Heredity **78** (1997) 433–444 *Received 28 May 1996*

Detection of genomic regions differentiating two closely related oak species *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L.

C. BODÉNÈS, S. JOANDET, F. LAIGRET† & A. KREMER*

I.N.R.A., Laboratoire de génétique et amélioration des arbres forestiers, B.P.45, F-33611 Gazinet Cedex and †I.N.R.A., Laboratoire de biologie cellulaire et moléculaire, B.P. 81, F-33883 Villenave D'Ornon Cedex, France

Genomic regions differentiating *Quercus petraea* and *Quercus robur* were detected by screening 2800 PCR amplification products using random primers on 22 trees of each species sampled in 11 natural populations. Only two per cent of the amplified fragments exhibited significant frequency differences between the two species and none of them was specific to a species. The nucleotide divergence between the two species estimated with RAPD data was 0.5 per cent in the overall genome and increased to 3.3 per cent in the discriminant regions. Twenty-three informative fragments were cloned and partially sequenced. New primers were derived from these sequences to obtain Sequence Characterized Amplified Region (SCAR) fragments. Southern blot experiments indicated that the SCARs were generally in low copy number in the genome. A search for similarity between SCAR sequences and sequences contained in data banks revealed that three of them corresponded to known DNA sequences.

Keywords: molecular differentiation, nucleotide divergence, oaks, RAPD, SCAR.

Introduction

Investigations on molecular differences between closely related species have usually been performed with two different approaches. The first one has attempted to quantify overall genomic changes between two genomes, as for example, the DNA–DNA hybridization technique (Sheldon, 1995) or the Representational Difference Analysis (RDA, Lisitsyn *et al.*, 1993). The second has aimed at calculating differences in any marker available irrespective of its implication in the speciation process. This has often been the case with allozymes, which are generally in limited number (Crawford, 1985) or with RFLP and RAPD data. For both approaches, genetic and evolutionary inferences have been restricted to the conclusions drawn from genetic distances calculated at the overall genome level (DNA–DNA hybridization) or at a few anonymous markers (allozymes, DNA markers). None of these methods has focused on the molecular organization and distribution of genomic regions specifically involved in species differentiation. Such a strategy involves detecting, as a first step, the informative regions, the 'hot spots' where higher interspecific molecular differences exist. Once they are detected, their organization in the genome can be depicted and they may be localized on a genetic map to analyse their distribution. This strategy opens new perspectives in evolutionary studies at the species level such as molecular analysis of speciation or interspecific gene flow. We report here on the detection of genomic regions differentiating two closely related white oak species.

Sessile (*Quercus petraea* (Matt.) Liebl.) and pedunculate (*Quercus robur* L.) oaks are two widespread European species. They are sympatric and generally occupy different but proximal ecological niches. Although leaf and fruit interspecific differences are clearly recognized (Dupouey & Badeau, 1993), the two species exhibit extremely low genetic differentiation (Müller-Starck *et al.*, 1993; Zanetto *et al.*, 1994). Gene diversity studies based either on allozymes (Zanetto *et al.*, 1994), or ribosomal DNA and chloroplast DNA (Petit *et al.*, 1993) have shown that both species share the same alleles, including rare alleles. This extremely low differentiation may originate from their interfertility. Extensive unidirectional hybridization has been shown in mixed stands (Bacilieri *et al.*, 1996) and confirmed by interspecific

*Correspondence. E-mail: antoine.kremer@pierroton.inra.fr

crosses (Steinhoff, 1993). The obvious discrepancy on the level of differentiation between the phenotypic and genetic levels raises the question of the molecular basis of species differentiation: how are the interspecific polymorphic regions distributed in the genomes of these two white oaks? Are they numerous or rare? Are they widespread or concentrated in a few groups? Answers to these questions require an intensive search of the genome in order to make an inventory of genomic regions that show species differentiation. We applied DNA amplification with randomly designed fragments (RAPD, Williams *et al.*, 1990) to identify the discriminant regions.

