Heredity (2005) 95, 457–465

www.nature.com/hdy

Gene flow and hybridisation in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in central Spain

M Valbuena-Carabaña[1], SC González-Martínez[2], VL Sork[3], C Collada[4], A Soto[2], PG Goicoechea[5] and L Gil[1]

[1]*Unidad de Anatomía, Fisiología y Genética, ETSIM, Ciudad Universidad s/n, Madrid 28040, Spain;* [2]*Unidad de Genética Forestal, Centro de Investigación Forestal, CIFOR-INIA, Carretera de La Coruña km 7.5, Madrid 28040, Spain;* [3]*Department of Ecology and Evolutionary Biology, University of California at Los Angeles (UCLA), CA 90095-1606, USA;* [4]*Departamento de Biotecnología, ETSIM, Ciudad Universidad s/n, Madrid 28040, Spain;* [5]*NEIKER A.B., Granja Modelo-Arkaute, Vitoria 01080, Spain*

Oaks are long-standing models for the study of gene flow and hybridisation. Temperate (*Quercus petraea*) and sub-Mediterranean (*Quercus pyrenaica*) oaks coexist in central Spain, showing remarkable differences in population size and structure. *Q. petraea* has a scattered distribution in central Spain, where it is at one of the southernmost limits of its range, and forms low-density stands; in contrast, *Q. pyrenaica* is widespread in the region. We selected a mixed population of the two species (~13 ha, 176 adults and 96 saplings) to compare the patterns of gene flow within each species and the extent of introgression between them. Using five nuclear microsatellite markers, we performed a parentage analysis and found considerable immigration from outside the stand (~38% for *Q. petraea* and ~34% for *Q. pyrenaica*), and estimated average seed-dispersal distances of 42 and 14 m for *Q. petraea* and *Q. pyrenaica*, respectively. Introgression between species was also estimated using our microsatellite battery. First, we developed a multivariate discriminant approach and, second, we compared our results with a widely used clustering method (STRUCTURE). Both analyses were consistent with a low level of introgression between *Q. petraea* and *Q. pyrenaica*. Indeed, only 15 adult trees, ~8.5%, were identified as putative hybrids when both methods of analysis were combined. Hybrids may be most common in contact zones due merely to physical proximity.
Heredity (2005) **95,** 457–465. doi:10.1038/sj.hdy.6800752; published online 26 October 2005

Keywords: microsatellites; gene flow; hybrid zones; introgression; genetic differentiation; *Quercus*

Introduction

In species with weak reproductive isolating mechanisms, such as oaks, long-distance gene flow and hybridisation are two potential sources of genetic variation. Hybridisation in the genus *Quercus* has attracted attention since Darwin (1859). Indeed, *Quercus* has been proposed as a model genus for a species concept that relies on ecological criteria, rather than reproductive isolation, to delimit species boundaries (Van Valen, 1976). In this genus, long-distance gene flow is suggested, then, to break the ecological isolation needed for new species formation, resulting in complex hybrid systems (eg, white oaks, including the *Quercus petraea–Q. robur* complex; Whittemore and Schaal, 1991; Muir *et al*, 2000; Petit *et al*, 2003).

The extent of hybridisation in oaks, as in other hybridising forest trees, is usually related to the phylogenetic distance between species (eg, Jiménez *et al*, 2004). When the ranges of two species overlap, hybrids occur more frequently at their geographical or ecological margins (Stebbins, 1950). Environmental conditions in such marginal areas might be more suitable for hybrids with intermediate characters or might cause breakdown of mate recognition systems in the parental species, due to the stress associated with the extreme ecological conditions at the edge of each species' range (Williams *et al*, 2001 and references therein). Hybridisation is often asymmetrical in oaks; a phenomenon which has been attributed to both genetical and ecological effects, including pollen-tube and ovary interactions (Boavida *et al*, 2001), decreased male fitness at hybrid zones – the environmental emasculation hypothesis (Williams *et al*, 2001) – and differences in flowering phenology (Belahbib *et al*, 2001). In the most intensively studied oak hybrid complex in Europe, the invasion by one species (*Q. petraea*) into the range of another (*Q. robur)* has been attributed to pollen swamping (Petit *et al*, 2003).

The potential for long-distance dispersal has been noted in many oak species (eg, Dow and Ashley, 1996, 1998; Streiff *et al*, 1999). Studies using a parentage analysis based on molecular markers have shown that mating occurs commonly among neighbours but a significant percentage of pollen donors come from outside the study area. For example, Dow and Ashley (1996, 1998) showed that more than 50% of mating events occurred at distances greater than 150–200 m in *Q. macrocarpa*, an American white oak. While the

Correspondence: L Gil, Unidad de Anatomía, Fisiología y Genética, Universidad Politécnica de Madrid, ETSI Montes, Ciudad Universidad s/n, Madrid 28040, Spain. E-mail: luis.gil@upm.es
Received 4 March 2005; accepted 26 July 2005; published online 26 October 2005

parentage approach illustrates the potential of long distant genes as a source of variation, recent studies based on the pollen pool structure among maternal trees (the TWOGENER approach; Smouse *et al*, 2001) suggest that the gene pool is dominated by the uneven reproductive contribution of local pollen donors (Smouse and Sork, 2004). To understand dispersal as an evolutionary process, gene flow studies using established offspring are required. To date, only a few examples are available in trees (see Dow and Ashley, 1996, for oaks), but the relevance of using estimates based on effective dispersal and not only on seed or pollen movements has been highlighted (Nathan and Muller-Landau, 2000).

The primary objective of this study is to understand the patterns of gene movement and hybridisation in a mixed stand of two *Quercus* species (*Quercus petraea* and *Q. pyrenaica*) with contrasting population size and ecological requirements. Our study plot (~13 ha) is located at the southern edge of the European distribution of *Q. petraea*, where stands of this species consist of only a few hundred mature individuals. *Quercus petraea* grows here under marginal ecological conditions, the water deficit during summer being a major limiting factor to growth and survival. In contrast, *Q. pyrenaica* is widespread in the region, the study plot being located in the central range of its distribution. *Quercus pyrenaica* is well known for its root and stem sprouting ability, which confers selective advantage after big ecological perturbations such as forest fires. The dense pilosity of its leaves confers protection from high levels of radiation and delayed bud burst protects the species from the effects of late frosts, a common climatic feature in Mediterranean mountains. These differences in abundance and in ecological suitability to the local environment might play an important role in reproductive patterns. Microhabitat segregation has been observed within the plot – *Q. petraea* occurs in deeper and less rocky soils – and contact zones between species are limited to only a few locations (Pardo *et al*, 2004).

