Molecular Ecology (2005) **14**, 549–561 doi: 10.1111/j.1365-294X.2004.02418.x

Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.)

GRAHAM MUIR* and CHRISTIAN SCHLÖTTERER
Institut für Tierzucht und Genetik, Veterinärmedizinische Universität Wien, Josef Baumann Gasse 1, 1210, Wien, Austria

Abstract

***Quercus petraea* and *Quercus robur* are two closely related oak species, considered to hybridize. Genetic markers, however, indicate that despite sharing most alleles, the two species remain separate genetic units. Analysis of 20 microsatellite loci in multiple populations from both species suggested a genome-wide differentiation. Thus, the allele sharing between both species could be explained either by low rates of gene flow or shared ancestral variation. We performed further analyses of population differentiation in a biogeographical setting and an admixture analysis in mixed oak stands to distinguish between both hypotheses. Based on our results we propose that the low genetic differentiation among these species results from shared ancestry rather than high rates of gene flow.**

Keywords: hybridization, microsatellites, oaks, *Quercus*, shared ancestral alleles

Received 11 June 2004; revision received 25 October 2004; accepted 25 October 2004

Introduction

Gene flow among populations is a central parameter of some speciation models. Nevertheless, gene flow is notoriously difficult to distinguish from shared ancestral variation as both scenarios produce a similar pattern of allele sharing. One example of this phenomenon occurs within the *Quercus* genus. Allele sharing is omnipresent among closely related species pairs in the group and the differentiation among species results from allele frequencies rather than private alleles (e.g. Howard *et al*. 1997). In Europe, *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. are a species pair considered to hybridize over most of their sympatric range while retaining their genetic, morphological and ecological identity (e.g. Bodénès *et al*. 1997; Ponton *et al*. 2001; Kremer *et al*. 2002). Radiocarbon-dated pollen identifies refugia located in the Iberian, Italian and Balkan Peninsulas as centres for the expansion of these species following the last glaciation (Brewer *et al*. 2002). Spatially, the two species share chloroplast haplotypes more readily than by chance. This nonrandom distribution of haplotypes has been interpreted as evidence for hybridization between the two taxa before and after expansion (Ferris *et al*. 1993; Dumolin-Lapègue *et al*. 1997). Regardless of the number of sites assayed, however, gene flow is difficult to quantify in the chloroplast genome as all polymorphisms recover only one genealogy (the clonally inherited chloroplast, Dumolin *et al*. 1995). Chloroplast capture or 'pollen swamping', an asymmetrical bias in the fertilization of *Q. robur* ovules by *Q. petraea* pollen, has been invoked to account for the low nuclear divergence between the species (Bacilieri *et al*. 1996; Petit *et al*. 1997). Attempts to quantify the direction of gene flow, however, may be compromised by the inherent circularity of a priori assignments to parental species (based on morphology). Despite a substantial fraction of shared alleles (Bodénès *et al*. 1997; Muir *et al*. 2001), the combined analysis of multiple polymorphic markers indicates that *Q. petraea* and *Q. robur* remain two separate genetic entities (Zanetto *et al*. 1994; Muir *et al*. 2000; Gömöry *et al*. 2001; Coart *et al*. 2002).

Three hypotheses can be proposed to reconcile the significant genetic differentiation among *Q. petraea* and *Q. robur* despite a high proportion of shared alleles. (1) Low gene flow. Rates of gene flow may be too low to compensate for genetic drift within the two species. Thus, gene flow may prevent drift generating species-specific alleles, but drift produces different allele frequencies in both species. The range of parameters for which this scenario holds, however, has not been explored. Generally speaking, low

Correspondence: Christian Schlötterer, Fax: + 43 1 25077 5693; E-mail: Christian.Schloetterer@vu-wien.ac.at
*Present address: School of Biosciences, University of Birmingham, Birmingham, BI5 2TT, UK.

levels of migration (i.e. > one migrant per generation) are sufficient to prevent population differentiation by drift (Slatkin 1987). Hence, it can be assumed that the low gene flow hypothesis only holds for a restricted set of parameters. (2) Selection. This hypothesis assumes gene flow among species, but species integrity is maintained by strong selection operating on a few loci. High levels of differentiation are predicted from this model for genomic regions linked to selected loci maintaining species integrity. No (or low) differentiation is predicted for most of the genome. (3) Shared polymorphism: this hypothesis assumes a recent speciation event and large effective population sizes (Pamilo & Nei 1988). Consequently, even in the absence of gene flow, alleles will be shared among species. Nevertheless, genetic differentiation among *Q. petraea* and *Q. robur* should be observed across the entire genome.

In order to differentiate between gene flow and shared ancestral variation, we examined the genetic differentiation among species pairs with different geographical spacing. In a biogeographical context, gene flow is expected to result in less differentiation between neighbouring populations than among distantly located ones. If both species retain shared ancestral alleles (scenario 3) however, geographical neighbours are expected to be no more differentiated than distantly located ones. Biogeography could then be used to distinguish gene flow from shared ancestral variation. In order to discriminate between the hypotheses, we evaluated the differentiation between *Q. petraea* and *Q. robur* at 20 microsatellite loci using multiple populations from both species in a biogeographical setting. Consistent with the hypothesis of shared ancestral variation we found similar levels of differentiation between *Q. petraea* and *Q. robur* populations, irrespective of their geographical origin.

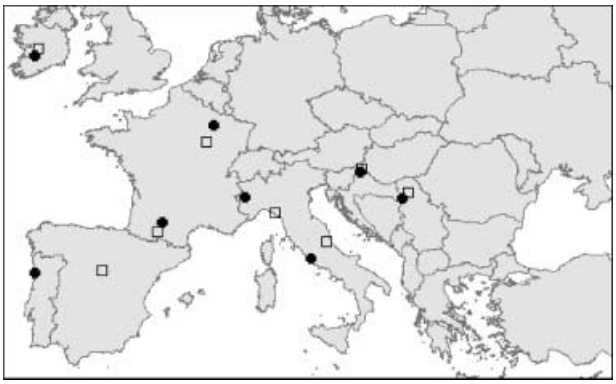

Fig. 1 Sampling localities across Europe. *Quercus petraea* (squares), *Quercus robur* (circles). While not included in the locus-wise F_{ST} analysis, the two species pairs from Italy were genotyped to check the presence of specific alleles (see Results, Table 1).

