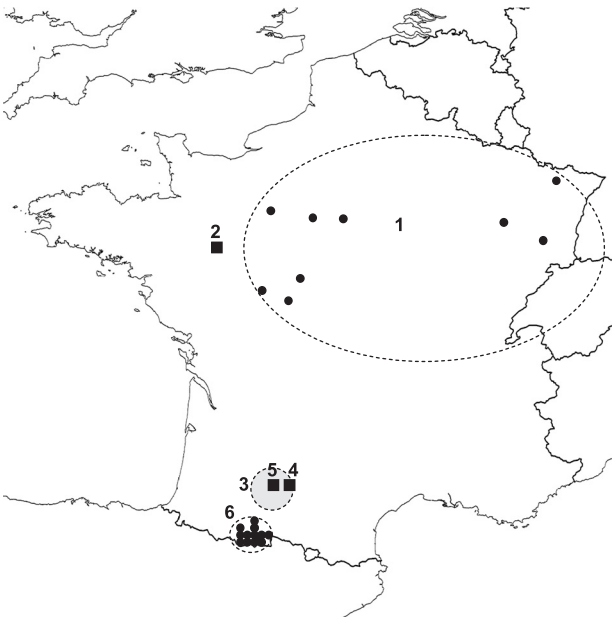

Fig. 1 Location map of the intensively studied stands (squares) and the other sampled populations (encircled) in France (see Table S1 for more details). 1, ONF (National Forest Office) populations; 2, Petite Charnie stand; 3, Aurignac region; 4, Briouant stand; 5, Paguères stand; and 6, Pyrenean populations.

pubescens is similar to *Q. petraea* but the leaves have a higher number of lobes and the abaxial part is densely hairy (Dupouey & Badeau 1993; Curtu *et al*. 2007). *Quercus pyrenaica* leaves are hairy on both sides and have a particular leaf shape with numerous lobes and deep sinuses.

Sampling strategy

A total of 2107 oak trees belonging to the species complex described above were sampled in 53 populations in France (Fig. 1, Table S1, Supporting information). This material had been sampled in the frame of several studies with different objectives, so the sampling strategies are contrasted. The large size of the combined data set should help improve assignment tests (Pritchard *et al*. 2000; Waples & Gaggiotti 2006). In three regions, 10–79 individuals were collected from many populations in France: in the south (Aurignac and Pyrenean stands) and in the north (ONF stands), representing a total of 889 individuals in 50 populations (see Table S1 for more details). In the other areas, stands were more intensively sampled with two stands exhaustively collected, regardless of leaf morphology (Petite Charnie and Briouant) and a third one regularly sampled along a grid (Paguères).

ONF populations consisted in high forests composed mostly of *Q. robur* and *Q. petraea*. Oaks showing typical species morphology were sampled whenever possible. Pyrenean populations were sampled in two valleys at an altitude ranging form 100 to 1600 m. Only *petraea*-like individuals were collected in this study. The Petite Charnie stand has been intensively studied for a long time (Bacilieri *et al*. 1995; Streiff *et al*. 1998; Streiff *et al*. 1999) and only *Q. robur* and *Q. petraea* have been described in this stand, which is part of a continuous high forest. In Aurignac, oak trees showing typical morphology of all three locally abundant oak species (*Q. robur*, *Q. petraea* and *Q. pubescens*) were collected. We sampled one to three individuals by stand (2.5 on average) in 29 forest fragments located within a radius of 30 km around Paguères stand. Briouant and Paguères are two coppice stands localised with Aurignac populations in the long term Ecological Research (LTER-Europe) site 'Vallées et Coteaux de Gascogne'. Paguères includes *Q. robur*, *Q. pubescens* and few *Q. petraea* oaks whereas in Briouant *Q. pyrenaica* is the most frequent species, followed by *Q. robur*, *Q. pubescens* and only few *Q. petraea*.

Two leaves per tree were sampled and kept at 4 °C until stored at –80 °C in the laboratory or immediately dried in silica gel and kept at room temperature. Global positioning system coordinates and morphological species identification using the morphological criteria described above were recorded for each collected tree. Moreover, a detailed morphological analysis was available for the trees from the Petite Charnie (Bacilieri *et al*. 1995) and Briouant (Viscosi *et al*. 2009). Either a discriminant function based on two morphological characters (Kremer *et al*. 2002) was used to distinguish *Q. robur* and *Q. petraea* in the ONF stands or 10 morphological characters were measured to perform a morphological analysis in the case of Pyrenean populations (E. Guichoux, unpublished data and F. Alberto, unpublished data, respectively). When species status was uncertain, oaks were recorded as undetermined species.

Genetic analyses

DNA isolation was performed with a cetyltrimethyl ammonium bromide (CTAB) protocol as previously described (Lepais *et al*. 2006) except for the ONF populations for which the QIAGEN DNeasy Plant Mini Kit was used following the manufacturer's instructions. Ten microsatellite loci selected for their relatively high degree of genetic differentiation between species (Scotti-Saintagne *et al*. 2004; P. G. Goicoechea, unpublished data) were analysed using a multiplex protocol (Lepais *et al*. 2006). Briefly, two polymerase chain reaction were carried out with an MJ Research DNA Engine Tetrad2 thermocycler to amplify the 10 microsatellites: QpZAG110 (Steinkellner *et al*. 1997), QrZAG11, QrZAG112, QrZAG39, QrZAG96, QrZAG7, QrZAG87, QrZAG65, QrZAG5, QrZAG20 (Kampfer *et al*. 1998). Amplified fragments were analysed with an Amersham MegaBace1000 capillary sequencer and individual genotypes were determined with the Fragment Profiler software version 1.2 using the same parameters for all populations.

© 2009 Blackwell Publishing Ltd

Table 1 Number of simulated individuals (rows) assigned to the different species or hybrid classes (columns) and computed efficiency, accuracy and global performance of the assignment method (at the bottom). Correct assignments are highlighted in bold

Simulated/assigned	Rob	Pet	Pyr	Pub	Hyb RobPet	Hyb RobPyr	Hyb RobPub	Hyb PetPyr	Hyb PetPub	Hyb PyrPub	Total
Rob	**996**				3	1					1000
Pet		**988**			7			4	1		1000
Pyr			**992**			3		2		3	1000
Pub				**972**			4		14	10	1000
F_1_RobPet		3			**27**						30
bc_RobPet	19	1			**38**	2					60
bc_PetRob		20			**38**			2			60
F_1_RobPyr			1			**28**				1	30
bc_RobPyr	15				2	**40**	3				60
bc_PyrRob	1		16			**41**		1		1	60
F_1_RobPub				1			**29**				30
bc_RobPub	17				4	1	**38**				60
bc_PubRob				11	1		**44**		1	3	60
F_1_PetPyr			1					**29**			30
bc_PetPyr		16			2	1		**38**	3		60
bc_PyrPet			23			2		**32**		3	60
F_1_PetPub							1		**29**		30
bc_PetPub		19	1		1			4	**35**		60
bc_PubPet				29			2		**27**	2	60
F_1_PyrPub								1	2	**27**	30
bc_PyrPub			16			1		4	1	**38**	60
bc_PubPyr			1	25				1	2	**31**	60
Total	1048	1047	1051	1038	123	120	121	118	115	119	4900
Efficiency (percentage)	99.6	99.8	99.2	97.2	68.7	72.7	74.0	66.0	60.7	64.0	
Accuracy (percentage)	95.0	94.4	94.4	93.6	83.7	90.8	91.7	83.9	79.1	80.7	
Performance (percentage)	94.7	93.2	93.6	91.0	57.5	66.0	67.9	55.4	48.0	51.6	

Hyb, hybrids; F_1, first generation hybrids; bc, backcrosses; Rob, *Q. robur*; Pet, *Q. petraea*; Pub, *Q. pubescens*; Pyr, *Q. pyrenaica*.