A first screening of the genome showed that only a few discriminant RAPDs were present (Moreau *et al.*, 1994). We extended the screening procedure to a larger coverage of the genome. The detection of informative RAPD fragments was then followed by their sequencing. New specific primers were derived and Sequence Characterized Amplified Regions (SCAR, Paran & Michelmore, 1993) were obtained. SCARs were then used as probes to analyse further the organization of genomic regions differentiating the two species. The objectives of these experiments were threefold: (1) to compare the level of differentiation between informative and noninformative genomic regions by deriving nucleotide divergence from RAPD data, (2) to evaluate the degree of repetition of informative regions in the genome and (3) to estimate the similarity of informative regions in the two species.

Materials and methods

Plant materials

The sample consisted of 22 populations of *Q. petraea* and *Q. robur* clustered in 11 geographical pairs (Fig. 1). These populations were distributed through much of the natural range of the species from Spain to Poland. Buds or leaves were collected on two trees per population for DNA extraction, on adult trees for pairs 10 and 11, and on seedlings grown in the nurseries for others pairs. One sample tree, which appeared to belong to a different species, had to be discarded during the experiment, leaving the total number of trees as 43 (22 *Q. petraea* and 21 *Q. robur*).

DNA extraction, RAPD amplification conditions

Genomic DNA was extracted from buds or young leaves frozen at −80°C according to Saghai-Maroof *et al.* (1984) with minor modifications (Moreau *et al.*, 1994). Amplification reactions were carried out in 15 μL reaction mixtures containing 16.6 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl pH 7.5, 2 mM $MgCl_2$, 0.0005 per cent anionic detergent, 0.2 per cent β-mercaptoethanol, 4.4 μM/mL bovine serum albumin, 100 μM each of dATP, dGTP, dCTP and dTTP, 0.2 μM of 10-base primers from Operon Technologies, 2–5 ng of template DNA and 0.8 U of Taq DNA polymerase (GibcoBRL) on a PHC3 Techne thermal cycler. PCR conditions were: 4 min at 94°C followed by 35 cycles, each of 45 s at 93°C, 45 s at 40°C and 1 min 30 s at 72°C. Amplification products were separated in 1.5 per cent agarose gels using a 0.5 × TBE running buffer at 3 V/cm. The gels were stained with ethidium bromide.

Cloning and sequencing of RAPD fragments

RAPD fragments to be cloned were rescued from the agarose gel by touching the corresponding DNA fragment with a pipette tip, rinsing the tip into 200 μL of sterile H_2O and carrying out a new PCR amplification using the RAPD assay conditions described above; two μL of the RAPD fragment sample were used as template. On some occasions, it was necessary to repeat this process several times until a pure single fragment was obtained. The amplified product was purified using Promega Wizard PCR preps and was subsequently cloned into the pGEM-T^R plasmid vector (Promega) following the manufacturer's instructions. The identity of the cloned RAPD products was verified by hybridization of the cloned fragments to Southern blots of RAPD fragments. The nucleotide sequences of the cloned RAPD fragments were obtained using the T7 sequencing kit (Pharmacia) and the universal forward and reverse primers.

SCAR design and analysis

For each RAPD fragment, three identical clones were sequenced. New primers (15–20 bp length) were derived from the sequences (Table 1). Amplification of genomic DNA with SCAR primers was performed as described above except for the primer concentration (0.1 μM) and for the annealing temperature (Table 1). The SCAR products were analysed on 1 per cent agarose gels in 0.5 × TBE running buffer. Following amplification, monomorphic fragments were digested with restriction enzymes or analysed with the Single Strand Conformation Polymorphism (SSCP) method (Bodénès *et al.*, 1996) to detect polymorphism.