Given the different ecological requirements of *Q. petraea* and *Q. pyrenaica*, it is important to evaluate the sources of genetic variation for local adaptation. In particular, it is imperative to ask whether the marginal populations of the temperate *Q. petraea* have recently exchanged genes with other remnant populations of the same species and/or with the more predominant oak species, *Q. pyrenaica*. Thus, this paper will address three questions:

(i) What patterns of gene movement can be detected within the stand, using samples of extablished saplings?
(ii) How much gene immigration has occurred from outside the plot, and do the two species differ in their rates of immigration?
(iii) How much introgression has occurred between the two species at this site (measured for adults and saplings)?

Materials and methods

Study area and plant material

The sampling plot is located at 'El Chaparral de Montejo de la Sierra' (hereafter, Montejo) in central Spain. Montejo forest is a protected area and represents the transition zone between the Atlantic and Mediterranean bioclimatic regions. In Montejo, typical tree species from central European temperate forests (eg, *Fagus sylvatica*, *Ilex aquifolium*, *Prunus avium* and *Q. petraea*) coexist with Mediterranean mountain species (eg, *Q. pyrenaica* and *Pinus sylvestris*). Apart from its bioclimatic location, the historical use of the land has determined the composition and structure of the present-day Montejo forest. Documents dating back to the 16th century report the intensive use of forest resources: agriculture, cattle grazing and logging for firewood. Open oak woodlands (*dehesas*) such as Montejo were managed to provide food and shelter for cattle. Once the traditional management was abandoned, partially in the 1960s and completely in 1974, the area was declared Natural Site of National Interest, and a dense regeneration of saplings established.

Montejo is a heterogeneous stand, and both *Q. pyrenaica* and *Q. petraea* individuals are scattered in low- to medium-density patches. A study plot, covering ~13 ha, was defined in a contact zone between *Q. petraea* and *Q. pyrenaica* on a hillside (Figure 1). Taxonomic identification of adults and saplings of *Q. petraea* and *Q. pyrenaica* was carried out during winter based on bud pilosity. Bud and leaves from *Q. petraea* are glabrous, whereas those from *Q. pyrenaica* show a clear pilosity. Differences in hirsuteness between these two species were further confirmed by using Kissling's gradient system (1977) in adult mature leaves collected in early summer. Confidence intervals for pilosity scores in *Q. petraea* and *Q. pyrenaica* did not overlap and no evidence of morphologically intermediate individuals was found at the study area. All adult individuals within the study plot, 93 *Q. petraea* and 83 *Q. pyrenaica*, were permanently marked with a chip embedded in the heartwood and mapped with the GIS Arc View system (ESRI Company, USA), once tree-coordinates had been measured with a GPS unit (Geoexplorer 3 Trimble Navigation, Trimble Navigation Company, USA) for long-term monitoring. Adult trees were between 200 and 300 years old (Alonso, 2001). A subplot of 900 m^2 was also sampled, including 52 *Q. petraea* and 44 *Q. pyrenaica* saplings. This sapling subplot was located in one of the main contact zones between species, on the uphill side of the study area (see Figure 1).

DNA extraction and molecular markers

Leaves from each of the 176 adults and 96 saplings were collected and stored at −80°C until DNA extraction. About 0.5 g of leaf material was ground in liquid nitrogen to a fine frozen powder, from which we extracted DNA, following a slightly modified protocol from Doyle and Doyle (1990). Five highly polymorphic microsatellites, which are located on different linkage groups, were analysed. Three microsatellite loci (*Qp*ZAG9, *Qp*ZAG36 and *Qp*ZAG110) were developed by Steinkellner *et al*, (1997) for *Q. petraea* and two more loci (MSQ4 and MSQ13) were by Dow *et al*, (1995) for *Quercus macrocarpa*.

The PCR was carried out in a Thermal Cycler Perkin Elmer GeneAmp PCR system 9700, using 0.4 U of Ecogen *Taq* DNA-Polymerase, and approximately 5 ng of genomic DNA in a total volume of 10 µl. The PCR mix also contained 0.2 µM of each primer (forward primers were

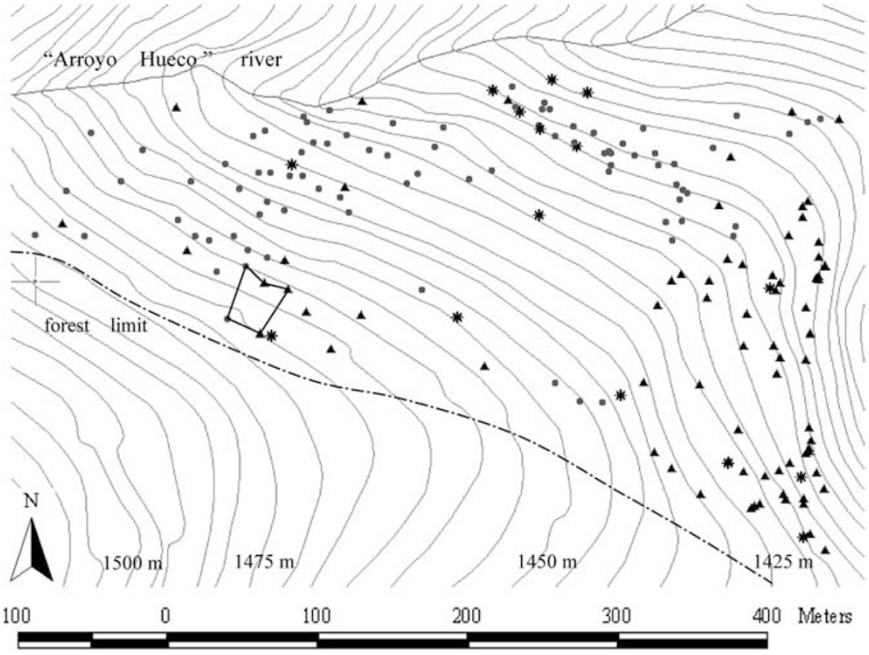

Figure 1 Intensive study plot at 'El Chaparral de Montejo de La Sierra'. *Quercus petraea* adult trees are represented by circles, and *Q. pyrenaica* by triangles. Filled stars represent hybrids detected using molecular markers during the course of this study (see the text for details). The continuous line delimits the subplot where natural regeneration was collected. Contour lines, the forest limit and the 'Arroyo Hueco' river are also shown.