Materials and methods

Plant material

To account for geographical differences, one population from each species was sampled at six sites across Europe (Fig. 1, Table 1), including putative glacial refugia (see also Muir *et al*. 2000). Collections were made from individuals in northern Spain and Portugal (*Quercus petraea* 41°07′ N 3°30′ W, *Quercus robur* 41°00′ N 8°41′ W), respectively, southwest France (*Q. petraea* 43°11′ N 0°40′ E, *Q. robur* 43°42′ N 1°06′ E),

Table 1 Species, populations and individuals surveyed in this study

Sampling location*	Species	Description	Loci†	*n*‡	Analyses
Portugal	*Q. robur*	Population	21	18	Locus-wise F_{ST}
Spain	*Q. petraea*	Population	21	17	Locus-wise F_{ST}
Southwest France	*Q. petraea*/*Q. robur*	Species pair	21	27	Locus-wise F_{ST}, local vs. global F_{ST}
Eastern France	*Q. petraea*/*Q. robur*	Species pair	21	30	Locus-wise F_{ST}, local vs. global F_{ST}
Slovenia	*Q. petraea*/*Q. robur*	Species pair	21	32	Locus-wise F_{ST}, local vs. global F_{ST}
Serbia	*Q. petraea*/*Q. robur*	Species pair	21	34	Locus-wise F_{ST}, local vs. global F_{ST}
Ireland	*Q. petraea*/*Q. robur*	Species pair	20	54	Locus-wise F_{ST}, local vs. global F_{ST}
Approximately equidistant to pair above	*Q. petraea*	Population	20	29	In tandem with pair above, STRUCTURE (admixture)
Northern Italy	*Q. petraea*/*Q. robur*	Populations	5	15	Analysis of loci with high F_{ST} values: 96, 7, 112 & reference loci (9, 110)
Central Italy	*Q. petraea*/*Q. robur*	Populations	5	16	Analysis of loci with high F_{ST} values: 96, 7, 112 & reference loci (9, 110)
Vienna	*Q. palustris*	Individual	5	1	Outgroup (ancestral states)
Vienna	*Q. cerris*	Individuals	5	2	Outgroup (ancestral states)

*Grid references given in Materials and methods, see also Fig. 1; †Locus 962 (71 base pairs downstream of 96) was surveyed in 158 individuals in addition to the 20 microsatellites genotyped in the main survey; ‡Number of analysed individuals.

Table 2 Diversity indices for the 21 investigated microsatellite loci in *Quercus petraea* and *Quercus robur*

		H_O		H_E		V		Average repeat no.		No. of alleles		n	
Locus*	Repeat†	*Q. pet.*	*Q. rob.*	*Q. pet.*	*Q. rob.*	*Q. pet.*	*Q. rob.*	*Q. pet.*	*Q. rob.*	*Q. pet.*	*Q. rob.*	*Q. pet.*	*Q. rob.*
110(8)	$(AG)_{15}$	0.80	0.71	0.86	0.76	26	15	11	9	22	17	103	105
1/5(7)	$(GT)_5(GA)_9$	0.70	0.71	0.89	0.81	8	3	14	13	21	13	97	100
7	$(AG)_{13}(AAAG)_3$	0.16	0.20	0.89	0.68	32	21	9	10	17	18	79	98
16(6)	$(AG)_{21}$	0.81	0.84	0.92	0.93	9	22	17	20	21	28	103	105
9(7)	$(AG)_{12}$	0.90	0.88	0.87	0.88	11	25	14	14	15	15	101	103
119(2)	$(GA)_{24}$	0.79	0.59	0.94	0.87	78	90	20	17	32	30	87	101
15(9)	$(AG)_{23}$	0.63	0.71	0.82	0.78	21	10	11	10	14	13	87	103
1/2	$(AG)_{16}$	0.67	0.51	0.79	0.59	12	4	7	5	11	9	81	98
46(2)	$(AG)_{13}$	0.89	0.80	0.92	0.86	29	15	9	9	22	28	97	101
7§	$(TC)_{17}$	0.87	0.87	0.92	0.93	30	26	8	7	22	20	91	104
87	$(TC)_{20}$	0.92	0.82	0.86	0.89	9	25	19	20	17	21	102	101
96	$(TC)_{20}$	0.81	0.34	0.90	0.42	25	21	12	8	20	14	99	101
962‡	$(TC)_8$	0.40	0.22	0.57	0.34	0.4	0.3	8	9	4	5	65	78
112	$(GA)_{32}$	0.28	0.63	0.30	0.75	0.9	12	10	14	7	17	103	102
108	$(GA)_{19}(GGGA)_3$	0.39	0.40	0.66	0.81	17	27	9	12	11	13	95	96
101	$(TC)_{20}(AC)_{15}$	0.79	0.78	0.85	0.84	19	11	31	31	17	16	105	105
20	$(TC)_{18}$	0.91	0.83	0.89	0.83	31	19	15	13	21	18	98	105
31	$(GA)_{31}$	0.56	0.43	0.96	0.95	57	62	19	20	42	41	90	89
58(5)	$(GA)_{34}$	0.48	0.58	0.89	0.96	39	69	44	47	30	41	91	100
36(2)	$(AG)_{19}$	0.89	0.80	0.91	0.87	17	12	14	13	19	15	105	105
102(9)	$(AG)_5AA(AG)_{13}$	0.87	0.82	0.94	0.94	28	26	16	15	32	36	97	96
	Average	0.69	0.64	0.84	0.79	24	25	15	15	20	20		

*When available, linkage groups mapped in *Quercus robur* by Barreneche *et al*. (1998) are shown in brackets; †Repeat units and motif for each cloned allele, after Steinkellner *et al*. (1997) and Kampfer *et al*. (1998); ‡Calculations based on individuals from Portugal, Spain, France Slovenia and Serbia only. Locus 7 is from Steinkellner *et al*. (1997), while Locus 7§ is from Kampfer *et al*. (1998). Locus 16 amplifies locus 3/64 and vice versa (this corrects a minor error in Steinkellner *et al*. (1997)). Locus 962 is located 71 base pairs downstream of locus 96. H_O = observed heterozygosity, H_E = expected heterozygosity, V = variance in repeat number and n = number of analysed individuals.

eastern France (*Q. petraea* 48°18′ N 4°28′ E, *Q. robur* 48°56′ N 5°01′ E), Serbia (*Q. petraea* 45°20′ N 19°50′ E, *Q. robur* 44°59′ N 19°23′ E) and Slovenia (*Q. petraea* 46°36′ N 16°20′ E, *Q. robur* 46°25′N 16°10′E). Individuals were also sampled randomly in mixed populations from Ireland (53°03′ N 8°24′ W, predominantly *Q. petraea*, 53°04′ N 8°50′ W, predominantly *Q. robur*). These leaves were then classified morphologically in the laboratory according to: petiole length (expressed as a ratio of leaf lamina length), basal shape of the lamina and abaxial laminar pubescence. This sampling scheme allows a biogeographical perspective to be given to the interpretation of the genetic differences among the species. A total of 12 populations were genotyped with 20 microsatellite loci. These loci were derived in part from Steinkellner *et al*. (1997) and in part from Kampfer *et al*. (1998). PCR (polymerase chain reaction) primers for three loci (36, 58 and 102) were redesigned using the OLIGO (National Biosciences, Inc.) primer analysis software. These primer sequences were as follows: 36f 5′-GAT CAA AAT TTG GAA TAT TAA-3′, 36r 5′-TCG TGG AGG TTA GTC CCN TTT-3′; 58f 5′-AAT TGA GAG TGA CAG AAA GAG-3′, 58r 5′-TTC TTT TTC CTA ATC TCA ACT-3′ and 102f 5′-AAG CTT TCC AAT TGC ATA AAC-3′, 102r 5′-ATG TAC AAT GTG TTG ACT ACT-3′. Ten microsatellite loci used in this study have been mapped to six of 12 haploid chromosomes in *Q. robur* by Barreneche *et al*. (1998). Linkage groups for these loci and the repeat units/motifs for each locus are shown in Table 2. In addition to the 20 microsatellites genotyped, one neighbouring microsatellite to locus 96 was amplified using a nested PCR to eliminate PCR artefacts. PCR products from external primers 962f 5′-GTG ATG AAT TGG ACA GTA G-3′, 962r 5′-CTA AAA CTG NTT CTC TTT C-3′ were reamplified with an internal primer 962f int. 5′-GAT GTG ACT GGG AAT TTC-3′ and primer 962r.