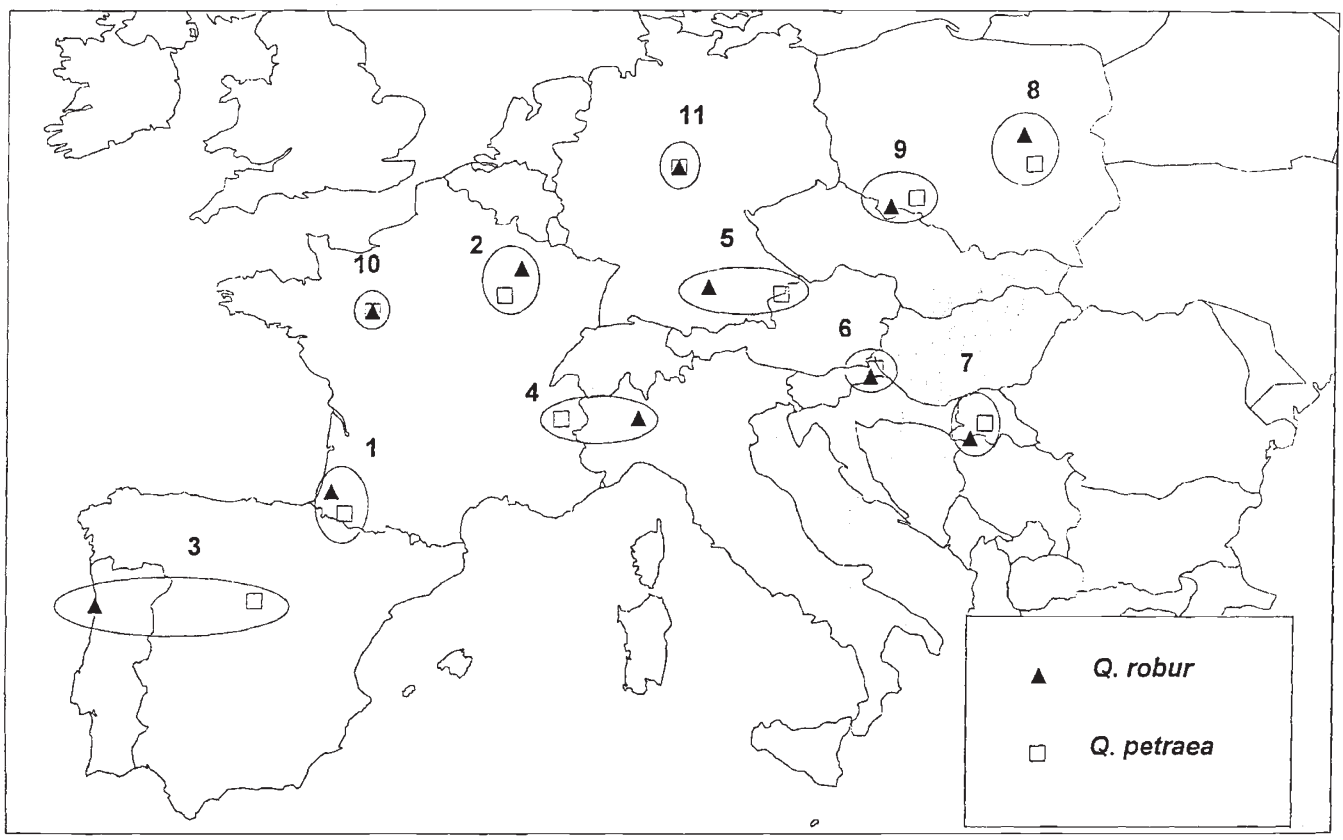

Fig. 1 Geographical distribution of *Quercus robur* and *Quercus petraea* populations within Europe.