labelled with the infra-red fluorescent IRDye-800 on the 5′ end, purchased from MWG Biotech), 0.2 mM dNTPs, 10 mM Tris/HCl pH 9 and 2 mM $MgCl_2$. After a preliminary denaturation at 94°C for 5 min, PCR amplifications were performed for 25 cycles under the following conditions: 1 min at 94°C, 30 s at the annealing temperature of 50°C and 1 min at 72°C. The amplification conditions for *Qp*ZAG36 were slightly different: $MgCl_2$ concentration was 3 mM and PCRs were performed for 40 s at 94°C, 45 s at the annealing temperature of 48°C and 45 s at 72°C. After the amplification, 2 µl of PCR product were mixed with 2 µl of loading buffer (78% formamide, 10 mM EDTA pH 7.6, 0.1% bromophenol blue and 0.1% xylene cyanol), heated for 5 min at 94°C, and quickly cooled on ice. Finally, 0.75 µl of denatured SSR fragments were loaded onto 25-cm denaturing gels, containing 8% acrylamide/bisacrylamide (19:1, W/V), 6 M urea and 0.4 × TBE buffer. Electrophoresis and scoring of fragments were performed on a 4200 Li-Cor automated DNA sequencer (Li-Cor Biosciences, Lincoln, NE, USA), using a 1 × TBE running buffer, with run parameters of 1500 V, 40 mA and 45°C plate temperature.

Gene movement within the plot

The microsatellite markers were highly polymorphic, resulting in high exclusion probabilities, computed following Jamieson and Taylor (1997), for parentage analysis (see Table 1). We used LOD scores and a simulation procedure to infer single parents and parent pairs for each sapling, as implemented in FAMOZ software (see details in Gerber *et al*, 2000, 2003). Using this method and considering no mistyping, we have complete discrimination among candidate parents within the plot and between them and gene flow coming from outside the plot. However, mistyping is very likely to occur when scoring microsatellites and any parentage analysis should allow for at least at a low rate (Gerber *et al*, 2000 and references therein). In our study, error due to mistyping was introduced at a rate of 0.0001 – both in the simulation procedure and in the assignation of the most-likely parents and parent pairs (Sophie Gerber, personal communication). In order to evaluate the influence of error rate on parental assignment and gene flow estimates, additional parentage analyses were conducted with even higher error rates (0.01–0.001).

Parentage was attributed to those candidate parent trees that had the highest LOD-score above the threshold value estimated using the simulation procedure. In some cases, both members of the parental couple could be attributed. Such pairs were always classified as the parents, even in the rare cases when one of them did not have the highest individual LOD-score, because the combined LOD-score for the pair was sufficiently high to be more convincing.

To test whether effective gametic dispersal was restricted by distance and whether differences between species can be explained by the heterogeneous location of adults within the plot, we simulated 10 000 reproduction events for the adult cohort and built a random dispersal distribution for each species. This distribution was compared with the actual values and confidence intervals were computed at the 95% probability level.

Immigration from outside the plot

The incoming gene flow was computed directly from the parentage analysis described above. Immigration was estimated as the number of gametes that originated from outside the plot (ie, number of gametes for which no

parent was found within the plot) over the total number of gametes sampled (ie, twice the number of offspring) and corrected by an estimate of cryptic gene flow extracted from simulated data (Gerber *et al*, 2000, 2003). The computation of cryptic gene flow required value for the census size of the population at the time of reproduction and the assumption of equal reproductive success between sampled and nonsampled trees. We used the census size provided by Alonso (2001) in the forest inventory of Montejo. This inventory reported an adult population census size of 600 for *Q. petraea* and 1000 for *Q. pyrenaica* in the complete Montejo forest (122 ha), where our study plot is located.

Genetic differentiation and introgression between species

A general estimate of genetic differentiation between species (F_{ST}) was computed using an analysis of variance framework (Weir and Cockerham, 1984). We also calculated a R_{ST} statistic, which takes account of allele size and was designed for molecular markers following the stepwise mutation model (Michalakis and Excoffier, 1996). To test whether estimates based on allele size were more informative than those following an infinite allele mutation (IAM) model with respect to species differentiation, we used a new test developed by Hardy *et al* (2003). This test is based on a randomisation procedure of allele sizes to determine whether stepwise-like mutations contributed to genetic differentiation. A significant test (ie, one indicating that R_{ST} performs better than F_{ST}) is expected when the stepwise mutation rate is higher or equal to the migration rate among populations. We used 10 000 random permutations of allele sizes among allelic states. The analyses were carried out using the program SPAGeDi (version 1.1) developed by Hardy and Vekemans (2002).

To confirm the species identity of each individual, both adults and saplings, and to study the patterns of introgression between species, we used a Bayesian approach, based on a discriminant procedure using the multilocus genotypes of mature trees. First, we generated a multivariate genotype of each adult tree by transforming the diploid genotypes into scores of allelic traits for each allele at a locus, following Smouse *et al* (1982). Our data set included 176 observations (adult sample size) and 102 different allelic variants (10–24 alleles per microsatellite and species). Because of the high number of allelic variants with respect to the number of observations, we reduced the number of variables using a Principal Component Analysis (procedure PRINCOMP of SAS version 8.1, SAS Institute Inc., NC, USA). We selected the first 40 principal components, which explained 82.9% of variation of the genetic data. After conducting a preliminary discriminant analysis, we excluded those principal components (seven) having a small correlation (<0.02) with the discriminant canonical axis. The remaining 33 principal components were used to perform the final discriminant analysis (procedure DISCRIM of SAS version 8.1, SAS Institute Inc., NC, USA). The discriminant analysis allowed us to construct a classification criterion to evaluate whether a particular individual in the sample belongs to one or the other species. Using the covariance matrix of the pooled data, we developed a linear discriminant function based on the genetic distances between the two species. With 33 variables, the generalised squared Mahalanobis distance between adults of both species was 15.1237 and the total squared canonical correlation was 0.7922. We established an *a priori* probability of belonging to one or the other species, using the average distance between the species, and using posterior probability, we reclassified each adult individual, as a gauge of taxonomic resolution. Next, to investigate levels of introgression in next-generation trees, we applied this discriminant criterion to the sapling cohort.