For loci with high F_{ST} values (at least twofold higher than the average F_{ST}), additional individuals from both species in four Italian populations were genotyped to check the presence of specific alleles in this refugial peninsula. *Q. petraea* individuals were sampled from Monte San Nicolao 44°14′ N 9°42′ E and Monti della Laga 42°41′ N 13°36′ E (n = 17), while *Q. robur* individuals were sampled from Torbiera di Trana 45°02′ N 7°25′ E and Castel Porziano 41°45′ N 12°25′ E (n = 14) (Fig. 1, Table 1).

Microsatellite analysis

Microsatellite loci were amplified following standard protocols (Schlötterer 1998). Briefly, end labelled (^{32}P) PCR primers were used in a 10 μL reaction volume (1.5 mM $MgCl_2$, 200 μM, dNTPs, 1 μM of each primer, 20–100 ng template DNA and 0.5 U *Taq* polymerase). The cycling profile consisted of an initial denaturation step of 3 min followed by 35 cycles of 1 min at 94 °C, 1 min at 45–60 °C (depending on the primer combination) and 1 min at 72 °C. A final extension at 72 °C for 45 min was used to assure a quantitative terminal transferase activity of the *Taq* polymerase. PCR products were then separated on a 7% denaturing polyacrylamide gel (32% formamide, 5.6 M urea) and sized by running a two-base pair (bp) 'slippage-ladder' as well as a PCR product of a M13 clone of known length next to the amplified microsatellites (Schlötterer & Zangerl 1999).

Cloning and sequencing

Microsatellite loci with a base substitution interrupting the repeat unit have a lower mutation rate (Jin *et al*. 1996; Petes *et al*. 1997). To exclude the possibility that a base substitution interrupting a microsatellite was responsible for any differential patterns of variability in either species, we sequenced a sample of alleles ($n = 37$) from individuals across Europe at the three most differentiated loci (as determined by F_{ST}), and two reference loci (exhibiting low F_{ST} values). Homozygous individuals were amplified and sequenced in both directions. Direct sequencing was performed using BigDye terminator chemistry (Applied Biosystems) according to manufacturer's instructions and run on an ABI 377 automated sequencer.

To determine the ancestral character states of base substitutions within the microsatellite array and their flanking regions, PCR products of the three most differentiated loci and two reference loci were amplified and subsequently cloned into M13mp19 (Yanisch-Perron *et al*. 1985). These products were amplified from single individuals of *Quercus cerris* L. and *Quercus palustris* Muench (growing in the Vienna region). *Q. palustris* is regarded as a closer relative to *Q. petraea* and *Q. robur* than *Q. cerris* (Manos *et al*. 1999).

The standard 20 μL ligation mix contained 20–100 ng of phosphorylated PCR product, 100 ng M13mp19, 1 U T4 ligase (Promega) and 1 × T4 ligation buffer (Promega). Ligation was carried out at 18 °C overnight. Clones were transformed into XL1 blue *Escherichia coli* competent cells (Stratagene). Clones carrying inserts were identified with blue/white selection. Sequencing templates were prepared from overnight cultures of positive clones using standard protocols (Sambrook *et al*. 1989). Multiple clones from each locus were sequenced using BigDye chemistry.

Data analyses

Individuals from different geographical populations were pooled into two groups based on phenotypic classification, one for *Q. petraea* and one for *Q. robur* (Muir *et al*. 2000). Diversity indices and general statistics were calculated using MSANALYSER 3.12 (Dieringer & Schlötterer 2003). Using the same program, we determined Θ, an unbiased estimate of Wright's Fixation Index (Weir & Cockerham 1984), to identify the most informative loci separating the two species. The significance of pairwise Θ-values (referred to as F_{ST} in the text) was tested by permuting genotypes among groups. This method of permutation does not rely on Hardy–Weinberg assumptions (Goudet *et al*. 1996).

A hierarchical analysis of molecular variance (AMOVA) using the program, ARLEQUIN (Schneider *et al*. 2000) was employed to examine the partitioning of microsatellite variance into components derived from species differences and those derived from geographical location. The total variance was partitioned into covariance components resulting from: within populations, among populations within species and among species.

The average repeat length per locus was calculated either indirectly by subtracting the flanking region or directly, when the microsatellites had been sequenced. Estimates of θ (the neutral mutation parameter $4N_e\mu$, where N_e is the effective population size and μ is the neutral mutation rate) were calculated in DNASP 3.52 (Rozas & Rozas 1999) from the average pairwise nucleotide diversity (π; Tajima 1983).

Testing for gene flow in a biogeographical setting

Gene flow was analysed in a biogeographical setting by comparing the F_{ST} values of neighbouring species pairs (focal groups) with more distantly located pairs (as a result of the geographical distance between populations, the Spanish/Portuguese species pair was omitted). One F_{ST} value was chosen randomly from a distantly located species pair to use as an independent distance measure with respect to each of the five focal groups. Five F_{ST} values (one from each of the focal groups) and five values from long-distance comparisons were used to compare differences.

Bayesian admixture analysis

A model-based clustering method implemented in the program STRUCTURE (Pritchard *et al*. 2000) was used to assign individuals probabilistically to homogenous clusters (K populations) without consideration of sampling localities. Estimated posterior probabilities for the simulated model fitting the data were calculated assuming a uniform prior for K, where $K \in \{1, 2, 3, 4\}$. To minimize the effect of the starting configuration during the Monte Carlo simulation,

we simulated 50×10^3 updates of the Markov chain before data for the parameter estimation were collected from another 10^6 iterations. Three independent runs of the Markov chain, each of least 10^6 updates were performed to assure convergence of the chain and homogeneity among runs for each prior of K. The posterior probabilities of K were then calculated using Bayes' rule. The program was run without population identifiers (USEPOPINFO = 0) and in the admixture mode (NOADMIX = 0). As larger sample sizes were available in Ireland, we used one population from each species located 29 km apart. A third approximately equidistant population of *Q. petraea* was added as a 'distance' control. If demography rather than species differences was responsible for any population structure, then the prediction for K would be three rather than two. A minimum of 24 individuals (83 in total) from each Irish population was genotyped.