Southern hybridizations

Restricted genomic DNA from *Q. petraea* and *Q. robur* or RAPD products was transferred onto N+ membrane (Appligene) according to the standard alkaline Southern blot procedure (Southern, 1975; Reed & Mann, 1985). After purification, SCAR fragments were used as probes as described above and labelled with α-^{32}P-dATP (3000 mCi/mmole) using the random primed DNA labelling kit (Biolabs). The probes were purified by gel-filtration chromatography on Sephadex G50 columns. Hybridizations were performed overnight at 42°C in 5 × SSC, 5 × Denhart's (Denhart, 1966), 0.05 per cent SDS, 0.1 M Tris-HCl, 50 per cent formamide and 300 μg/mL yeast RNA. Filters were washed for 15 min at 50°C twice in 2 × SSC, 0.1 per cent SDS and twice in 0.2 × SSC, 0.1 per cent SDS, rinsed 2 min in 2 × SSC and then autoradiographed using X-ray film.

Sequence analysis

SCAR sequences were compared to those existing in databases (Genbank/EMBL) to detect potential similarities. Nucleotide sequence similarities were searched using BLASTX and BLASTN (Altshul *et al.*, 1990) at the National Center of Biotechnology Information (Wilbur & Lipman, 1983).

Data analysis

For each sample, RAPD fragments were scored as present (1) or absent (0). The screening procedure was based on the identification of the primers which produced RAPD profiles allowing differentiation of the two species and giving reproducible results (Fig. 2). The Fisher exact test was used to identify discriminant fragments, i.e. fragments exhibiting frequency differences between the two species. Nucleotide diversity was inferred from RAPD data by using the method of Nei & Miller (1990) extended to diploid data.

Results

Identification of informative RAPD fragments

A total of 2800 amplified fragments, obtained with 250 primers, was analysed and 64 fragments (2 per cent), generated by 40 primers (14.4 per cent), were

Table 1 Primer sequences, PCR conditions for the SCAR amplifications and type of polymorphism obtained from oaks

SCAR name†	Sequence 5′ 3′‡	Product length (bp)	Annealing temperature	Type of polymorphisms§
A1 500	S: ATG AGT GGG ATG AAG AGC E: CAG TTG TAG AAC CAC CTG	470	56	M4*
A17 700	S: GAC CGC TTG TTG GTA E: GAC CGC TTG TGA TTA	700	55	P2
B11 1500	S: AGA CCC GTA GAG GAG ACA TT E: CCC GTG TAG TAT TCC ACA AA	1500	52	P2
B12 500	S: GTC TTC GAC TGG GGT GAA E: TCA GTG CAT TCC GAA AGA	520	55	M3
B12 750	S: GTT TAA GCC CAA TTT TTA TT E: TTT GAA GTT GAT ACA TAT TC	700	46	M4
B12 800	S: CCT TGA ACG CAT TAT GAC AT E: CCT TGA ACG CAG CAC AAT TC	800	48	P1
B19 800	S: ACC CCC GAA GTA GCG TTT CT E: CGT CAC CGA CGA CTG ATT CA	800	54	P1
E6 1600	S: ACC CCT CCC CTA AAT CTC TA E: AAC AAT GCA TAC CTC TAT GG	1600	57	M4*
F1 1000	S: CTT TAC AAA ATG GGT AGA GA E: CAG GTA TCC TTC CTT AAA TC	1000	55	M4*
F14 700	S: CAG AAG AAG CAA TGG TAA CA E: CAA CAT TTG GTG TGT CTT AG	750	50	M3
174 400	S: ATT TGG TTT TGG GCT ACA AC E: CGG AGC CAT ATT ATC TAC CT	420	55	M4*
I13 300	S: GTG TGG TGC AGA AAA E: ACT CCA GGT CTA TCC	260	52	P2
I13 500	S: CTG GGG CTG ACC AAC GTG AA E: ATT AAG GGT GGG TTG TGT GG	500	52	M5
I14 250	S: GCG TGG AGG TTG CCA CTG AT E: CGG CGT CAC TTC AAC TAA TG	250	57	M3
I14 780	S: GGC GGT ACA ACA AAC TTA TA E: CAG AGG TCC ATT TGA GAT TA	790	57	M3
I16 500	S: ATG GAA TAG TGA AAA GAG AC E: TAA AGC ACA TAA GAA TAA GA	450	48	M4
P14 450	S: TGC AAA ACA AAC TAA ACA TT E: AAT CCA GTG GCA AGT TTT AA	450	48	M3
P14 1000	S: ACT GTG CAA CAA CCG ACT CA E: TAT CGG AAA AAT CAT GTA CC	1000	52	P1
P17 1400	S: CGC ATG GAT ATA CTA TTC E: CTA CCT AGT GTC CTA TGG	1400	54	P1
R11 570	S: GGA GGT TAG GGC TTT TTA E: CTG TAC CAA ACA AGA AAC A	570	57	M5
R12 500	S: CGA GCG TTG ATA GCC AAT AG E: TCG GAA GCA AAA GGG TAA TT	500	51	P1
U1 500	S: GGA AGC TAA CAC AAA CTC E: ACG GAC GTC AAC ACA CTA	500	55	P1
U7 790	S: GCC CTA ACA AAT CAT CTC E: TAA TAT AGA AAG GGA AAG	790	48	M3