The pattern of introgression was further analyzed using a model-based clustering method (STRUCTURE version 2 software, Pritchard *et al*, 2000). This method assigns individuals (probabilistically) to populations (species in our case), using a Bayesian clustering approach. Estimated allele frequencies are used to compute the likelihood that a given genotype originated from a given population (see details in Pritchard *et al*, 2000). This method allows for the presence of admixed individuals in the clusters, giving for them a posterior likelihood of belonging to one or the other group. Several preliminary runs were performed to find the number of existing populations (K) giving the highest posterior probabilities. Once we knew the most-likely number of existing clusters in our data (two; see Results), we performed the final run, using a burn-in period of 10^6 (to minimise the effect of the starting configuration) and a running length period of 10^6. This running length is considered sufficient to obtain accurate estimates of posterior likelihood and admixture proportions for each individual (see documentation for STRUCTURE version 2 software).

Results

In *Q. pyrenaica* and *Q. petraea*, our survey showed high levels of heterozygosity for both adults and saplings (see Table 1 for adult trees). Using only five microsatellite markers, we were able to distinguish any single adult

Table 1 Genetic diversity and exclusion probabilities in adult trees for five microsatellite loci used in this study

Locus	A	A_p	H_e	EP_{sp}	EP_{pp}
Quercus petraea					
MSQ4	16	5	0.830	0.504	0.857
MSQ13	10	1	0.809	0.455	0.819
*Qp*ZAG9	18	3	0.866	0.566	0.888
*Qp*ZAG36	24	10	0.878	0.616	0.922
*Qp*ZAG110	20	4	0.847	0.546	0.887
Overall	88	23	0.846	0.980	1.000
Quercus pyrenaica					
MSQ4	20	9	0.833	0.509	0.860
MSQ13	16	7	0.586	0.204	0.591
*Qp*ZAG9	18	3	0.884	0.610	0.912
*Qp*ZAG36	15	1	0.855	0.539	0.872
*Qp*ZAG110	23	7	0.935	0.749	0.965
Overall	92	27	0.818	0.982	1.000

A: allelic richness; A_P: number of private alleles; H_e: Nei's expected heterozygosity; EP_{SP}: single-parent exclusion probability; EP_{PP}: parent-pair exclusion probability.
Estimates are given separately for *Q. petraea* ($n=93$) and *Q. pyrenaica* ($n=83$).

tree (ie, each one had a different genotype) and to assign the offspring in the parentage analysis with high statistical confidence. Different levels of putative vegetative reproduction were found for *Q. pyrenaica* and *Q. petraea*. In *Q. pyrenaica*, we found six groups (of two to four saplings each) sharing the same genotype and one sapling whose genotype was the same as an adult located less than 1 m away (probably a root sprout). In *Q. petraea*, only two saplings (out of 52 sampled) shared the same genotype, being located only 50 cm apart. In total, we found that ~25% of *Q. pyrenaica* saplings were of vegetative origin (*vs* ~4% of *Q. petraea*), which agrees with the high root-sprouting ability previously described for this species. It is worth noting that among *Q. pyrenaica* adults we have not found any tree sharing the same genotype.

Gene movement within the plot

Using parentage analysis, we identified at least one parent (among those genotyped) for ~96% of *Q. petraea* and 100% of *Q. pyrenaica* saplings. In addition, we found a second parent (ie, a matching parent pair) for ~31% and ~36% of *Q. petraea* and *Q. pyrenaica* saplings, respectively. Only two *Q. petraea* saplings (~4%) had no putative parent among the genotyped adults (Table 2). As a general pattern of gene flow from parents to offspring, progenitors were found nearby the sapling plot. Indeed, adults located closer than 45 m in *Q. petraea* and 12 m in *Q. pyrenaica* produced about 50% of the gametes with identified parents. Nevertheless, some long-distance dispersal events occurred in both species. In fact, ~12% of *Q. petraea* and ~19% of *Q. pyrenaica* gametes were originated from adults located more than 200 m apart. The longest dispersal event found in this study was of 381 m, in an East to West direction.

Comparing random (ie, affected only by adult spatial location) and the inferred dispersal distributions, we found restricted gene dispersal up to 75 m for *Q. petraea* and to 25 m for *Q. pyrenaica* (Figure 2). Within these distance classes, a similar percentage of gametes was produced in both species (41% for *Q. petraea* and 44% for *Q. pyrenaica*). The long-distance dispersal events described above appeared to be slightly less numerous than expected from adult location alone.

Table 2 Number of offspring and gametes (considering only one offspring from each clone assembly) for which a single parent or a parent pair was found among the genotyped adult trees

	n	*Parentage analysis*				*Immigration (%)*	
		No match	*Single parent*	*Parent pair*	*Total*	*Apparent*	*True*
Quercus petraea							
Offspring	51	2	34	15	49		
Gametes	102	38	34	30	64	37.26	38.43
Quercus pyrenaica							
Offspring	33	0	21	12	33		
Gametes	66	21	21	24	45	31.82	34.29

Gametes with identified origin equal twice the number of offspring when a parent pair is found and the number of offspring when only a single parent is found. Apparent gene flow and true gene flow (corrected by cryptic gene flow) at the gamete level are also given.

Immigration from outside the plot

Apparent gene flow (as gametes) from outside the plot (~13 ha) was ~37% for *Q. petraea* and ~32% for *Q. pyrenaica* (Table 2). The rate of cryptic gene flow, as estimated by FAMOZ software, was relatively small; the difference between cryptic and apparent gene flow was ~3% for *Q. petraea* and ~8% for *Q. pyrenaica*. Thus, corrected gene flow estimates are close to the direct estimates: ~38% for *Q. petraea* and ~34% for *Q. pyrenaica*. Assuming higher mistyping and simulation error rates resulted in higher gene flow estimates, as a consequence of increased estimates of the cryptic gene flow. Thus, the immigration rates reported here should be considered as a lower bound.