Admixture analyses

Bayesian clustering of the genetic data was performed using Structure version 2.1 (Pritchard *et al*. 2000; Falush *et al*. 2003). To determine the optimal number of groups (K), we ran Structure with K varying from 1 to 10, with 10 runs for each K value, to find the K value with the highest posterior probabilities. We also used the ΔK statistics to evaluate the change in likelihood (Evanno *et al*. 2005). Our parameters were 50 000 burn-in periods and 100 000 Markov chain Monte Carlo repetitions after burn-in with admixture and correlated allele models without any prior information. For the most likely number of clusters ($K = 4$), we calculated the average result over 10 runs to get the final admixture analysis.

Hybrid simulation and genetic assignment

For each of the four species, we selected at random 65 individuals that had high probabilities (admixture coefficient, $Q > 0.90$) to belong to each of the four corresponding clusters identified in the admixture analysis. This allowed us to estimate allelic frequencies of the four species. We then simulated pure species and hybrid genotypes using these allele frequencies and the R statistic software (R Development Core Team 2005). We simulated 1000 genotypes for each species, 30 F_1 hybrids and 60 backcrosses for all combinations of possible crosses between each pair of species. The number of simulated hybrids is somewhat arbitrary but reflects the expected hybrid percentage observed in real populations (see Results section). We analysed these simulated data set with the Structure software, with $K = 4$ and the same parameters as before, to test the performance of the software to distinguish between pure species and hybrids, and to determine thresholds to assign individuals to these categories to reach a high correct classification rate. We then assigned individuals with the determined threshold (see Results section) and computed efficiency (the proportion of correctly assigned individual), accuracy (the proportion of true hybrids or purebreds assigned in each hybrid or purebred classes) and overall performance (the product of efficiency and accuracy) of the assignment procedure (Vaha & Primmer 2006).

Distance-based analyses

Using the individual tree assignment results, we computed Cavalli-Sforza and Edwards genetic distances (DS;

© 2009 Blackwell Publishing Ltd

Cavalli-Sforza & Edwards 1967) between each pair of species or hybrid classes in each population (provided there were a minimum of 10 individuals) with the Populations software (Langella 1999). The resulting distance matrix was used to build an unrooted neighbour-joining tree using the R package APE (Analysis of Phylogenetics and Evolution; Paradis *et al.* 2004).

Hybridisation characteristics and direction of introgression

We further analysed the three intensively sampled stands (Briouant, Petite Charnie and Paguères) to characterise introgression between species. We first performed global analyses to check if there was a difference in the contribution of each species to hybridisation. For $K = 4$, each individual is characterised by a vector of four admixture coefficients. In each stand, we defined two groups of individuals: purebred (whatever their species) and hybrids. We then computed the average of each of the four individual admixture coefficient within groups, resulting in a vector of four averaged admixture coefficients for purebred and a vector of averaged admixture coefficients for hybrids. These two vectors characterised the global genetic composition of purebreds and hybrids in each stand. The null expectation was that each species would contribute to the hybrid gene pool in proportion to its abundance in the stand; that is, the global genetic composition of purebreds should be the same as the global genetic composition of hybrids. To test this hypothesis, we compared the differences between averaged admixture coefficients in purebred and in hybrids using a Student *t*-test. We then investigated the effect of species abundance on differences in genetic composition between hybrids and pure categories. We computed the difference between hybrids and purebreds of each averaged admixture coefficients, considered as an estimate of hybrid excess. This measure of hybrid excess was correlated to the corresponding species relative abundance and tested with a linear model using the R package effects (Fox 2003) to estimate the confidence interval of the linear regression.

We then performed a detailed analysis to test for an effect of parental species relative abundance on introgression directionality. In each stand, we grouped hybrid individuals in one of the six plausible hybrid classes (each characterised by their two parental species). We first computed the average admixture coefficient of each hybrid class in each stand. The genetic composition of each hybrid class is characterised by a vector of four averaged admixture coefficients, among them, the two corresponding to the parental species have a high value while the other have a very low value. We then computed parental species relative abundance for each hybrid class (ratio between the number of oaks of the most abundant parental species and the total number of oaks of the two parental species) in each stand

Table 2 Number (and percentage) of pure species and hybrid oaks as assigned by the Structure software in the different studied stands and populations

Population	Sampling strategy	*N*	Rob	Pet	Pub	Pyr	Hyb RobPet	Hyb RobPub	Hyb RobPyr	Hyb PetPub	Hyb PetPyr	Hyb PubPyr	Total species	Total hybrids
Briouant	Exhaustive, stand	807	240 (29.7%)	3 (0.4%)	83 (10.3%)	235 (29.1%)	28 (3.5%)	35 (4.3%)	48 (5.9%)	7 (0.9%)	40 (5.0%)	88 (10.9%)	561 (69.5%)	246 (30.5%)
Petite Charnie	Exhaustive, stand	262	128 (48.9%)	84 (32.1%)	–	–	15 (5.7%)	11 (4.2%)	12 (4.6%)	7 (2.7%)	5 (1.9%)	–	212 (80.9%)	50 (19.1%)
Paguères	Partial grid-based, stand	149	87 (58.4%)	1 (0.7%)	28 (18.8%)	–	12 (8.1%)	16 (10.7%)	–	2 (1.3%)	–	3 (2.0%)	116 (77.9%)	33 (22.1%)
Aurignac	Partial, 29 populations	75	24 (32.0%)	14 (18.7%)	29 (38.7%)	–	4 (5.3%)	1 (1.3%)	1 (1.3%)	1 (1.3%)	–	1 (1.3%)	67 (89.3%)	8 (10.7%)
Pyrenees	Partial, 12 populations	288	1 (0.3%)	223 (77.4%)	4 (1.4%)	–	31 (10.8%)	3 (1.0%)	1 (0.3%)	9 (3.1%)	15 (5.2%)	1 (0.3%)	228 (79.2%)	60 (20.8%)
ONF	Partial, 9 populations	526	117 (22.2%)	321 (61.0%)	2 (0.4%)	–	24 (4.6%)	11 (2.1%)	4 (0.8%)	26 (4.9%)	21 (4.0%)	–	440 (83.7%)	86 (16.3%)
Total		2107	597 (28.3%)	646 (30.7%)	146 (6.9%)	235 (11.2%)	114 (5.4%)	77 (3.6%)	66 (3.1%)	52 (2.5%)	81 (3.8%)	93 (4.4%)	1624 (77.1%)	483 (22.9%)

N indicates the number of sampled oaks; Hyb, hybrids; Rob, *Q. robur;* Pet, *Q. petraea;* Pub, *Q. pubescens;* Pyr, *Q. pyrenaica* and all hybrid classes between these species by pairs.