S corresponds to the upstream primer. E corresponds to the downstream primer.
†SCAR primer pairs are denominated according to the Operon primer used for the random amplification followed by the size of the discriminant fragment.
‡The underlined sequences represent the sequence of the original RAPD primer.
§The letters and the values correspond to the types of polymorphism as illustrated in Fig. 3.
*Indicates that these SCAR fragments have not been analysed by RFLP or SSCP.

selected because they showed significant frequency differences between the two species (Fisher exact test, data not presented). Among these primers, 34 gave reproducible results and produced 36 informative fragments among a total of 412. Seven fragments only among the 412 were monomorphic for the 43 trees tested.

Differentiation among the 43 sampled trees was analysed with Factorial Correspondence Analysis (FCA) (Lebart *et al.*, 1984) which is a weighted principal correspondence analysis. The calculations were performed with the 36 informative fragments selected during the screening procedure. The first axis of FCA explaining more than 31 per cent of the total variation clearly subdivides the 43 trees in two sets partly overlapping and corresponding to the two species (Fig. 2). This subdivision appears to be the most discriminant with the variables available. Indeed, an empirical statistical test showed that no other more discrete subdivision among the 43 trees was possible with the data set. This test was performed by randomly selecting 36 different variables among the 412 available and running the FCA analysis. Two thousand different random combinations of 36 variables were tested and none of them resulted in a higher discrimination than 19 per cent (per cent of the variation explained by the first axis of FCA). One may therefore conclude that (1) the highest differentiation among the 43 trees is the one corresponding to the 36 fragments obtained after the screening procedure and (2) this differentiation corresponds to the separation into two species (Fig. 2).

Nucleotide diversity and divergence between Q. petraea *and* Q. robur

Nucleotide diversity and divergence between the two species were estimated from the RAPD data (Table 2). Separate analyses were conducted with three different sets of fragments: (1) a total of 412 fragments resulting from the screening procedure (34 primers); (2) 376 fragments excluding the most discriminant fragments among the previous set; and (3) the 36 discriminant fragments only. Different values of the fixation index (from 0 to 0.20) were considered in concordance with previous results obtained with isozymes (Zanetto *et al.*, 1994). Fixation indices are assumed to be known to infer nucleotide diversity measures estimated from RAPD data. The level of the fixation index has no strong effect on the level of diversity and divergence (Table