Genetic differentiation and introgression between species

Genetic differentiation between *Q. petraea* and *Q. pyrenaica* was moderate but significant, as shown by *F*- and *R*-statistics ($F_{ST} = 0.081$, SD = 0.039; $R_{ST} = 0.109$, SD = 0.046). A 10 000 iteration permutation test of the alleles sizes indicated that differentiation estimated by R_{ST} was not significantly different from that estimated by F_{ST}. These results are consistent with the fact that most alleles were shared by both species (Figure 3). Species-private alleles were typically found at low frequency (<0.05), except for *Q. petraea* alleles 226 bp at MSQ13 (frequency = 0.11) and 224 bp at *Qp*ZAG36 (frequency = 0.07), which might be useful for discriminating species.

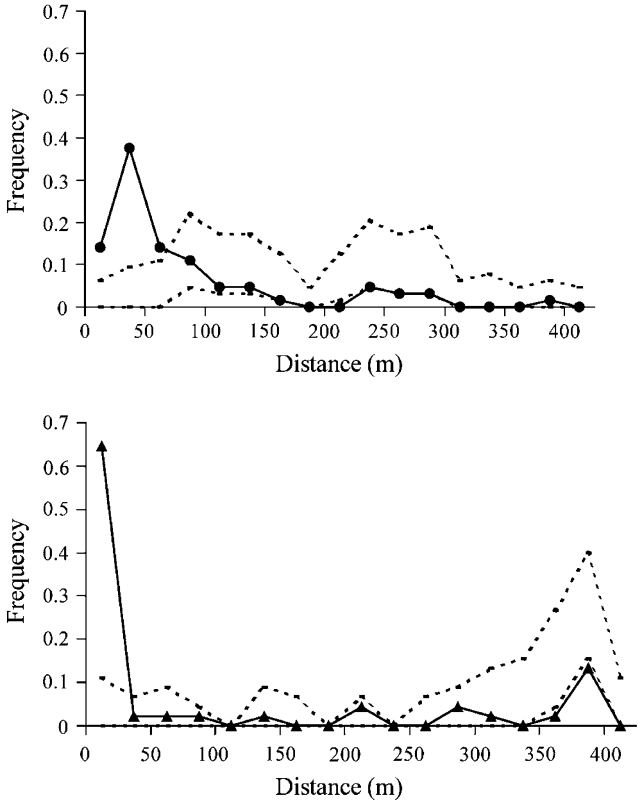

Figure 2 Distribution of gamete dispersal distances of *Q. petraea* (above) and *Q. pyrenaica* (below) as inferred using a combined parentage analysis for single parents and parent pairs. Confidence intervals at 95% level based on random dispersal distributions are shown with discontinuous lines.

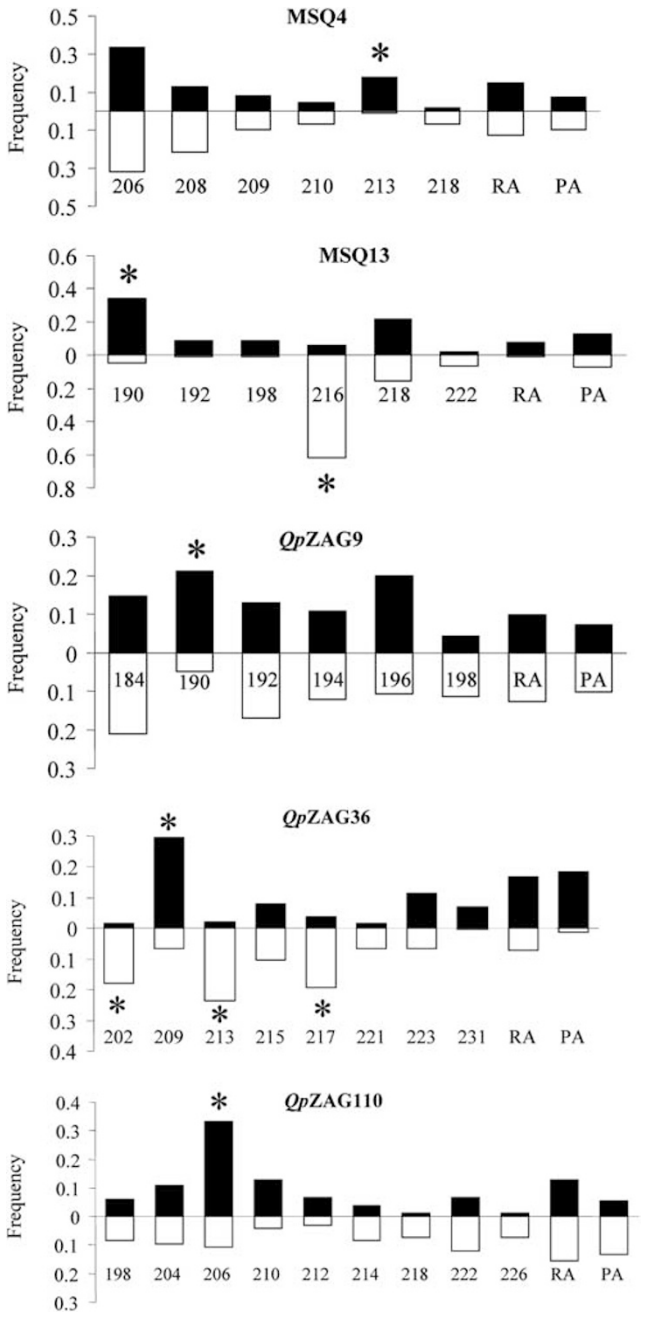

Figure 3 Allelic frequency for shared and species-private alleles in *Q. petraea–pyrenaica* hybrid system. Allele sizes are given in base pairs number. Shared low-frequency alleles (<0.05) and private alleles were pooled in single categories named RA and PA, respectively. Asterisks are used to mark those alleles useful for species discrimination. Bars in black represent *Q. petraea* adult trees and bars in white *Q. pyrenaica* adult trees.