© 2009 Blackwell Publishing Ltd

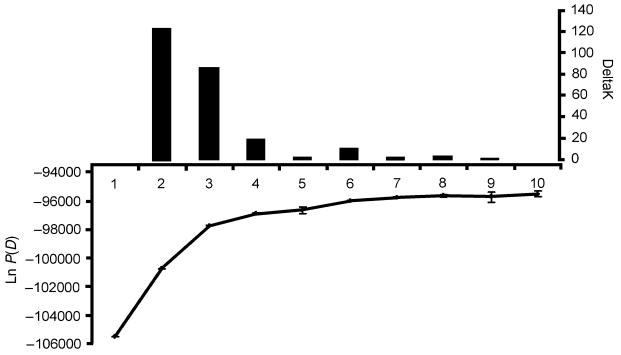

Fig. 2 Estimated number of populations (*K*) derived from the Structure clustering analyses. Mean and standard deviation probabilities of the data over 10 replicated runs (below) and Δ*K* (above) are plotted as a function of the number of clusters (*K* from 1 to 10).

and plotted it against the averaged admixture coefficient of hybrid class that corresponds to the most abundant parental species. If hybridisation is strictly bidirectional or restricted to the first generation (F_1 only), one would expect that the hybrids have an average admixture coefficient value of 0.5. However, if hybridisation is not restricted to the first generation and hybrids themselves can reproduce freely with their parental species, one would expect relative parental species abundance to affect hybrid genetic composition, that is, hybrids would be genetically more similar to the more frequent parental species.

Results

Admixture analysis

The likelihood of the partition of the data increased sharply from $K = 1$ to $K = 3$ and then increased only slightly from $K = 3$ to $K = 6$, where it reached a plateau (Fig. 2). The statistics Δ*K* indicates that $K = 2$ corresponds to the optimal number of groups, but the statistics also gives some support for $K = 3$ or even for $K = 4$ or $K = 6$. We thus report admixture results for $K = 2$, $K = 3$, $K = 4$ and $K = 6$ to compare them (Fig. 3). For $K = 2$, one cluster corresponds to *Quercus robur* (green) and the second to the three remaining morphological species. When adding a third cluster ($K = 3$), *Quercus petraea* is grouped into a specific cluster (yellow) while *Quercus pubescens* and *Q. pyrenaica* are grouped together in the third cluster (pink). For $K = 4$, we get different solutions depending on the run. In seven out of the 10 runs, each species is grouped in one cluster ($K = 4$, Fig. 3: *Q. robur* in the green cluster, *Q. petraea* in the yellow, *Q. pubescens* in the blue and *Q. pyrenaica* in the violet). The other solutions for $K = 4$ (not shown) group *Q. pubescens* and *Q. pyrenaica* in the same cluster while partitioning *Q. robur* and *Q. petraea* in three clusters. Finally, for $K = 6$, only one solution was found: *Q. pubescens* and *Q. pyrenaica* were distinguished as before but *Q. robur* and *Q. petraea* occupied two clusters each. This substructure in *Q. petraea* and *Q. robur* follows a north–south trend with one intraspecific cluster (dark green for *Q. robur* and brown for *Q. petraea*) more frequent in the northern populations while the other (light green for *Q. robur* and orange for *Q. petraea*) is more frequent among southern populations. The genetic distances between the intraspecific clusters are 10-fold smaller than the distances between clusters corresponding to different species, giving strength to the $K = 4$ clustering solution (Fig. S1, Supporting information).

Performance of assignment methods

Distribution of admixture coefficients (*Q*) of simulated individuals (Fig. S2, Supporting information) shows that a threshold value of 0.90 allows separating pure species from hybrids (including F_1 and backcrosses) with the lowest misclassification rate. We thus classified each individual with $Q > 0.90$ as pure species and $Q < 0.90$ as hybrids. However, individuals with $Q < 0.90$ for one cluster but $Q < 0.10$ for each of the three remaining clusters (2.1% of simulated individuals) were supposed to have the majority of their

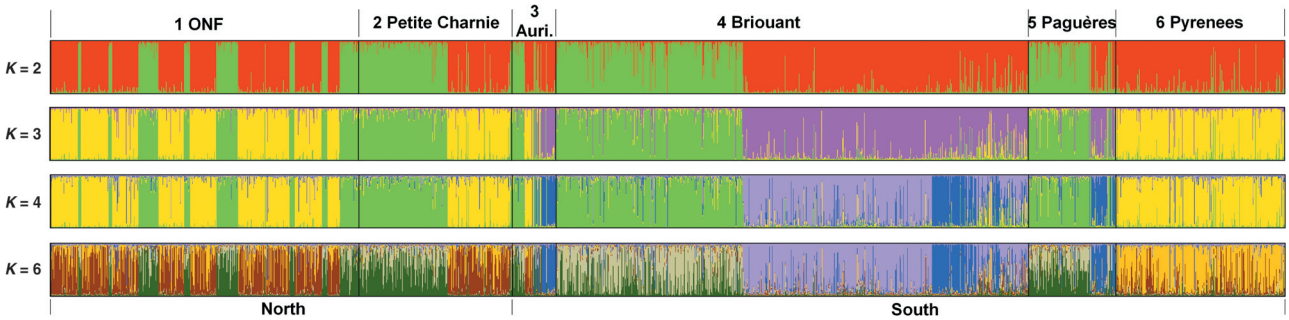

Fig. 3 Structure clustering results obtained for 2, 3, 4 and 6 clusters (*K*). Each individual is represented by a thin vertical line partitioned into *K* coloured segments proportional to its membership in the corresponding genetic cluster. Black lines separate individuals from different populations as indicated at the top, classified according to their latitude, indicated at the bottom. Within populations, individuals are grouped according to their species morphological aspect as determined in the forest (information not used in the clustering analysis).

© 2009 Blackwell Publishing Ltd

genome from one species without any significant influence from other species, and they were thus also classified as pure species (changing this rule did not affect the main conclusions of this work, results not shown). For hybrids, we considered that the two species with the highest assignment probability correspond to the hybrid parental species, whatever the probabilities of the third cluster (i.e. the existence of tri-hybrid individuals was ruled out). Note however, that among assigned hybrids 4.9% show a significant contribution ($Q > 0.10$) from a third cluster. Nevertheless, this assignment strategy provides high efficiency and accuracy (Table 1). The overall performance of the method varies from 94.7% to 91.0% depending on the species. Only 0.4% of pure simulated *Q. robur* individuals are wrongly assigned to a hybrid class but the proportion reaches 2.8% for pure *Q. pubescens*. The overall performance is lower for hybrid identification. The majority of simulated F_1 hybrids are correctly assigned to their hybrid class but simulated first-generation backcrosses often fall into the corresponding pure species category (Table 1, Fig. S2). This results in a decrease in the accuracy of pure species identification and in the efficiency of hybrid assignment, as 32% of these backcrosses are wrongly assigned to a pure species class. However, these wrongly assigned individuals are always classified into their parental species class (the species to which the hybrid is backcrossed). Moreover, 2.7% of F_1 and 6.9% of backcrosses are assigned to another hybrid class. Overall, this strategy should result in a conservative approach to hybrid identification (high accuracy at the expense of a decreased efficiency).

Hybridisation between oak species across populations

We assigned all individuals from natural populations using the method indicated above. Among the 1624 trees assigned to pure species, 226 (14%) showed signs of slight introgression (less than 0.90 probability to belong to their own species but less than 0.10 probability to belong to any other species). Among the 483 assigned hybrids (23%), 96 (20%) have a probability, higher than 0.10, to belong to a third species. Those individuals that escape the strict 0.90 threshold rule are far more numerous than in the case of simulated individuals (2.1% and 4.9% in simulated genotypes, respectively, as described above). This result indicates that in real populations, interspecific crosses may be more complex than the ones modelled in simulations. First, the existence of third-generation or later-generation hybrids could explain the high percentage of slightly introgressed trees in nature. Second, hybridisation involving more than two species seems to happen in natural populations.