The species, populations and individuals used for each analysis, respectively, are shown in Table 1.

Results

Twenty polymorphic microsatellites were screened for both species in six populations covering the species distribution in Europe (Fig. 1, see also Muir *et al*. 2000). Genetic differentiation among both species was low ($F_{ST} = 0.050$), but highly significant ($P < 0.0001$) when all loci were considered jointly. A separate analysis of individual loci indicated that all 20 microsatellite loci detected a significant ($P < 0.05$) differentiation between *Quercus petraea* and *Quercus robur* (Table 3). To discount the possibility that this differentiation merely reflects substructure within species, we took advantage of the hierarchical sampling while partitioning the total variance according to species and geographical location. This hierarchical analysis (AMOVA) indicated that 15 of the 20 loci were significantly different among species. Given the moderate sample sizes for individual populations, the slight reduction in the number of significant loci under the AMOVA is not unexpected. Ten of the microsatellite loci surveyed in this study are mapped in *Q. robur* (Barreneche *et al*. 1998). Based on these loci, we observed significant differences among the two species for six (out of 12) haploid chromosomes. Both analyses (F_{ST} and AMOVA) suggest that the differentiation between the two species is not limited to a few loci, but is spread throughout the genome.

Testing for gene flow in a biogeographic setting

To investigate the extent of gene flow and drift among the two species, we compared the F_{ST} values of neighbouring species pairs with more distantly located pairs. If gene flow is an important evolutionary force, then the differentiation of neighbouring species pairs (e.g. *Q, robur* vs. *Q. petraea* in Ireland, 29 km apart), should be lower than for distantly located pairs (e.g. *Q. robur* in Ireland vs. *Q. petraea* in Slovenia, 896 km apart). However, we found no significant difference among neighbouring species pairs ($\bar{F}_{ST} = 0.082$) and more distant species pairs ($\bar{F}_{ST} = 0.104$, $P = 0.42$ Mann–Whitney U-test). Within the limits of low statistical power (limited number of species pairs), this is an observation which suggests low differentiation among populations within species and also low rates of gene flow between species.

Table 3 Pairwise F_{ST} values between *Quercus petraea* and *Quercus robur* for the 20 investigated microsatellite loci

Locus	Wright's fixation index (F_{ST})	P-values (10 000 replicates)
110	0.023	< 0.0001
1/5	0.032	< 0.0001
7	0.110	< 0.0001
16	0.022	< 0.0001
9	0.017	< 0.0001
119	0.056	< 0.0001
15	0.011	0.0205
1/2	0.054	< 0.0001
46	0.062	< 0.0001
*7	0.023	< 0.0001
87	0.067	< 0.0001
96	0.232	< 0.0001
112	0.162	< 0.0001
108	0.036	< 0.0001
101	0.012	0.0031
20	0.017	0.0003
31	0.010	0.0008
58	0.041	< 0.0001
36	0.040	< 0.0001
102	0.008	0.0017
All	0.050	< 0.0001

Locus 7 is from Steinkellner *et al*. (1997) while Locus *7 is from Kampfer *et al*. (1998).

Estimates of admixture in mixed populations

To further substantiate the absence of gene flow among the two species, we derived estimates of admixture probabilistically using Bayesian statistics in individuals from both species sampled in Ireland. In total, 83 individuals were genotyped from three equidistant (*c*. 40 km) populations. Based on phenotypic classification, one population sample consisted only of *Q. petraea* individuals, while the *Q. robur* population contained 10% *Q. petraea* individuals. The third population was predominantly *Q. petraea* with 1% *Q. robur*. This mixture allowed interspecies gene flow to occur within populations as well as among populations. Hence, the sampled individuals should reflect the ongoing level of hybridization between the two species. The highest posterior probability was obtained for two clusters ($K = 2$, Table 4)

Table 4 Estimated posterior probabilities of K assuming four genetic clusters

K*	ln $P(X \mid K)$	$P(X \mid K)$†
1	–7409.7	~0
2	–7176.0	~1
3	–7196.2	~0
4	–7200.5	~0

Calculations based on 83 individuals of *Quercus petraea* and *Quercus robur* from Ireland. *Number of clusters assumed using STRUCTURE. †Assuming a uniform prior for K, where $K \in \{1, 2, 3, 4\}$.

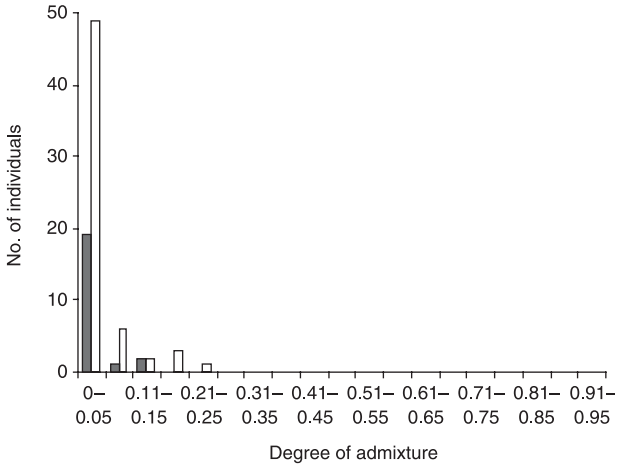

Fig. 2 Frequency distribution of admixed genotypes in two populations of *Quercus petraea* and one population of *Quercus robur* from Ireland. Admixture proportions are for an optimal $K = 2$ (Table 4), determined using the program, STRUCTURE. The degree of admixture for each individual represents the frequency of non-*Q. petraea* alleles (unshaded) and non-*Q. robur* alleles (shaded).

rather than three, confirming that species differences rather than demography were responsible for the observed population structure. Using the admixture model in the STRUCTURE program, we also estimated the probability of ancestry for each individual in one of two groups: *Q. petraea* vs. *Q. robur*. The majority of individuals had high probabilities of ancestry in only one of the two groups corresponding to their phenotype. A low number of individuals were found to score in both groups. Hence, we detected little evidence of admixture among the two species. The degree of individual admixture varied between 0.5% and 20.3% (mean admixture coefficient = 0.037, Fig. 2). No individual had admixture coefficients larger than 21%, a proportion that is unlikely to have been realized by an F_1 hybrid. Furthermore, the individual with the highest admixture proportion occurred in the only population consisting of phenotypically pure *Q. petraea* (i.e. the population with the lowest opportunity for interspecific gene flow). Hence, this analysis supports the low rates of gene flow inferred from the comparison of F_{ST} values among neighbouring species pairs and distantly located pairs.

Loci displaying high F_{ST} *values*

A closer inspection of the locus-specific F_{ST} values revealed three loci (96, 112 and 7), which had values up to one order of magnitude higher than the remaining 17 (Table 3). As such loci may be good candidates for genomic regions carrying genes responsible for adaptive differences between the two species, we characterized these loci at the sequence level. For comparison, we included two loci that had low F_{ST} values (9 and 110). Several alleles were sequenced covering the allelic spectrum in both species. Cloned sequences from *Q. palustris* and *Q. cerris* individuals were used to infer the ancestral character states. Interestingly, for all loci sequenced, we failed to detect species-specific alleles. All sequence variants were shared among *Q. petraea* and *Q. robur*, a pattern that is consistent with either shared ancestral sequence polymorphism or gene flow among the two species. Hence, based on the sequence data, no unambiguous differentiation between the two species for all five loci emerged. A formal test for differentiation is not possible, as the sequenced alleles were selected based on their size class and occurrence in the two species.