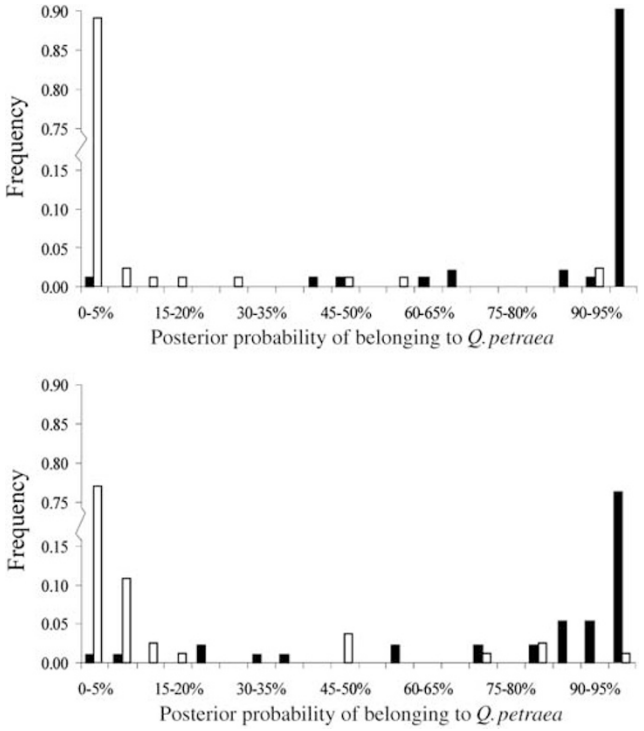

Figure 4 Individual posterior probability of belonging to its own species cluster for adult trees. Two different methods were used: a multivariate discriminant procedure (above) and STRUCTURE model-based method (below). Bars in black represent *Q. petraea* adult trees and bars in white *Q. pyrenaica* adult trees (as classified by bud hirsuteness).

Using our multivariate discriminant criterion (at the 95% level) and the battery of molecular markers assayed, we found ~10% of introgressed adult trees and ~1.2% of introgressed saplings. Introgressed individuals are defined here as those having less than a 95% probability of belonging to its own species (as shown by bud hirsuteness). Introgression was apparently bidirectional, since it occurred with approximately equal frequency in each morphological class. Only four mature trees (~2%) had a high likelihood of being F_1 hybrids (ie, those having a posterior probability of 40–60%).

The model-based clustering (STRUCTURE software) showed an optimal number of populations (K) of two within our data set, which agrees with the existence of two differentiated species. We compared the assignments with those based on our discriminant function (Figure 4). The fraction of potential introgressed trees was two-fold higher (~23%), although half of them had still high probability (90–95%) of belonging to its own species class. The number of potential F_1 hybrids (five out of 176 trees) was similar to that found using the discriminant analysis. Considering only the individuals classified as hybrids by both methods (represented by filled stars in Figure 1), a reliable minimum estimate of introgression for this hybrid system would be ~8.5% (15/176).

Discussion

To the best of our knowledge, this is the first study that addresses gene flow and hybridisation in *Q. pyrenaica*, a widespread oak species in southwestern France, the Iberian Peninsula, and northern Morocco. In contrast, *Q. petraea*, which is broadly distributed in Europe, has been the object of many studies of gene diversity, population structure and mating system in the past decade (Dumolin-Lapègue *et al*, 1999 and references therein). In the present study, we analysed a marginal population of *Q. petraea* from its southernmost range, where limiting ecological conditions might have significant effects on gene flow and genetic diversity. Introgression between

Q. petraea and *Q. pyrenaica* was also studied here for the first time, and provides a comparison with the well-studied *Quercus petraea* – *Q. robur*, hybrid system.

Gene movement into and within the plot

The estimated immigration rates were similar for the two species, being ~38% for *Q. petraea* and ~34% for *Q. pyrenaica*. In contrast, gametic movements within the stand revealed differences: the average distance for a successful mating (ie, effective pollen flow) in *Q. pyrenaica* was three-fold that in *Q. petraea* (270 and 92 m, respectively), being consistent with pairwise distances between trees of the same species at Montejo (329 m for *Q. pyrenaica* and 76 m for *Q. petraea*). Conversely, if we assume that the closest parent is the mother tree (as in Dow and Ashley, 1996), average effective seed movement was 14 m for *Q. pyrenaica* and 42 m for *Q. petraea*. Differences in acorn weight between species, 2.14 g (SD = 0.68) in *Q. petraea* and 3.26 g (SD = 0.67) in *Q. pyrenaica* (our unpublished results), might account for this disparity in seed displacement. Overall, gametic movement within the plot was more restricted in *Q. pyrenaica* (up to 25 m) than in *Q. petraea* (up to 75 m).

Our results may be explained by the existence of a short- and a long-range seed dispersal component. Most acorns are probably dispersed over short distances, while a few are dispersed over large distances. Assuming that the mother tree is the closest parent, a more detailed analysis of seed movement was possible. Under this assumption, only ~4% (2/49) of *Q. petraea* sapling and none (0/33) of *Q. pyrenaica* were found below their mother's crown (~6 m from the trunk), which might imply a low level of effective primary seed dispersal by gravity. For only five saplings, ~2% (1/49) in *Q. petraea* and ~12% (4/33) in *Q. pyrenaica*, a secondary transportation down the slope had taken place, carrying seeds 23 to 37 m from its mother's trunk. The rest of the saplings with identified parents, ~94% (46/49) in *Q. petraea* and ~88% (29/33) in *Q. pyrenaica*, were located at higher elevation than their putative mother trees (slopes up to 10%). Birds and rodents might be responsible for these up-slope seed movements, and for those coming from outside the plot. Gómez (2003) found that jays (*Garrulus glandarius*) moved acorns of *Q. ilex* from 5 m to 1 km, with an average of 72 ± 26 m, within the same oak woodland. Although, we have not observed jays foraging at Montejo, we have at the nearest *Q. pyrenaica* population, and jay-mediated seed movements have been reported in other oak forests from Central Spain (L Carrascal, personal communication). Also, long-distance acorn movements could have been caused by common squirrels (*Sciurus vulgaris*), frequently observed at Montejo, or by other animals such as, golden orioles (*Oriolus oriolus*), woodpigeons (*Columba palumbus*), turtledoves (*Streptopelia turtur*) and badgers (*Meles meles*, abundant in Montejo in former times). Short-distance acorn displacement could be affected by blackbirds (*Turdus merula*) and song thrushes (*Turdus philomelos*) (JL Ceresuela, personal communication).