Overall, we detected a high occurrence of hybrids in all studied populations (Table 2). The percentage of hybrids was higher in the intensively studied stands (Briouant, Petite Charnie and Paguères), ranging from 19.1% to 30.5% (23.9% on average) compared with 10.7% to 20.8% (15.9% on average) in populations where we sampled a limited number of individuals per stand (Aurignac, Pyrenees and ONF) (Table 2).

We identified hybrids between all pairs of species investigated, in particular in Briouant where the four species co-occur (Table 2). Additionally, we detected a number of hybrids involving a species present in the population and another species not identified during field work. This finding is particularly remarkable in the well characterised Petite Charnie stand where only pedunculate and sessile oaks had been described but where hybrids involving *Q. pubescens* and *Q. pyrenaica* were detected using molecular markers (Table 2). A similar finding was made in populations from the Pyrenees and in the ONF stands where hybrids with *Q. pyrenaica* (not known in these areas) were observed. To test if these results can be explained by assignment error, we used the results from the simulated data set (Table 1). Among 2000 simulated pure *Q. robur* and *Q. petraea* trees, six individuals were wrongly assigned to *Q. pubescens* or *Q. pyrenaica* hybrids (0.3%). Out of 150 simulated *Q. robur* × *Q. petraea* hybrids, we wrongly assigned four trees considered to represent *Q. pubescens* or *Q. pyrenaica* hybrids (3%). Assuming that we only have *Q. robur* and *Q. petraea* species and their hybrids in Petite Charnie, we expect to falsely assign less than one individual from the 212 pure species trees to *Q. pubescens* or *Q. pyrenaica* hybrids and less than 1.5 tree from the 50 hybrids to *Q. pubescens* or *Q. pyrenaica* hybrids (Table 2). Thus in total, if the Petite Charnie stand was only composed by *Q. robur* and *Q. petraea* and their hybrids, we would expect less than three erroneous assignments to *Q. pubescens* or *Q. pyrenaica* hybrids. By contrast, we identified 35 hybrid types involving these species (Table 2), a figure that cannot be explained by assignment errors alone.

Analyses of genetic distances between groups confirmed species and hybrid identification. Pure species oaks identified in each population group together in the same common node (Fig. 4). Furthermore, hybrids involving the same pair of species, whatever their geographical origin, share a common node or are localised in the same part of the tree. This is clearly the case for *Q. robur* × *Q. petraea*, *Q. robur* × *Q. pubescens* and *Q. robur* × *Q. pyrenaica* hybrids (Fig. 4).

Genetic composition of species and hybrids

We computed the average of each of the four admixture coefficients for the two categories (pure species and hybrids). In the three intensively studied stands, the overall genetic composition differed between pure species and hybrids (Fig. 5). The fact that the genetic composition of the pure species category differs from that of the hybrid category indicates that the four species are not involved proportionally in the formation of hybrids and backcrosses. In Petite Charnie, *Q. robur* and *Q. petraea* genes seem to be equally

© 2009 Blackwell Publishing Ltd

Fig. 4 Phylogenetic neighbour-joining tree based on Cavalli-Sforza and Edwards genetic distances (Cavalli-Sforza & Edwards 1967) between pure species and hybrids as assigned by the Structure software in the different populations. Only groups with more than 10 individuals were used to build the tree, the scale line represents a genetic distance of 0.05. Large branches represent pure oak species with colours corresponding to Fig. 3 at $K = 4$. Thinner branches illustrate hybrid groups with each colour corresponding to a specific hybrid type. Labels at the tip of the branches indicate the corresponding species or hybrid type (Rob, *Quercus robur*; Pet, *Q. petraea*; Pub, *Q. pubescens*; Pyr, *Q. pyrenaica*; and hyb, hybrid) and populations' names are given in the subscript (Bri, Briouant stand; PC, Petite Charnie stand; Pag, Paguères stand; Auri, Aurignac populations; ONF, ONF populations; Pyr, Pyrenean populations).

represented in species and hybrid trees but *Q. pubescens* and *Q. pyrenaica* genes are significantly overrepresented among hybrids ($P < 0.001$ and $P < 0.01$, respectively). In Briouant, *Q. robur* genes are far less present in the hybrid category than in the pure species category ($P < 0.001$) whereas *Q. petraea* and *Q. pubescens* genes are significantly more frequent among the hybrid category ($P < 0.001$ and $P < 0.001$, respectively). In Paguères, we also found that *Q. robur* genes are under-represented among hybrid trees ($P < 0.001$), whereas *Q. petraea* and *Q. pyrenaica* genes are over-represented among hybrids ($P < 0.001$ and $P < 0.001$, respectively).

Species frequency-dependent hybridisation and introgression

Differences in genetic composition between hybrids and pure-bred individuals suggest that genes of the more abundant species are under-represented in hybrids (Fig. 5). To formally test this hypothesis, we have plotted the species relative abundance in each stand against the difference in its genetic composition in hybrids vs. purebreds (Fig. 6). There is a clear negative relationship (Fig. 6, $R^2 = 0.83$, $F_{1,10} = 52.86$, $P < 0.001$). This result comforts our observation that abundant species are proportionally less involved in hybridisation

© 2009 Blackwell Publishing Ltd

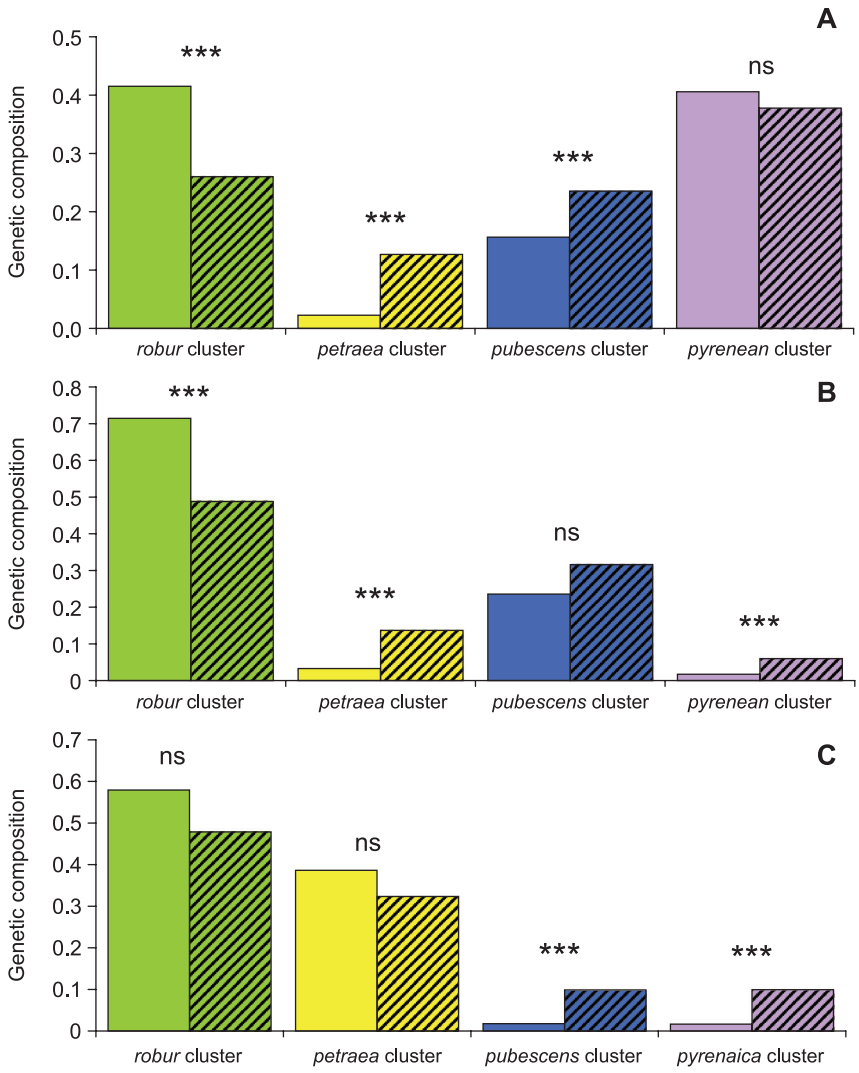

Fig. 5 Comparisons of genetic composition (averaged admixture coefficients from each of the four clusters) for pure species (plain colours) and hybrids (dashed colours) in Briouant (A), Paguères (B) and Petite Charnie (C) stands. Differences were tested with a Student's *t*-test (***: $P < 0.001$, NS: not significant).