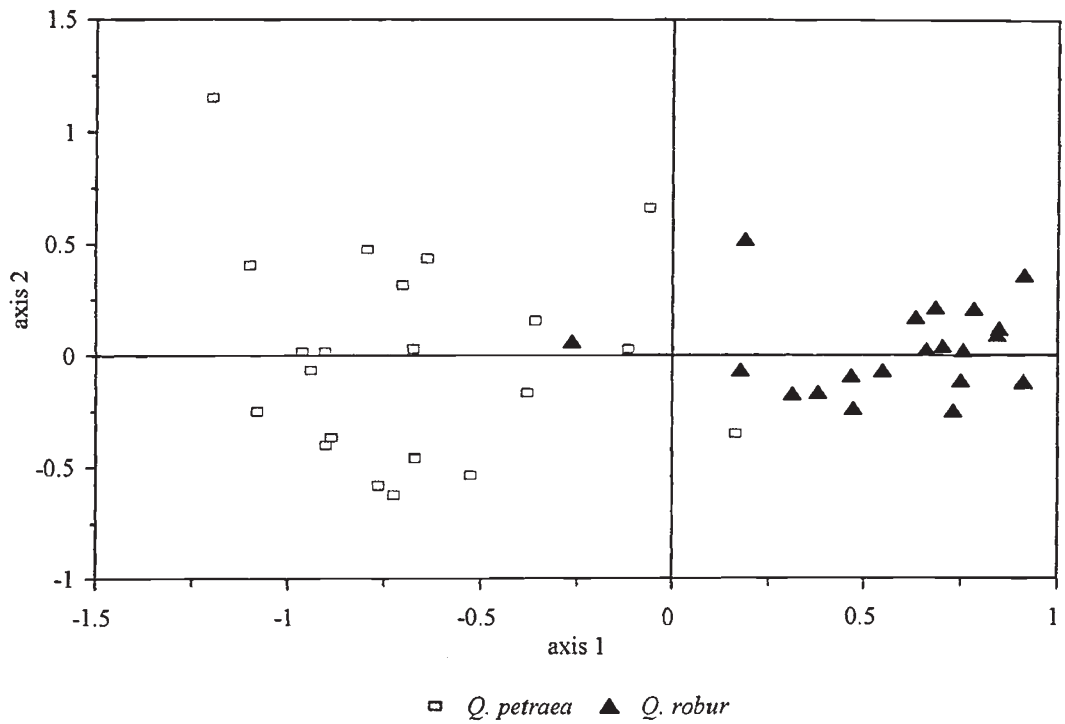

Fig. 2 Graph of the Factorial Correspondence Analysis (FCA). FCA was computed with 36 informative fragments. The first axis allows the separation of the two species and explains 31 per cent of the total variation. In each species, two points are overlapping.

2). The results clearly indicated that *Q. petraea* is more variable than *Q. robur* and that this difference increased when only discriminant fragments were considered, as shown also by the FCA (Fig. 2). The comparison of nucleotide diversity among the different sets of fragments (especially 376 vs. 36) indicated that the selection of discriminant fragments resulted in the selection of genomic regions that exhibit also higher *within*-species diversity. Finally, the screening of discriminant fragments resulted in the selection of genomic regions that showed 10 times higher nucleotide divergence between the two species than the rest of the genome (0.03 vs. 0.003).

Conversion of RAPD fragments to SCARs

The amplified products of 36 RAPDs showing significant frequency differences between the two species were cloned and sequenced at their extremities or in totality depending on their sizes. SCAR primers were designed when three cloned fragments had the same DNA sequences. The similarity of the cloned products with the selected RAPD fragments was verified by Southern blot hybridizations. The RAPD profiles and the hybridization pattern with the cloned fragments of the same individuals were identical indicating that the SCAR fragment was derived from the amplified product. However, 13 fragments among the 36 selected were discarded during the conversion of RAPDs to SCARs because they were a mixture of DNA fragments of the same size or because the reamplification of the RAPD fragments (see methods) repeatedly produced a multibanded pattern. Improvement of the techniques to overcome these difficulties was not attempted.