Genetic differentiation and introgression between species

Molecular differentiation between *Q. petraea* and *Q. pyrenaica* was high with respect to other hybridising oaks, as estimated assuming both SMM and IAM models. The differentiation between the two species was six-fold the average differentiation between *Q. petraea* and *Q. robur* – the more intensively studied hybrid system in oaks – (Mariette *et al*, 2002). Bruschi *et al* (2000) reported little molecular differentiation between *Q. petraea* and *Q. pubescens* in northern and central Italy ($R_{ST} = 0.048$), an oak complex where overlapping morphological traits is often complete and multiple hybrids have been described. The differentiation that we have found between *Q. petraea* and *Q. pyrenaica* is therefore the largest reported for any oak species. This differentiation might indicate a greater phylogenetic distance between *Q. petraea* and *Q. pyrenaica*. Differences in ecological requirements support this assumption. However, *Q. pyrenaica* is not usually included in phylogenetic studies and its relationship with other European white oaks remains poorly characterised. The much lower sympatry between *Q. petraea* and *Q. pyrenaica*, compared to *Q. petraea* and *Q. robur*, is another fact that might explain their higher genetic differentiation, as opportunities for extensive hybridisation and coadaptation are scarce. In the Iberian Peninsula, for instance, contact areas are reduced to the southern slopes of the Cantabric Mountain Range and some isolated populations in the northern Castilian Plateau.

Introgression between *Q. petraea* and *Q. pyrenaica* is not frequent at Montejo, although it takes place at a low level (minimum of ~8.5%). This low rate was consistent with the relatively large genetic differentiation between these two oak species. Whilst Muir *et al* (2000) failed in the assignment of 22% of individuals to *Q. petraea* or *Q. robur* using 20 microsatellites, four microsatellites were sufficient to assign almost all individuals in a mixed *Q. lobata* and *Q. douglasii* population (Craft *et al*, 2002). In our study, using five microsatellite markers, we were able to infer the degree of introgression in Montejo, which is similar to that reported by Craft *et al* (2002). Differences in phenology between species might be responsible for the low level of introgression. *Quercus pyrenaica* flowering takes place about 2 weeks later than *Q. petraea*'s, although some overlap has been observed. The protandrous character of *Q. pyrenaica* might allow its early pollen to fertilise late female *Q. petraea*'s flowers, leading to asymmetrical hybridisation that might also be favoured by the larger population size of *Q. pyrenaica*. Phenological evidence and apparently bidirectional hybridisation (as shown by our study) are against a model of asymmetrical hybridisation favoring *Q. petraea* colonisation of new ranges through pollen swamping, a process described for the *Quercus petraea–Q. robur* hybrid system (Petit *et al*, 2003). Most introgressed trees were located in contact areas either in the central part of the plot or in the northern part, near the river, which is the limit of the studied area (see Figure 1). Upstream the river, a large population of *Q. pyrenaica* exists, while the closest population of mature *Q. petraea* is located some kilometres away. Hybrids may be most common in contact zones due merely to physical proximity. Alternatively, contact areas might represent ecotones for edaphic or hydrological parameters favouring the establishment of intermediate individuals (Williams *et al*, 2001), given the different ecological requirements of *Q. pyrenaica* and *Q. petraea*.

Reports of natural hybrids among the studied species are scarce, which agrees with the low level of introgression found in our study. Vicioso (1950) cited the presence of hybrids in a few localities in the northern Castilian Plateau in Spain (named *Q.* × *legionensis*). Hy (1895) gave the name of *Q.* × *trabuti* to some infertile and morphological intermediate individuals found in the French region of Angers. Interspecific gene flow between oak species probably occurs at higher rate than the proportion of intermediate individuals found (see Petit *et al*, 2003 for *Q. petraea–robur* and our own results). Morphologically intermediate individuals are generally scarce in mixed populations of hybridising oaks (Kremer *et al*, 2002; Mariette *et al*, 2002) and less frequent than inferred from cytoplasmic gene flow analyses (Whittemore and Schaal, 1991; Jiménez *et al*, 2004). The relatively quick reversion of morphological traits to parental types in the progeny of natural and artificial hybrids might be responsible for these difficulties in finding intermediate morphological traits (Stebbins, 1950).

In the Iberian Peninsula, altitude and orography, along with latitudinal climate changes are responsible for present-day distribution of oak species. Recent studies have shown the existence of quaternary refugia for the white oaks, not only in southern Iberia but also in northern parts of the territory (Olalde *et al*, 2002, and references therein). Both *Q. petraea* and *Q. pyrenaica* are fixed for haplotype *H10a* (using the extended nomenclature of Dumolin-Lapègue *et al*, 1997) at Montejo (our unpublished results), the most common haplotype of the Iberian *B* cpDNA lineage. As previously shown (Dumolin-Lapègue *et al*, 1997, 1999), cpDNA variation is geographically structured in European white oaks, implying that related haplotypes have similar geographical distribution, irrespective of the oak species within which they are found. Common glacial refugia for *Q. petraea* and *Q. pyrenaica* in the Iberian Peninsula might explain haplotype sharing between these two species, but whether coincident postglacial recolonisation routes from the same refugia or posterior contact of different-species migration routes in central Spain are responsible for the existence of a fixed haplotype in Montejo, remains uncertain. Nevertheless, both scenarios would indicate an ancient origin of hybridisation between *Q. petraea* and *Q. pyrenaica* in central Spain.

In conclusion, we have found similar immigration rates for both species, so the scattered distribution of *Q. petraea* in the area does not imply greater isolation than for *Q. pyrenaica*, which is growing in its core range. Not only acorn size and weight, but also the preferences of disperser animals (perhaps due to acorn palatability), might account for the differences in the oak species' dispersal patterns. Lastly, our battery of microsatellite markers was able to (i) characterise each tree individually, (ii) identify clones, (iii) discriminate between species, and (iv) estimate individual introgression levels, thus providing a useful tool for the management of forest genetic resources of these two hybridising oak species.

Acknowledgements

We are grateful to Ricardo Alía for valuable discussion on the original idea and design of this study. Comments on an early version of the manuscript by Peter E. Smouse, Olivier J. Hardy and an anonymous reviewer are much appreciated. María Valbuena-Carabaña was supported by a PhD scholarship from the 'Comunidad Autónoma de Madrid' and Santiago C González-Martínez by a Fulbright/MECD scholarship at University of California (Davis) and the 'Ramón y Cajal' fellowship RC02-2941. Victoria L Sork was supported by NSF-DEB-0242422 Grant. This work was funded by CAM 07M/0011/2000 and CAM 07M/0012/2002.