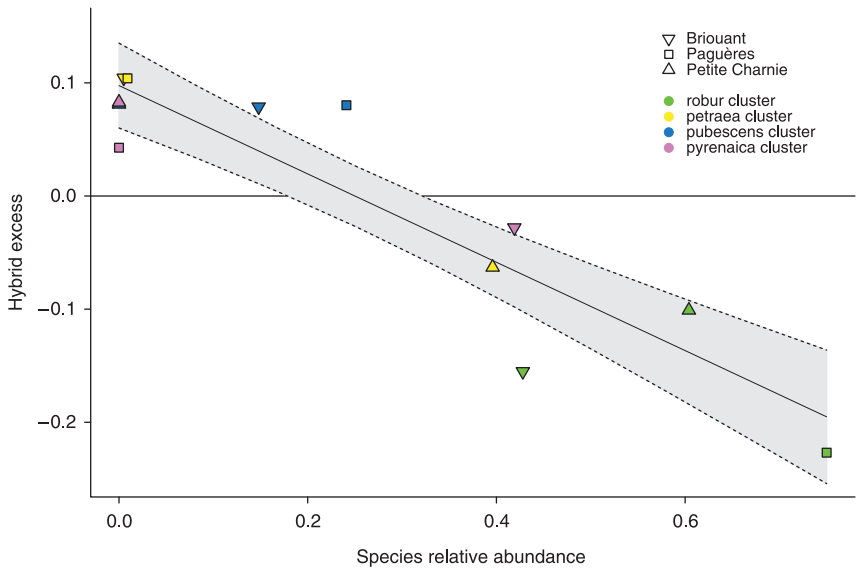

Fig. 6 Change in admixture coefficient between hybrids and purebreds as a function of the corresponding species relative abundance in the stand. The continuous black line indicates no difference between averaged admixture coefficients for hybrids and purebreds. A positive value indicates over-representation of the corresponding cluster in hybrid individuals whereas a negative value indicates over-representation of the corresponding cluster in purebred oaks. Dashed lines and grey shading indicate the confidence interval of the linear regression (large black line; $R^2 = 0.83$, $F_{1,10} = 52.86$, $P < 0.001$). The shapes of the symbols represent the different stands (down-pointing triangle, Briouant; square, Paguères; up-pointing triangle, Petite Charnie) and colours represent clusters (green, *Q. robur* cluster; yellow, *Q. petraea* cluster; blue, *Q. pubescens* cluster; and purple, *Q. pyrenaica* cluster).

© 2009 Blackwell Publishing Ltd

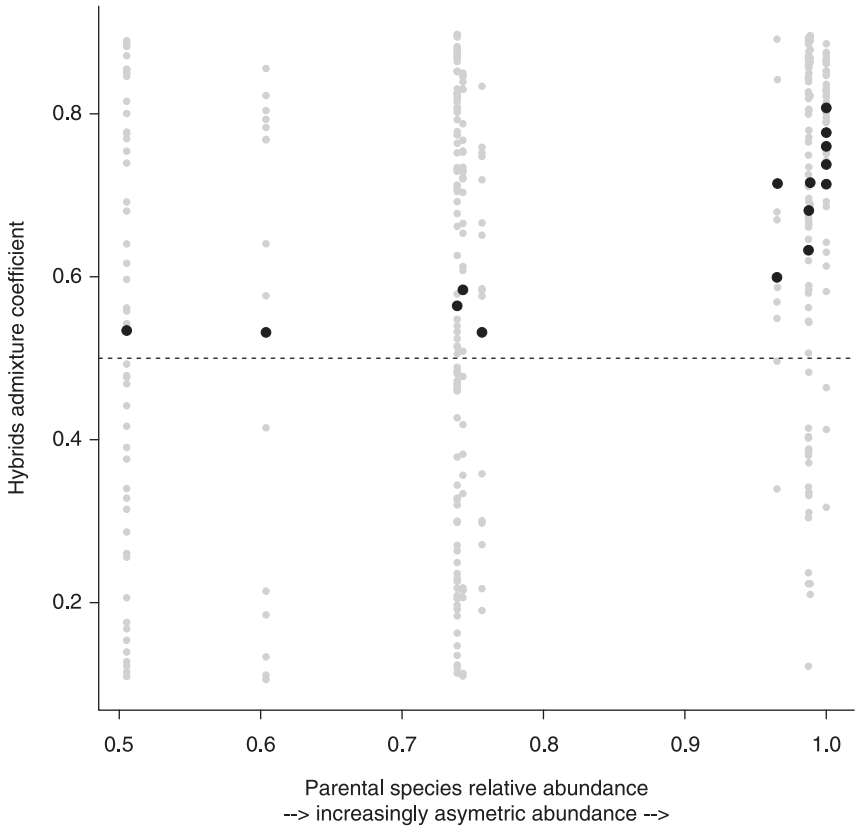

Fig. 7 Effect of parental species relative abundance on hybrid admixture coefficients. Small grey points represent admixture coefficient of each hybrid individuals whereas large black points represent the averaged admixture coefficient for each hybrid class in each stand. For each hybrid class, we used the admixture coefficient corresponding to the most abundant parental species. The horizontal dashed line gives the expected admixture coefficient if introgression was not directional.

than minority species. We then performed a detailed analysis of genetic composition of hybrid classes by using admixture coefficients (Fig. 7). Hybrid individuals admixture coefficients have a large distribution, indicating that hybridisation is not restricted to the first generation (i.e. numerous hybrids had admixture coefficient between 0.65 and 0.9, values that are unlikely for F_1 hybrids, see Fig. S2). Moreover, the averaged admixture coefficient of hybrid classes showed that some classes have an intermediate admixture value, pointing to balanced bidirectional introgression, whereas others hybrid classes have a genetic composition closer to one of the parental species (Fig. 7), indicating directional introgression. Hence, bidirectional introgression seems to take place when parental species are equally represented, whereas directional introgression appears to predominate when parental species differ greatly in abundance (Fig. 7).

Discussion

Our work has addressed the effect of species relative abundance on natural hybridisation and introgression. There are surprisingly few such studies in natural populations. We showed that relative species abundance affects both hybridisation rates and introgression directionality. Previous studies have reported hybridisation patterns between pairs of oak species (Muir *et al*. 2000; Muir & Schlötterer 2005; Valbuena-Carabaña *et al*. 2005, 2007; Gugerli *et al*. 2007) or have studied more species but in one restricted area (Curtu *et al*. 2007). Our extended analyses of 2107 oaks belonging to four species and several populations provide new insights into hybridisation and introgression dynamics within the European white oak species complex. Such large sample sizes should provide accurate estimates of allelic frequencies in the different oak species for use in species delineation and hybrid identification (Waples & Gaggiotti 2006). Using genetic clustering and simulations, we assigned the species or hybrid origin of each sampled oak. We found that hybrids (*sensu lato*: including introgressed individuals) are common in all studied populations, supporting previous claims that hybridisation is ongoing among these oak species (Gugerli *et al*. 2007). Moreover, intensive sampling in three stands allowed us to demonstrate the importance of stand species composition in hybridisation patterns and introgression dynamics.