Four of the sequenced loci had only low levels of polymorphism in the flanking region (π varied between 0 and 0.0056 within species). In contrast, locus 7 exhibited high levels of polymorphism among the sequenced alleles. Excluding the microsatellite region, the average pairwise nucleotide difference within *Q. petraea* and *Q. robur* sequences was 0.057 and 0.063, respectively. Comparison to a more distant relative, *Q. cerris*, revealed that this outgroup was identical to some *Q. petraea*/*Q. robur* sequences, indicating that the second class predates the split of *Q. petraea*/*Q, robur* (see Appendix for table of polymorphic sites).

Polymorphism at microsatellite loci with high F_{ST} *values*

Two of the three loci with the highest F_{ST} values also showed the most pronounced differences in heterozygosity among the two species. Locus 96 and 112 differed by 0.48 and 0.45, respectively, while the average difference across loci was 0.11. Locus 96 had a reduced heterozygosity in *Q. robur* while locus 112 had low levels of variability in *Q. petraea*. A separate analysis of the individual populations of each species indicated that this reduction in heterozygosity is a species-wide phenomenon (Table 5). Gene diversity for most geographical populations shows a pattern of lower variation in one species at these loci across Europe. This pattern was confirmed in other refugial populations additionally sampled from Italy (Table 5).

The statistical significance of the reduction in variability was evaluated by using the lnRH test statistic, which is

Table 5 Diversity indices by species and population for locus 96 and 112

Locus 96	H_O*	H_E†	V‡	No. of alleles	n§
Species/population					
Q. robur Ireland	0.50	0.55	24	7	24
Q. petraea Ireland	0.73	0.89	23	16	30
Q. robur Portugal	0.29	0.37	21	6	17
Q. petraea Spain	0.81	0.85	28	10	16
Q. robur sw France¶	0.29	0.33	23	6	14
Q. petraea sw France	0.82	0.84	25	8	11
Q. robur eastern France	0.50	0.54	26	6	14
Q. petraea eastern France	0.93	0.91	17	10	14
Q. robur Italy	0.38	0.41	33	5	13
Q. petraea Italy	1.00	0.91	27	13	16
Q. robur Slovenia	0.06	0.32	20	4	17
Q. petraea Slovenia	0.67	0.84	17	9	12
Q. robur Serbia	0.33	0.36	14	5	15
Q. petraea Serbia	0.88	0.92	38	12	17

Locus 112	H_O	H_E	V	No. of alleles	n
Species/population					
Q. robur Ireland	0.70	0.74	11	8	24
Q. petraea Ireland	0.20	0.25	1	5	30
Q. robur Portugal	0	0	0	1	18
Q. petraea Spain	0.12	0.11	0	2	17
Q. robur sw France	0.60	0.51	5	5	15
Q. petraea sw France	0.36	0.32	0	3	11
Q. robur eastern France	0.93	0.80	17	7	15
Q. petraea eastern France	0.67	0.52	0	3	15
Q. robur Italy	0.71	0.89	28	11	14
Q. petraea Italy	0.57	0.54	3	7	14
Q. robur Slovenia	0.81	0.87	13	9	16
Q. petraea Slovenia	0.31	0.35	2	4	13
Q. robur Serbia	0.80	0.89	15	9	15
Q. petraea Serbia	0.18	0.27	1	3	17

*H_O, observed heterozygosity; †H_E, expected heterozygosity; ‡V, variance in repeat number; §n = number of analysed individuals; and ¶sw, southwestern.

specifically designed to identify loci with a reduction in variability below neutral expectations (Kauer *et al*. 2003; Schlötterer & Dieringer in press). Both loci (96 and 112) with pronounced differences in heterozygosity between *Q. petraea* and *Q. robur* significantly deviated from the remainder of the genome ($P = 0.016$ and $P = 0.030$, respectively).

In each case, for the species displaying reduced polymorphism, the predominant allele class coincided with short microsatellite alleles. The 136 allele at locus 96 consisted of six repeats, while the 88 allele at locus 112 consisted of nine repeats (Fig. 3, Appendix). The same allele structure was found in both species. The high frequency of these shorter alleles resulted in a smaller average repeat number in one species at these two loci (Table 2). Given that the microsatellite mutation rate is highly correlated with repeat number (Brinkmann *et al*. 1998), the low variability in one of the species could be explained by the fixation of a short allele by genetic drift and an associated low mutation rate, rather than a selective sweep. To test this, we analysed an additional microsatellite locus in 158 individuals (Table 1, Fig. 3), which is located 71 bp downstream of locus 96. If the significant lnRH value is an artefact of different mutation rates in the two species, a flanking locus should show neither the same pattern of reduced variability nor a high F_{ST} value. However, we detected a high F_{ST} value (0.33, $P < 0.0001$) between the two species at this flanking locus and, in common with its neighbour (96), low gene diversity in *Q. robur* (Table 2). The mean repeat number was almost identical in both species (locus 962, Table 2). Hence, while the differentiation between the species at locus 96 and 112 is associated with the high frequency alleles with shorter repeats, the differentiation is most likely not an artefact of the shorter repeat length in one species. Rather, the strong reduction in variability may be caused by a selective sweep.

Discussion

We employed a microsatellite screen to investigate the magnitude and extent of genome differentiation among the two species. Consistent with previous allozyme, RAPD (randomly amplified polymorphic DNA) and AFLP (amplified fragment length polymorphism) studies (Zanetto *et al*. 1994; Bodénès *et al*. 1997; Coart *et al*. 2002), we observed that most of the variation was shared between the two species. In the following, we discuss several scenarios and the extent to which they may account for the high proportion of shared alleles between *Q. robur* and *Q. petraea* and for the significant genetic differentiation between the two species.