For each cloned RAPD product, a pair of primers (15–20 base-pairs) was synthesized (Table 2). The size and the sequence (including or not the sequence of the RAPD primer) were designed according to the OLIGO program (Rychlik *et al.*, 1990). Among the 23 SCARs obtained, different types of polymorphisms were observed (Table 2, Figs 3 and 4). In two cases, complicated PCR patterns were obtained and they were ignored for further analysis. In seven cases, the same polymorphism as with RAPD primers was obtained: the presence or absence of bands when the corresponding SCAR primers were used (dominant markers) was concordant with the presence or absence of bands on RAPD profiles. In three cases, two amplified fragments were obtained with the SCAR primers that may be two allelic products of a given locus. With 11 out of the 23 SCARs, the primer pairs amplified fragments from all the individuals, even those which did not present the RAPD fragment. In this case, digestion of the SCAR fragments with restriction enzymes or SSCP was used to detect polymorphism within the fragment (codominant markers) (Fig. 3) (Bodénès *et al.*, 1996).

Table 2 Nucleotide diversity and divergence calculated from RAPD data

Set of fragments† *f robur/f petraea*	h_r	h_p	h_{pr}	d_{pr}
412 fragments				
0.0/0.0	0.0230	0.0247	0.0289	0.0050
0.1/0.1	0.0221	0.0236	0.0279	0.0051
0.2/0.2	0.0213	0.0228	0.0272	0.0051
376 fragments				
0.0/0.0	0.0226	0.0236	0.0261	0.0030
0.1/0.1	0.0218	0.0226	0.0252	0.0031
0.2/0.2	0.0211	0.0218	0.0245	0.0031
36 fragments				
0.0/0.0	0.0257	0.0366	0.0640	0.0339
0.1/0.1	0.0243	0.0346	0.0622	0.0328
0.2/0.2	0.0231	0.0327	0.0605	0.0326

h_r, nucleotide diversity within *Quercus robur*; h_p, nucleotide diversity within *Q. petraea*; h_{pr}, nucleotide diversity between *Q. robur* and *Q. petraea*; d_{pr}, nucleotide diversity between *Q. robur* and *Q. petraea*.
†Value of the fixation index in *Q. robur* and *Q. petraea*.

Determination of genomic copy number and origin of SCAR amplification products

In RAPD profiles, it is not known if fragments of the same size shown by two different trees are similar. Here, this was tested by using SCAR fragments as probes and hybridizing them to RAPD patterns. The hybridizations were carried out on Southern blots of the two species (Table 3). Hybridizations between 19 SCARs used as probes and the 43 samples resulted in 817 positions that could be compared to the RAPD patterns. Only 3.1 per cent of the comparisons showed a discrepancy between RAPD and hybridization profiles. As a general rule, SCAR probes hybridized with the corresponding RAPD fragment and sometimes with additional fragments (generally with a second fragment and occasionally with several other fragments). The selected fragment always exhibited the strongest intensity. In two cases (F1–1000 and A1–500), the probes hybridized with a DNA fragment of an unexpected size present in each sample, indicating