From clustering to assignment analysis

In the clustering analyses, we found stable results for $K = 6$, highlighting not only differences between species but also a geographical structure within *Quercus robur* and *Quercus petraea*. Such a result might be due to a geographical gradient in allele frequencies, as demonstrated for allozyme data in

© 2009 Blackwell Publishing Ltd

Q. petraea (Zanetto & Kremer 1995; Kremer & Zanetto 1997; Le Corre *et al*. 1998). Using more loci on a wider sampling area covering the distribution range of the species could improve the understanding of these subspecific genetic patterns. In any case, it is clear that intraspecific differences are subsidiary to species differences, and thus intraspecific variation does not compromise species identification. The leaf morphology of a subset of the individuals had been previously analysed (Viscosi *et al*. 2009), showing a clear concordance between genetic cluster and morphological features in these oak species.

We then tested the performance of species assignment and hybrid identification using data-based simulations. Our results show that classes of pure and admixed individuals detected with Structure had been reconstructed with good accuracy and efficiency. However, our 10 microsatellites were not able to differentiate first from second-generation hybrids, an objective that has been shown to require more than 48 loci in cases of low genetic divergence, such as the one observed in these oaks (Vaha & Primmer 2006). Note that our estimates of hybrid abundance are conservative since the threshold we selected ($Q = 0.90$) to distinguish pure species from hybrids should slightly underestimate hybrid proportions and minimise assignment error rate among hybrid classes. Altogether, the results indicate that assignment methods, if used with caution, can be efficient to delimitate species across broad geographical ranges, without prior morphological information, as already shown by Duminil *et al*. (2006). They further indicate that assignments are still relevant when more than two species are present and when an intraspecific geographical structure is detected.

Widespread occurrence of hybrids in the European white oak species complex

Our genetic assignment analysis also confirms that sympatric species from the European white oaks complex do hybridise. Overall hybrid frequencies differ among areas (11–30%, Table 2) with more hybrids detected in intensively sampled stands (19–30%) than in less intensively sampled populations (11–21%). Sampling a small proportion of individuals in a stand can lead to an underestimation of hybridisation if oaks with typical leaf morphology are preferentially sampled. In a detailed multivariate analysis of leaf morphology, hybrid individuals were on average morphologically intermediate between parental species (Viscosi *et al*. 2009). Hence, some (but not all) hybrid oaks could be characterised by an intermediate leaf morphology and intentionally (or not) avoided during sampling (Lexer *et al*. 2006). Estimated hybridisation rates based on non-exhaustive sampling should thus be taken with caution.

The hybrid frequencies found in our populations are comparable with, although slightly higher than, previously found in other studies using comparable approaches. An analysis of three stands in Spain comprising *Q. petraea* and *Q. pyrenaica* detected between 6% and 22% of hybrids depending on the stand (Valbuena-Carabaña *et al*. 2007). Likewise, genetic assignment in a four-oak-species stand in Romania detected between 2% and 16% hybrids depending on the species pairs (Curtu *et al*. 2007). These estimates suggest that hybridisation is not a rare event in oaks and that it is a contemporary process. We were able to identify hybrids between all species pairs studied, indicating that no strict reproductive barriers exist. However, the frequency of the different hybrid classes varies among stands, suggesting that local conditions can affect the outcome of hybridisation. The simultaneous analysis of forests located far apart, with material from all four species included as reference, allowed us to detect hybridisation between species pairs in situations where one of the parental species is locally absent. In the Petite Charnie stand, for instance, only *Q. robur* and *Q. petraea* oaks have been described so far (Bacilieri *et al*. 1995; Streiff *et al*. 1998; Streiff *et al*. 1999) but we identified 13% of *Q. pubescens* and *Q. pyrenaica* hybrid types in this stand (Table 2, Fig. 4), compared with only 6% of *Q. robur* × *Q. petraea* hybrids. This finding highlights the importance of including all species potentially connected by gene flow when studying hybridisation with genetic assignment methods. A separate analysis of the Petite Charnie stand, for example, would have resulted in the detection of only two clusters without any chance to identify *Q. pubescens* and *Q. pyrenaica* hybrids.

The presence of hybrids in the absence of one parental species has also been demonstrated in American red oaks (Dodd & Afzal-Rafii 2004), pinyon pines (Lanner & Phillips 1992) and *Aesculus* tree species (DePamphilis & Wyatt 1989; Thomas *et al*. 2008). Two hypotheses can explain such observations: hybridisation by long-distance pollen dispersal or past local extinction of one of the two parental species (Buggs 2007; Thomas *et al*. 2008). Massive deforestation during the last 3000 years by human exploitation and land clearing for agriculture render difficult to estimate original species distribution ranges and thus the possibility of local extinction of *Q. pubescens* and *Q. pyrenaica* to explain the occurrence of their hybrids. Occasional long-distance hybridisation is not unlikely in these highly outcrossing wind pollinated species. The nearest *Q. pubescens* or *Q. pyrenaica* populations are localised some tens of kilometres from Petite Charnie. Because *Q. pubescens* and *Q. pyrenaica* are more drought tolerant and thermophilous than *Q. robur* and *Q. petraea*, dispersal by long-distance pollen hybridisation could be a mechanism to speed up their northern migration facing climate warming.

Frequency-dependent hybridisation and introgression

Species relative abundance is one of the factors that can affect hybridisation pattern and introgression dynamics

© 2009 Blackwell Publishing Ltd

(Anderson & Hubricht 1938; Nason *et al*. 1992; Burgess *et al*. 2005). Our detailed analysis of three stands differing in species composition allowed us to estimate the relative species abundance and its impact on the outcome of hybridisation.

Hybridisation rate

We found a deficit of hybrids involving locally dominant species (e.g. *Q. robur* and *Q. pyrenaica* in Briouant and *Q. robur* and *Q. petraea* in Petite Charnie), whereas less frequent or rare species tend to be over-represented among hybrids (Figs 5 and 6). Several hypotheses could account for this observation. First, dominant species are expected to be well adapted to local environmental conditions; their hybrids may therefore have a lower competitive ability. Limited hybrid formation between dominant species in a stand would then be caused by differential selection between hybrid and parental species. Second, if these hybrids were selected against, the strength of reproductive barriers between dominant species could increase as a result of reinforcement (Dobzhansky 1937; Butlin 1987). This would lead to a higher reproductive isolation and a lower hybridisation rate between dominant species, compared with species that came more recently in contact, for which reinforcement would not have time to develop. Comparative analyses of open-pollinated progenies with contrasted species abundance situations would be useful to test the hypothesis of reinforcement. Third, rare species could be over-represented among hybrids because of their difficulty to mate with other rare conspecific partners. Such minority species should receive abundant heterospecific pollen, which would increase hybridisation rate (Rieseberg & Gerber 1995). Relative species abundance and underlying causal factors such as local environment and forest management could have a major influence on hybridisation rate. However, this prediction should be tested by manipulating the proportion of pollen from several species received by female flowers using controlled crosses experiments.