High rates of gene flow

Previously, Streiff *et al*. (1999) documented relatively high rates of gene flow among the two species. Interspecific hybridization was detected in 23 out of 310 offspring (7% of cases in a parentage analysis of a mixed population). Such high rates of gene flow are expected to prevent genetic differentiation (Slatkin 1987). Nevertheless, we observed a significant differentiation between the two species not only for the joint analysis of all loci, but also for the majority of individual loci. This result clearly indicates that the differentiation between *Q. petraea* and *Q. robur* is not confined to some regions of the genome, but is putatively a genome-wide effect. Consequently, our results are not compatible with high rates of gene flow between the two species and strong selection acting on a few loci. One possible explanation for this apparent discrepancy is that the study of Streiff *et al*. (1999) was based on acorns that have

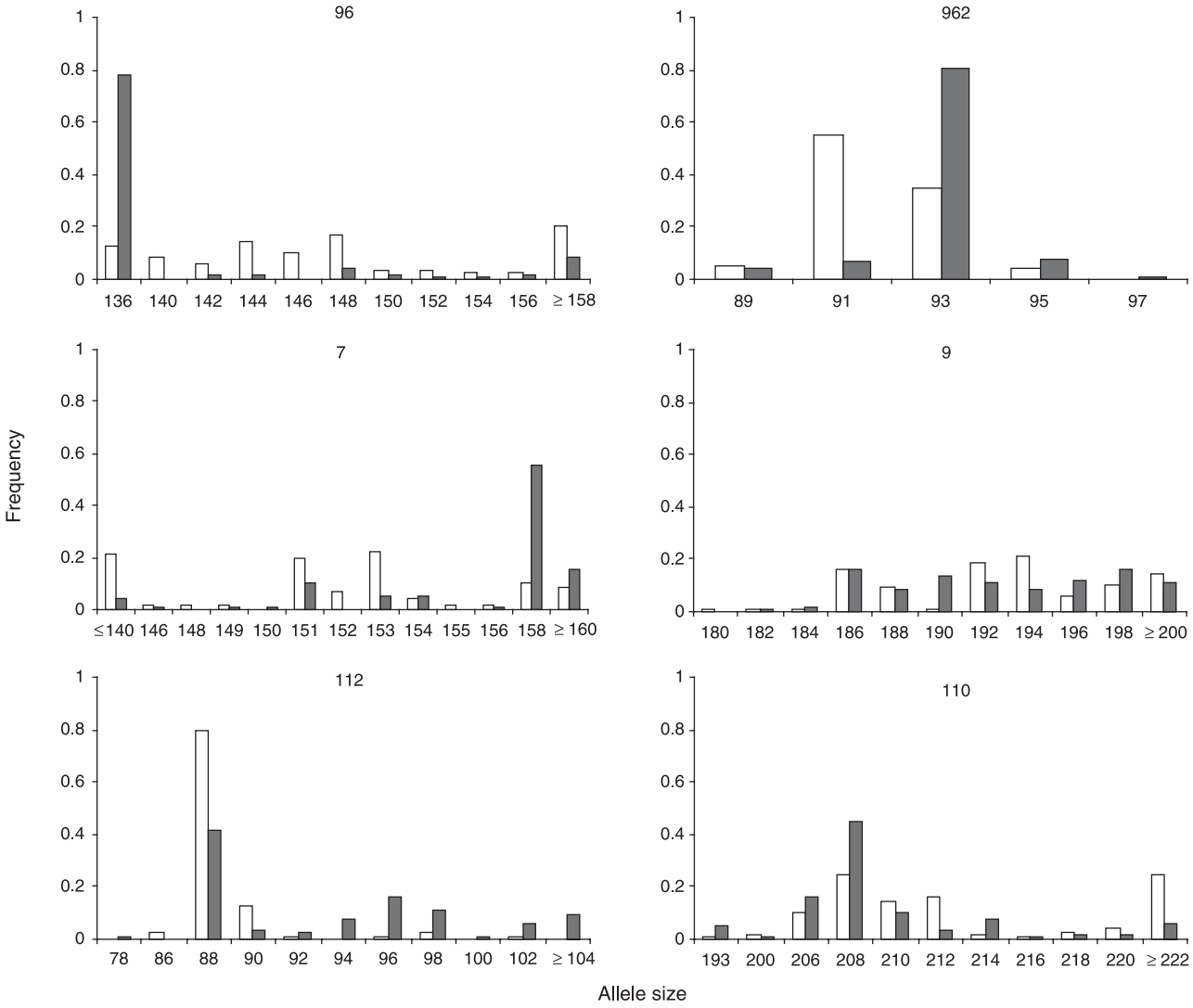

Fig. 3 Microsatellite allele frequency spectra at loci with high F_{ST} values and two reference loci (with low F_{ST} values: 9 and 110). Locus 962 is located 71 base pairs downstream of locus 96. *Quercus petraea* (unshaded), *Quercus robur* (shaded). Nucleotide sequences for a sample of these alleles (n = 37) are given in the Appendix.

not been subjected to environmental selection. Assuming postzygotic selection against hybrids in the wild, estimates derived from acorns may overestimate hybridization.

Low rates of hybridization

If genetic drift operates against the homogenizing effect of gene flow, the maintenance of a genome-wide difference between the two species could be explained by low rates of hybridization (e.g. Craft *et al*. 2002). While we note that the parameters for which such a scenario could be envisaged are not yet evaluated, for the sake of argument, we assume this is a realistic scenario. Given that a single migrant per generation is sufficient to prevent population differentiation by drift, even low levels of gene flow should be detectable. Hence, in the presence of low levels of gene flow, neighbouring populations should be more similar than distantly located ones. Our analysis did not indicate such an effect. Similarly, we detected very low levels of admixture between *Q. robur* and *Q. petraea* in three equidistant, partially mixed populations. While we cannot rule out that hybridization occurs at a low rate, we were not able to detect it in our analyses. To discount the possibility that our observations result from insufficient statistical power, more species pairs may be required to detect small but significant differences in F_{ST} among neighbouring and distantly located populations.

Shared ancestral variation

Allele sharing between two species may also result from shared ancestral variation. This phenomenon has been

extensively documented for closely related *Drosophila* species (Kliman & Hey 1993) and is particularly pronounced for recently diverged species with large effective population sizes (Pamilo & Nei 1988). Moreover, shared ancestral variation could also result in a pattern of low, but significant differentiation between species. Given that our analyses did not provide strong support for interspecific gene flow, shared ancestral variation should be seriously considered as an alternative explanation for the high frequency of shared alleles in combination with a genome-wide genetic differentiation among the two species. Support for shared ancestral polymorphism is provided by the DNA sequence analysis at locus 7. We detected substantial sequence divergence in the flanking region of this microsatellite and both species harboured divergent alleles. Furthermore, an allele from *Q. cerris* had more sequence similarity to one of the allelic classes than the other. These observations suggest that both *Q. petraea* and *Q. robur* have retained substantial levels of ancestral variation that could predate the species split to relatives such as *Q. cerris*.

The strongest evidence put forward for gene flow between *Q. petraea* and *Q. robur* is the phylogeographical pattern observed for the chloroplast. Spatially, individuals from both species share haplotypes more readily than would be expected by chance (Dumolin-Lapègue *et al*. 1997; Petit *et al*. 1997). This nonrandom distribution is difficult to reconcile with the shared ancestral polymorphism hypothesis. Nevertheless, gene flow is difficult to quantify in the chloroplast genome as the latter represents a single locus. Regardless of the number of sites assayed, all polymorphisms recover only one genealogy, the clonally inherited chloroplast (Dumolin *et al*. 1995). As such, the patterns of polymorphism at this locus may reflect selection acting on the chloroplast or even drift.