References

Alonso J (2001). Inventario Forestal del Sitio Natural de Interés Nacional 'Hayedo de Montejo de la Sierra', Monte no 89 del C.U.P. 'El Chaparral y La Solana' (Comunidad de Madrid). MSc Thesis, Universidad Politécnica de Madrid.

Belahbib N, Pemonge M-H, Ouassou A, Sbay H, Kremer A, Petit RJ (2001). Frequent cytoplasmic exchanges between oak species that are not closely related: *Quercus suber* and *Q. ilex* in Morocco. *Mol Ecol* **10**: 2003–2012.

Boavida LC, Silva JP, Feijó JA (2001). Sexual reproduction in the cork oak (*Quercus suber* L). II Crossing intra- and interspecific barriers. *Sexual Plant Reprod* **14**: 143–152.

Bruschi P, Vendramin GG, Bussotti F, Grossoni P (2000). Morphological and molecular differentiation between *Quercus petraea* (Matt.) Liebl. and *Quercus pubescens* Willd. (Fagaceae) in Northern and Central Italy. *Ann Botany* **85**: 325–333.

Craft KJ, Ashley MV, Koenig WD (2002). Limited hybridization between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a mixed stand in central coastal California. *Am J Botany* **89**: 1792–1798.

Darwin C (1859). *On the Origins of Species by Means of Natural Selection*. John Murray: London, UK.

Dow BD, Ashley MV (1996). Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. *Mol Ecol* **5**: 615–627.

Dow BD, Ashley MV (1998). High levels of gene flow in bur oak revealed by paternity analysis using microsatellites. *J Hered* **89**: 62–70.

Dow BD, Ashley MV, Howe HF (1995). Characterization of highly variable $(GA/CT)_n$ microsatellites in the bur oak, *Quercus macrocarpa*. *Theoret Appl Genet* **91**: 137–141.

Doyle JJ, Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus* **12**: 13–15.

Dumolin-Lapègue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997). Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146**: 1475–1487.

Dumolin-Lapègue S, Kremer A, Petit RJ (1999). Are chloroplast and mitochondrial DNA variation species-independent in oaks? *Evolution* **53**: 1406–1413.

Gerber S, Chabrier P, Kremer A (2003). FaMoz: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Mol Ecol Notes* **3**: 479–481.

Gerber S, Mariette S, Streiff R, Bodénès C, Kremer A (2000). Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Mol Ecol* **9**: 1037–1048.

Gómez JM (2003). Spatial patterns in long-distance dispersal of *Quercus ilex* acorns by jays in a heterogeneous landscape. *Ecography* **26**: 573–584.

Hardy OJ, Charbonnel N, Fréville H, Heuertz M (2003). Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* **163**: 1467–1482.

Hardy OJ, Vekemans X (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* **2**: 618–620.

Hy FC (1895). Sur quelques chênes hybrides des environs d'Angers. *Bull Soc Bot France* **42**: 552–559.

Jamieson A, Taylor SS (1997). Comparisons of three probability formulae for parentage exclusion. *Anim Genet* **28**: 397–400.

Jiménez P, López de Heredia U, Collada C, Lorenzo Z, Gil L (2004). High variability of chloroplast DNA in three Mediterranean evergreen oaks indicates complex evolutionary history. *Heredity* **93**: 510–515.

Kissling P (1977). Les poils des quatre espèces de chênes du Jura (*Quercus pubescens*, *Q. petraea*, *Q. robur* et *Q. cerris*). *Berich Schweiz Bot Ges* **87**: 1–18.

Kremer A, Kleinschmit J, Cottrell J, Cundall EP, Deans JD, Ducousso A *et al* (2002). 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 Ecol Manage* **156**: 75–87.

Mariette S, Cottrell J, Csaikl UM, Goicoechea PG, König A, Lowe AJ *et al* (2002). Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Q. petraea* (Matt.) Liebl. and *Q. robur* L. stands. *Silvae Genet* **51**: 72–79.

Michalakis Y, Excoffier L (1996). A genetic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142**: 1061–1064.

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

Nathan R, Muller-Landau HC (2000). Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends Ecol Evol* **15**: 278–285.

Olalde M, Herrán A, Espinel S, Goicoechea PG (2002). White oaks phylogeography in the Iberian Peninsula. *Forest Ecol Manage* **156**: 89–102.

Pardo F, Gil L, Pardos JA (2004). Structure and composition of pole-stage stands developed in an ancient wood pasture in central Spain. *Forestry* **77**: 67–74.

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

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

Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001). Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution* **55**: 260–271.

Smouse PE, Sork VL (2004). Measuring pollen flow in forest trees: a comparison of alternative approaches. *Forest Ecol Manage* **197**: 21–34.

Smouse PE, Spielman RS, Park MH (1982). Multiple-locus allocation of individuals to groups as a function of the genetic variation within and differences among human populations. *Am Naturalist* **119**: 445–463.

Stebbins GL (1950). *Variation and Evolution in Plants*. Columbia University Press: New York.

Steinkellner H, Fluch S, Turetschek E, Lexer C, Streiff R, Kremer A *et al* (1997). Identification and characterization of $(GA/CT)_n$-microsatellite loci from *Quercus petraea*. *Plant Mol Biol* **33**: 1093–1096.

Streiff R, Ducousso A, Lexer C, Steinkellner H, Gloessl J, Kremer A (1999). Pollen dispersal inferred from paternity analysis in a mixed stand of *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Mol Ecol* **8**: 831–841.

Van Valen L (1976). Ecological species, multispecies, and oaks. *Taxon* **25**: 233–239.

Vicioso C (1950). *Revisión del Género Quercus en España*. Ministerio de Agricultura: Madrid.

Weir BS, Cockerham CC (1984). Estimating F-statistic for the analysis of population structure. *Evolution* **38**: 1358–1370.

Whittemore AT, Schaal BA (1991). Interespecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences of USA* **88**: 2540–2544.

Williams JH, Williams JB, Howard DJ (2001). Reproductive processes in two oak (*Quercus*) contact zones with different levels of hybridisation. *Heredity* **87**: 680–690.