Direction of introgression

As we were unable to differentiate F_1 from backcrosses using direct genetic assignment, we computed the mean admixture coefficients of the different hybrid classes in each stand to get some insight into the genetic composition of hybrid individuals compared to their parental species. A mean admixture coefficient of 0.5 would imply that only first-generation hybrids exist or that each parental species mates in the same proportion with hybrids, producing a balanced number of each type of backcrosses. On the contrary, if the backcrosses were biased towards one of the parental species, we should observe a mean cluster value between 0.5 and 0.9 because a majority of the hybrids would be closer to the successfully backcrossing species. Clearly, the observed distribution of individual admixture coefficients in hybrids indicates that backcrosses are more numerous than F_1, as the majority of hybrids showed admixture coefficient between 0.65 and 0.90 (Fig. 7). These results show that hybridisation is not restricted to the formation of F_1 but instead involves further generations of backcrosses between pure species and F_1 hybrids.

Our results show that the direction of introgression strongly depends on the relative frequency of the parental species in the studied stands (Fig. 7). Knowledge of mating system of oak hybrids are lacking, with the exception of one study using controlled crosses on a fertile *Q. robur* × *Q. petraea* hybrid (Olrik & Kjaer 2007). In our study, we found that the direction of the backcrosses was predominantly towards the more numerous species. Additional analyses of hybrid reproductive behaviour would greatly improve our understanding of the hybridisation dynamic in this species complex. However, it is already clear that interspecific gene flow is a widespread and ongoing process among oak species. Since the species remain morphologically and ecologically distinct (Kremer *et al*. 2002; Petit *et al*. 2003), this observation indicates that collective evolution (*sensu* Morjan & Rieseberg 2004) takes place within these species in the face of extensive interspecific gene flow. It would be interesting now to study if collective evolution can simultaneously take place higher in the hierarchy, within groups of closely related species, as first suggested by Pernès (1984). The European white oaks would seem to be good candidates to test this idea, in view of the high rate of interspecific gene flow they experience. In any case, our results indicate that the rate of exchange between species belonging to the same species complex should not be viewed as a fixed parameter but as a variable one that depends on several factors such as the local composition of the community.

Acknowledgements

We thank Jean-Marc Louvet, Jérôme Willm, Maya Gonzalez and Alain Cabanettes for sampling assistance and sharing their field knowledge. We are grateful to Patrick Léger, Valerie Léger, Pierre-Yves Dumolin and Franck Salin for technical assistance. O.L. is grateful to Martin Lascoux for his invitation at the Evolutionary Biology Centre of Uppsala University. We thank Richard Abbott, Alex Buerkle and three anonymous reviewers for their suggestions that greatly improved the manuscript. Genotyping presented in this publication was performed at the Genotyping and Sequencing facility of Bordeaux (grants from the Conseil Régional d'Aquitaine n°20030304002FA, n°20040305003FA and from the European Union, FEDER n°2003227). Experiments were funded by the Interregional Project Aquitaine/Midi-Pyrénées: 'Évolution de la biodiversité des forêts sous l'effet des changements globaux (changements d'usage et changements climatiques)', by the French Research Agency (ANR) through the QDIV project: 'Quantification of the effects of global changes on plant diversity' (n°ANR-05-BDIV-009-01), and

© 2009 Blackwell Publishing Ltd

by the European Union supported project (QLRT-1999-30690) OAKFLOW 'Intra- and interspecific gene flow in oaks as mechanisms promoting genetic diversity and adaptive potential', as well as by the Office National des Forêts ('Traçabilité géographique et identification taxonomique du bois de chêne des forêts domaniales françaises').

References

Anderson E (1949) *Introgressive Hybridization*. Wiley & Sons, New York.

Anderson E, Hubricht L (1938) Hybridization in Tradescantia. III. The evidence for introgressive hybridization. *American Journal of Botany*, **25**, 396–402.

Arnold ML (2006) *Evolution through Genetic Exchange*. Oxford University Press, USA & Oxford, UK.

Bacilieri R, Ducousso A, Kremer A (1995) Genetic, morphological, ecological and phenological differentiation between *Quercus petraea* (Matt) Liebl and *Quercus robur* L. in a mixed stand of Northwest of France. *Silvae Genetica*, **44**, 1–10.

Bacilieri R, Ducousso A, Petit RJ, Kremer A (1996) Mating system and asymmetric hybridization in a mixed stand of European oaks. *Evolution*, **50**, 900–908.

Buggs RJA (2007) Empirical study of hybrid zone movement. *Heredity*, **99**, 301–312.

Burgess KS, Morgan M, Deverno L, Husband BC (2005) Asymmetrical introgression between two *Morus* species (*M. alba*, *M. rubra*) that differ in abundance. *Molecular Ecology*, **14**, 3471–3483.

Butlin R (1987) Speciation by reinforcement. *Trends in Ecology & Evolution*, **2**, 8–13.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Chan C, Ballantyne KN, Aikman H *et al*. (2006) Genetic analysis of interspecific hybridisation in the world's only Forbes' parakeet (*Cyanoramphus forbesi*) natural population. *Conservation Genetics*, **7**, 493–506.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, **62**, 1908–1920.

Curtu AL, Gailing O, Finkeldey R (2007) Evidence for hybridization and introgression whithin a species-rich oak (*Quercus* spp.) community. *BMC Evolutionary Biology*, **7**, 218.

DePamphilis CW, Wyatt R (1989) Hybridization and introgression in buckeyes (*Aesculus*, Hippocastanaceae): a review of the evidence and a hypothesis to explain long-distance gene flow. *Systematic Botany*, **14**, 593–611.

Dobzhansky T (1937) *Genetics and the Origin of Species*. Columbia University Press, New York.

Dodd RS, Afzal-Rafii Z (2004) Selection and dispersal in a multispecies oak hybrid zone. *Evolution*, **58**, 261–269.

Duminil J, Caron H, Scotti I, Cazal SO, Petit RJ (2006) Blind population genetics survey of tropical rainforest trees. *Molecular Ecology*, **15**, 3505–3513.

Dupouey JL, Badeau V (1993) Morphological variability of oaks (*Quercus robur* L., *Quercus petraea* (Matt) Liebl, *Quercus pubescens* Willd) in northern France: preliminary results. *Annales des Sciences Forestières*, **50**, 35s–40s.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Field DL, Ayre DJ, Whelan RJ, Young AG (2008) Relative frequency of sympatric species influences rates of interspecific hybridization, seed production and seedling performance in the uncommon *Eucalyptus aggregata*. *Journal of Ecology*, **96**, 1198–1210.

Fox J (2003) Effect display in R for generalised linear models. *Journal of Statistical Software*, **8**, 1–27.

Gonzalez M, Deconchat M, Balent G, Cabanettes A (2008) Diversity of woody plant seedling banks under closed canopy in fragmented coppice forests. *Annals of Forest Science*, **65**, 511.

Gugerli F, Walser JC, Dounavi K, Holderegger R, Finkeldey R (2007) Coincidence of small-scale spatial discontinuities in leaf morphology and nuclear microsatellite variation of *Quercus petraea* and *Q. robur* in a mixed forest. *Annals of Botany*, **99**, 713–722.

Hubbs CL (1955) Hybridization between fish in nature. *Systematic Zoology*, **4**, 1–20.

Kampfer S, Lexer C, Glossl J, Steinkellner H (1998) Characterization of $(GA)_n$ microsatellite loci from *Quercus robur*. *Hereditas*, **129**, 183–186.

Kim M, Cui M-L, Cubas P *et al*. (2008) Regulatory genes control a key morphological and ecological trait transferred between species. *Science*, **322**, 1116–1119.