Microsatellites have a relatively high mutation rate and are prone to produce homoplastic alleles (Ellegren 2000; Schlötterer 2000). The identification of species-specific microsatellite alleles is thus complicated by alleles which may be identical in state rather than by descent. Nevertheless, in *Q. petraea* and *Q. robur*, loci with high F_{ST} values have on average, short microsatellites. Hence, if the split between the two species is relatively recent, any resultant homoplasy may be small. Equally, assuming a lower mutation rate for nucleotides, then the time for species-specific mutations to have accumulated is small and all mutations are thus shared.

Evidence for directional selection

Two loci had a particularly strong reduction in variability in one of the two species. As selective sweeps reduce variability at genomic regions flanking the selected site, such reductions in variability are often interpreted as evidence for the recent spread of a linked beneficial mutation (Maynard Smith & Haigh 1974; Schlötterer 2003). However, while the reduction in variability is species-specific, it occurs in all three glacial refugia. Hence, one has to assume that the spread of the beneficial mutation occurred before the beginning of the last glaciation and the contraction to refugia. Assuming a mutation rate for plant microsatellites, which is in the order of 10^{-4} per generation (Thuillet *et al*. 2002; Vigouroux *et al*. 2002), it appears unlikely that the acquisition of new mutations would not have erased the signal of a selective sweep. DNA sequencing, however, indicated that the most frequent allele at both loci had only a very small (six and nine) number of repeats in the species with low heterozygosity. Hence, these alleles are probably significantly more stable as microsatellite mutation rate is highly correlated with the number of repeats at a given allele (Weber 1990; Brinkmann *et al*. 1998). This observation could explain why the signature of a (putative) selective sweep is still detectable, even if it predates the last glaciation.

We note that one of the loci exhibiting a reduction in variability (96) maps to leaf quantitative trait loci differentiating the two species identified by Saintagne *et al*. (2004).

Acknowledgements

We are grateful to H. Steinkellner and C. Lexer for providing information on microsatellite loci and access to flanking sequences. Special thanks to C. Vogl, J.-M. Gilbert and R. Ennos for numerous helpful discussions and to the reviewers for constructive comments on the manuscript. DNA samples were kindly provided by C. Bodénès, A. Kremer, C. Lexer and A. Ungaro. Financial support was provided by the European Science Foundation and Fonds zur Förderung der wissenschaftlichen Forschung (FWF) to G.M. and FWF and European Union grants to C.S.

Supplementary material

MSANALYSER input file containing the original data (PCR product lengths) is available from: http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC2418/MEC2418sm.htm

References

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.

Barreneche T, Bodénès C, Lexer C *et al.* (1998) A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. *Theoretical and Applied Genetics*, **97**, 1090–1103.

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

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

Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *American Journal of Human Genetics*, **62**, 1408–1415.

Coart E, Lamote V, De Loose M *et al.* (2002) AFLP markers demonstrate local genetic differentiation between two indigenous oak species [*Quercus robur* L. & *Quercus petraea* (Matt.) Liebl.] in Flemish populations. *Theoretical and Applied Genetics*, **105**, 431–439.

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

Dieringer D, Schlötterer C (2003) MICROSATELLITE ANALYSER (MSA): a platform-independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.

Dumolin S, Demesure B, Petit RJ (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics*, **91**, 1253–1256.

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.

Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, **16**, 551–558.

Ferris C, Oliver RP, Davy AJ, Hewitt GM (1993) Native oak chloroplasts reveal an ancient divide across Europe. *Molecular Ecology*, **2**, 337–344.

Gömöry D, Yakovlev I, Zhelev P, Jedináková J, Paule L (2001) Genetic differentiation of oak populations within the *Quercus robur*/*Quercus petraea* complex in Central and Eastern Europe. *Heredity*, **86**, 557–563.

Goudet J, Raymond M, de Meeüs T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933–1940.

Howard DJ, Preszler RW, Williams J, Fenchel S, Boecklen WJ (1997) How discrete are oak species? Insights from a hybrid zone between *Quercus grisea* and *Quercus gambelii*. *Evolution*, **51**, 747–755.

Jin L, Macaubas C, Hallmayer J, Kimura A, Mignot E (1996) Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 15285–15288.

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

Kauer MO, Dieringer D, Schlötterer C (2003) A microsatellite variability screen for positive selection associated with the 'out of Africa' habitat expansion of *Drosophila melanogaster*. *Genetics*, **165**, 1137–1148.

Kliman RM, Hey J (1993) DNA sequence variation at the period locus within and among species of the *Drosophila melanogaster* complex. *Genetics*, **133**, 375–387.

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.

Manos PS, Doyle JJ, Nixon KC (1999) Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution*, **12**, 333–349.

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favorable gene. *Genetical Research*, **23**, 23–35.

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

Muir G, Fleming CC, Schlötterer C (2001) Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. & *Quercus robur* L. *Molecular Biology and Evolution*, **18**, 112–119.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568–583.

Petes TD, Greenwell PW, Dominska M (1997) Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. *Genetics*, **146**, 491–498.

Petit RJ, Pineau E, Demesure B *et al.* (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 9996–10001.

Ponton S, Dupouey JL, Bréda N *et al.* (2001) Carbon isotope discrimination and wood anatomy variations in mixed stands of *Quercus robur* and *Quercus petraea*. *Plant, Cell and Environment*, **24**, 861–868.

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

Rozas J, Rozas R (1999) DNASP Version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15**, 174–175.

Saintagne C, Bodénès C, Barreneche T *et al.* (2004) Distribution of genomic regions differentiating oak species assessed by QTL detection. *Heredity*, **92**, 20–30.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.

Schlötterer C (1998) Microsatellites. In: *Molecular Genetic Analysis of Populations: a Practical Approach* (ed. Hoelzel RA), pp. 237–261. Oxford University Press, Oxford.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma*, **109**, 365–371.

Schlötterer C (2003) Hitch-hiking mapping — functional genomics from the population genetics perspective. *Trends in Genetics*, **19**, 31–38.

Schlötterer C, Dieringer D (in press) A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: *Selective Sweeps* (ed. Nurminsky D). Landes Bioscience, Georgetown, Texas.

Schlötterer C, Zangerl B (1999) The use of imperfect microsatellites for DNA fingerprinting and population genetics. In: *A Laboratory Guide to DNA Fingerprinting/Profiling* (eds Epplen JT, Lubjuhn T), pp. 153–165. Birkhäuser, Basel, Switzerland.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN: a software for population genetics data analysis. In: *Genetics and Biometry Laboratory*. Department of Anthropology, University of Geneva, Switzerland.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.

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

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Thuillet A-C, Bru D, David J *et al.* (2002) Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum*

turgidum (L.) Thell. ssp *durum* Desf. *Molecular Biology and Evolution*, **19**, 122–125.

Vigouroux Y, Jaqueth JS, Matsuoka Y *et al*. (2002) Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution*, **19**, 1251–1260.

Weber JL (1990) Informativeness of human $(dC\text{-}dA)_n(dG\text{-}dT)_n$ polymorphisms. *Genomics*, **7**, 524–530.