Kleinschmit J, Kleinschmit JGR (2000) *Quercus robur — Quercus petraea*: a critical review of the species concept. *Glasnik Za Smske Pokuse*, **37**, 441–452.

Kremer A, Zanetto A (1997) Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. II. Multilocus patterns of variation. *Heredity*, **78**, 476–489.

Kremer A, Dupouey JL, Deans JD *et al*. (2002) Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Annals of Forest Science*, **59**, 777–787.

Langella O (1999) *Populations*, Version 1.2.28. Available from URL: http://www.pge.cnrs-gif.fr/bioinfo/populations/index.php.

Lanner RM, Phillips AM III (1992) Natural hybridization and introgression of pinyon pines in northwestern Arizona. *International Journal of Plant Sciences*, **153**, 250–257.

Le Corre V, Roussel G, Zanetto A, Kremer A (1998) Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. III. Patterns of variation identified by geostatistical analyses. *Heredity*, **80**, 464–473.

Lepais O, Leger V, Gerber S (2006) Short note: high throughput microsatellite genotyping in oak species. *Silvae Genetica*, **55**, 238–240.

Levin DA, Francisco-Ortega J, Jansen RK (1996) Hybridization and the extinction of rare plant species. *Conservation Biology*, **10**, 10–16.

Lexer C, Kremer A, Petit RJ (2006) Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology*, **15**, 2007–2012.

Martinsen GD, Whitham TG, Turek RJ, Keim P (2001) Hybrid populations selectively filter gene introgression between species. *Evolution*, **55**, 1325–1335.

Mayr E (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge, Massachusetts.

Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, **13**, 1341–1356.

Muir G, Schlötterer C (2005) Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci

© 2009 Blackwell Publishing Ltd

differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology*, **14**, 549–561.

Muir G, Fleming CC, Schlötterer C (2000) Species status of hybridizing oaks. *Nature*, **405**, 1016–1016.

Nason JD, Ellstrand NC, Arnold ML (1992) Patterns of hybridization and introgression in populations of oaks, manzanitas and irises. *American Journal of Botany*, **79**, 101–111.

Olrik DC, Kjaer ED (2007) The reproductive success of a *Quercus petraea* × *Q. robur* F_1-hybrid in back-crossing situations. *Annals of Forest Science*, **64**, 37–45.

Paradis E, Claude J, Strimmer K (2004) APE. Analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Pernès J (1984) *Gestion des ressources génétiques. Tome 2: Manuel*. Agence de Coopération Culturelle et Technique, Paris, France.

Petit RJ, Bodenes C, Ducousso A, Roussel G, Kremer A (2003) Hybridization as a mechanism of invasion in oaks. *New Phytologist*, **161**, 151–164.

Prentis PJ, White EM, Radford IJ, Lowe AJ, Clarke AR (2007) Can hybridization cause local extinction: a case for demographic swamping of the Australian native *Senecio pinnatifolius* by the invasive *Senecio madagascariensis*? *New Phytologist*, **176**, 902–912.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Development Core Team (2005) R: a languaage and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from URL: http://www.R-project.org, edn.

Rameau JC, Mansion D, Dumé G (1989) *Flore forestière française: guide écologique illustré, 1: Plaines et collines*. Institut pour le Développement Forestier, Paris, France.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

Rieseberg LH (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systematics*, **28**, 359–389.

Rieseberg LH, Carney SE (1998) Plant hybridization. *New Phytologist*, **140**, 599–624.

Rieseberg LH, Gerber D (1995) Hybridization in the Catalina Island Mountain Mahogany (*Cercocarpus traskiae*). *RAPD Evidence. Conservation Biology*, **9**, 199–203.

Scotti-Saintagne C, Mariette S, Porth I *et al*. (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. Petraea* (Matt.) Liebl.]. *Genetics*, **168**, 1615–1626.

Steinhoff S (1993) Results of species hybridization with *Quercus robur* L. & *Quercus petraea* (Matt.) Liebl. *Annales Des Sciences Forestieres*, **50**, 137s–143s.

Steinhoff S (1998) Controlled crosses between pendunculate and sessile oak: results and conclusion. *Allgemeine Forst und Jagdzeitung*, **169**, 163–168.

Steinkellner H, Fluch S, Turetschek E *et al*. (1997) Identification and characterization of $(GA/CT)_n$-microsatellite loci from *Quercus petraea*. *Plant Molecular Biology*, **33**, 1093–1096.

Streiff R, Labbe T, Bacilieri R *et al*. (1998) Within-population genetic structure in *Quercus robur* L. & *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Molecular Ecology*, **7**, 317–328.

Streiff R, Ducousso A, Lexer C *et al*. (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. Petraea* (Matt.) Liebl. *Molecular Ecology*, **8**, 831–841.

Thomas DT, Ahedor AR, Williams CF *et al*. (2008) Genetic analysis of a broad hybrid zone in *Aesculus* (Sapindaceae): is there evidence of long-distance pollen dispersal? *International Journal of Plant Sciences*, **169**, 647–657.

Vaha JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63–72.

Valbuena-Carabaña M, González-Martínez SC, Sork VL *et al*. (2005) Gene flow and hybridisation in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in central Spain. *Heredity*, **95**, 457–465.

Valbuena-Carabaña M, González-Martínez SC, Hardy OJ, Gil L (2007) Fine-scale spatial genetic structure in mixed oak stands with different levels of hybridization. *Molecular Ecology*, **16**, 1207–1219.

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

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Wirtz P (1999) Mother species-father species: unidirectional hybridization in animals with female choice. *Animal Behaviour*, **58**, 1–12.

Zanetto A, Kremer A (1995) Geographical structure of gene diversity in *Quercus. petraea* (Matt.) Liebl. I. Monolocus patterns of variation. *Heredity*, **75**, 506–517.

Zhou R, Gong X, Boufford D, Wu CI, Shi S (2008) Testing a hypothesis of unidirectional hybridization in plants: observations on Sonneratia, Bruguiera and Ligularia. *BMC Evolutionary Biology*, **8**, 149.

This article is a part of O.L.'s PhD thesis focusing on hybridization dynamics between European white oak species. O.L. has a wide interest in application of molecular markers for studying the ecology, evolution and history of species. R.J.P. is a population geneticist with broad interest in evolution, phylogeography and mating system of trees. E.G. is a PhD student working on the characteristics of oak species used by the barrel industry. J.L. collaborated with O.L. during her Master; she is currently doing a PhD on the spatial and temporal variability of the mutualistic interaction between *Taxus baccata* L. and its frugivores' community. F.A. is a PhD student working on the adaptation of *Quercus petraea* (Matt.) Liebl. along an altitudinal gradient in the Pyrenean Mountains. A.K. has long standing interests in the evolution of temperate and tropical forest trees with particular emphasis on population differentiation at various levels where diversity is expressed (from genes to phenotypes). S.G. is a geneticist interested in population genetics and gene flow studies in forest trees, she supervised O.L.'s thesis.

© 2009 Blackwell Publishing Ltd

Supporting information

Additional supporting information may be found in the online version of this article:

Fig. S1 Neighbour-joining tree illustrating the net nucleotide genetic distances, as computed by the STRUCTURE software, between clusters at K = 6.

Fig. S2 Admixture coefficients distribution for simulated individuals: (A) pure species, (B) first generation hybrids (F1), (C) second generation hybrids (backcrosses) and (D) averaged distribution of pure species, first and second generation hybrids.

Table S1 Details of the sampled populations.

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 corresponding author for the article.

© 2009 Blackwell Publishing Ltd