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

Yanisch-Perron C, Vieria J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequence of M13 mp18 and pUC19 vectors. *Gene*, **33**, 103–119.

Zanetto A, Roussel G, Kremer A (1994) Geographic variation of inter-specific differentiation between *Quercus robur* L. & *Quercus petraea* (Matt.) Liebl. *Forest Genetics*, **1**, 111–123.

Graham Muir has diverse interests in plant population genetics and genome evolution. Currently, his work focuses on speciation and gene flow within the *Quercus* genus. Christian Schlötterer uses molecular markers to identify and characterize mutations in natural populations conferring a selective advantage in novel environments.

Appendix

Tables of polymorphic sites

Locus 96 (175 bp)

Allele size		pet_n	rob_n	*n*
136	tgg-TGAGAGAGAGAG----------	4	16	
140	...-AGAGAGAGAGAGAGAG----------	2		
142	...-AGAGAGAGAGAGAGAGAG----------	2		
144	...-AGAGAGAGAGAGAGAGAGAG----------	3		
148	...-TGAGAGAGAGAGAGAGAGAGAGAG----------	6	1	
168	...-TGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG---	2		
Q. pal.	...-TGAGAGAGAGAGAGAGAGAG----------aag			(5)
Q. cer.	aattTGAGAGAGAGAGAGAG----------aag			(6)

Locus 112 (106 bp)

Allele size		pet_n	rob_n	*n*
88	AGAGAAAA-GAGAGAGAGAGAGAGAG----------	5	5	
90	AGAGAAAAAGAGAGAGAGAGAGAGAG----------	2		
94	AGAGAGAGAGAGAGAGAGAGAGAGAGAGAG----------		2	
96	AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG---------		1	
98	AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG---------		5	
104	AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG---		1	
Q. pal.	AGAGAAAAGGAGAGAGAGAGAGAGAAAGAG----------			(5)
Q. cer.	AGAGAAAGGGAGAGAGAGAGAG----------taa			(3)
Q. cer.	AGAGAAAGAGAGAGAGAGAGAGAG----------taa			(2)

Locus 7 (204 bp)

Allele size		pet_n	rob_n	*n*
131	AAAAAAAAGAAAAAAAGAAAA-----GAGAGAGAGAGAGAG----------cccgtt	7		
132	**********GAAAAGAAAAGAGAGAGAGAGAG----------.....		1	
140	AAAAAA--GAAAA---GAAAAGAAAAGAGAGAGAGAGAGAGAGAGAG----------.....c	4		
151	AAAAAAAAGAAAAA--GAAAA-----GAGAGAGAGAGAGAG----------aagaacccagaacatcagag.....	3	4	
151	AAAAAAAAGAAAAA--GAAAA-----GAGAGAGAGATAGAG----------aagaacccagaacatcagag.....	2	1	
153	**********GAAAAGAAAAGAGAGAGAGAGAG----------aagaacccagaacatcagag.....	2	1	
158	AAAAAAA-GAAAA---GAAAAGAAAAGAGAGAGAGAGAGAGAG----------aagaacccagaacatcagagt.ta.c	3	10	
158	**********GAAAAGAAAAGAGAGAGAGAGAGAG----------aagaacccagaacatcagag.tt..c	1	1	
163	AAAAAA--GAAAA---GAAAAGAAAAGAGAGAGAGAGAGAGAGAGAG----------aagaacccagaacatcagag.....c		4	
177	**********----------aagaacccagaacatcagag..t..c		3	
181	**********----------aagaacccagaacatcagag..t..c	1		
181	AAAAAA--GAAAA---GAAAAGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG----------aagaacccagaacatcagag.t...c	1	1	
183	AAAAAA--GAAAA---GAAAAGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG----------aagaacccagaacatcagag.t..cc		1	
Q. cer.	AAAAAAAA-AAAAAAA----------GAGAGAGAGAGAGAG----------gaacccagaacatcagagaagaacccagaacatcagag.....			(5)

Appendix *Continued*

Locus 9 (264 bp)

```
Allele size                                                                                                              petn  robn  n
186      aAGAGAGAGAGAGAGAGAG-------------------tcagtc-----------------------------------------------------------acca-ggca   3     1
190      .AGAGAGAGAGAGAGAGAGAGAG---------------......-----------------------------------------------------------....-....         2
192      .AGAGAGAGAGAGAGAGAGAGAGAG-------------......-----------------------------------------------------------....-....   4     1
194      .AGAGAGAGAGAGAGAGAGAGAGAGAG-----------......-----------------------------------------------------------....-....         3
196      .AGAGAGAGAGAGAGAGAGAGAGAGAGAG---------......-----------------------------------------------------------....-....         4
198      .AGAGAGAGAGAGAGAGAGAGAGAGAGAGAG-------......-----------------------------------------------------------....-....   1     1
200      .AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG-----......-----------------------------------------------------------.....-...         1
Q. pal.  tAGAGAGAGAGAGAGAGAGA-------------cacag......agtcccttctcagggctgggcttggctggcctaataataactttatttggtttagaggggga..a...g           (2)
Q. cer.  .AGAGAGAGAG----------------------------......agttccttctctgggctggcct------------aataattgtatttggtttagaggggga----aa.           (6)
```

Locus 110 (260 bp)

```
Allele size                                                                                                              petn  robn  n
193      ac-AGAGAGAGAGAGAG----------------------------------------------------------------a-ag                                     1
206      ..-AGAGAGAGATAGAG---------------------------------------------------aagggttttattt.-..                               5     1
208      ..-AGAGAGAGAGAGAG---------------------------------------------------aagggttttattt.-..                               4     9
212      ..-AGAGAGAGAGAGAG---------------------------------------------------aagggttttattt.-..                               1
228      ..-AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG--------------------------------aagggttttattt.-..                               1
Q. pal.  ..aAGAGAGAGAGAGAGAGAGAGAGCGAGAG--------------------------------------aagggttttatt-.-.a                                          (2)
Q. cer.  ta-AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG------------aagggttttattttac.                                          (3)
Q. cer.  ..-AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGaagggttttattttac.                                          (3)
```

Sequence alignment for a sample of microsatellite alleles at locus 96, 112, 7, 9 and 110 (see also Fig. 3). The complete microsatellite is shown in uppercase for each allele, respectively, while polymorphic sites upstream and downstream of the microsatellite are shown in lower case letters. At locus 112 and 110, the two *Q. cerris* (cloned) alleles correspond to heterozygous individuals. The number in brackets after the locus number is the sequence length. 'Allele size' refers to the PCR product length determined by microsatellite typing, 'pet_n' and 'rob_n' are the number of individuals sequenced from each species, respectively. The numbers in brackets under the heading '*n*' are the number of sequenced clones (always from single individuals). Invariant sites are marked with dots and gaps are marked with dashes. At locus 7, the upstream sites from the microsatellite are missing from the 132, 153 and also one of the 158 alleles (marked with *). In addition, an ambiguous repeat length is present for the 177 and 181 alleles despite cloning of these alleles (also marked with *).