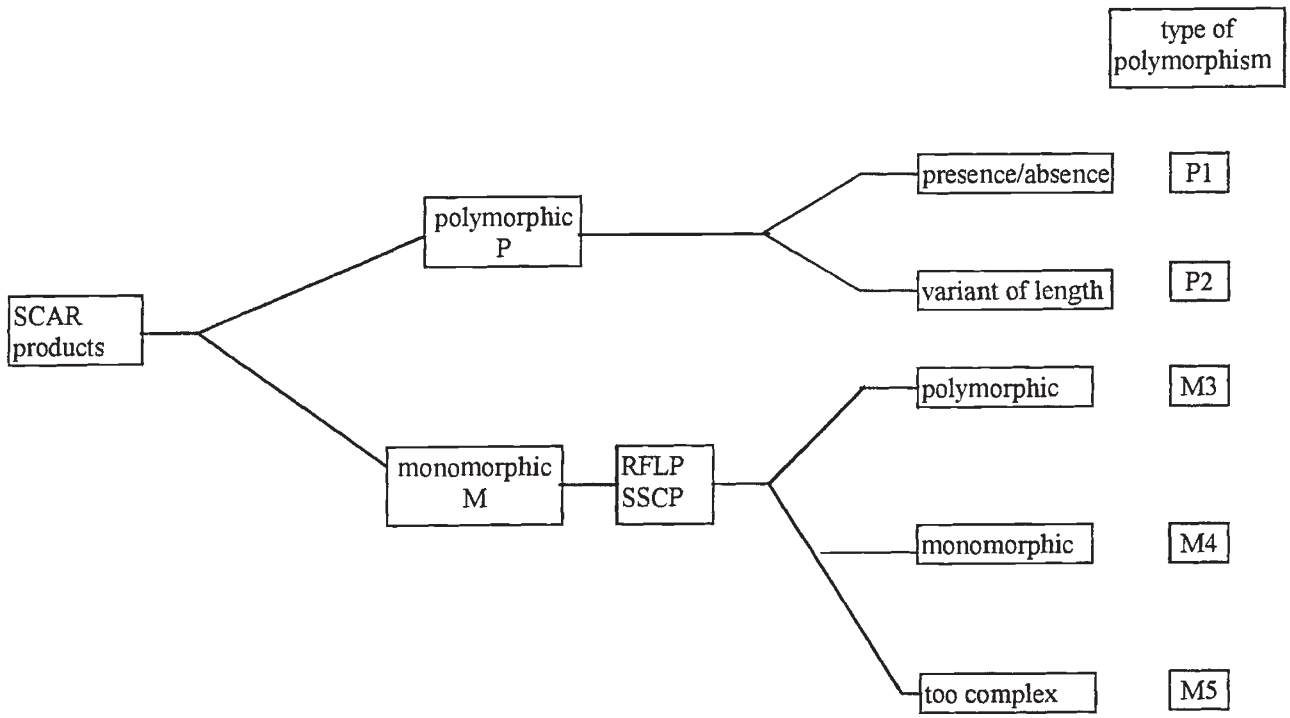

Fig. 3 Illustration of the types of polymorphisms detected in SCAR products.

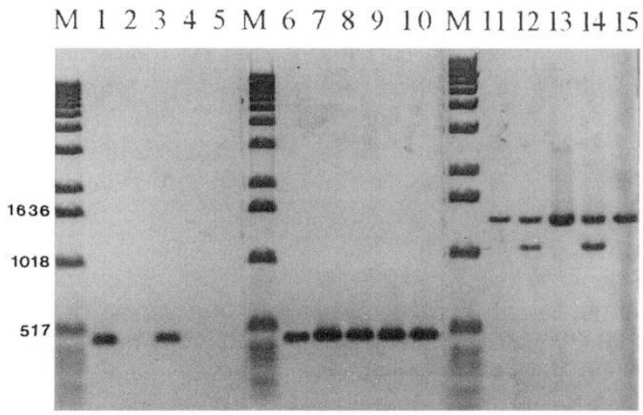

Fig. 4 Electrophoretic profiles of SCAR products. M, molecular weight ladder (1 kb). Lanes 1–5: presence or absence of R12-500 SCAR products (polymorphism P1). Lanes 6–10: presence of the P14-450 SCAR products in all individuals (monomorphism M). Lanes 11–15: one or two B11-1500 SCAR products (polymorphism P2).

that the sequenced fragments were not the selected fragments. This discrepancy was attributed to the difficulty of picking up the right band on the electrophoresis gel. Hybridizations were successful with RAPD patterns of the other species; therefore RAPD fragments of identical sizes in the two species were similar.

To evaluate the copy number of the informative SCARs, genomic DNA digested with endonucleases was hybridized with SCAR fragments. In most cases, the probes hybridized with a small number of fragments and they were classified as representing low copy numbers or moderately repeated regions (Table 3).

Sequence analysis

Among the 23 sequences, three presented significant nucleotide similarities with known sequences in the data banks (Genbank/EMBL data libraries) using BLASTN: B12-500, B19-800 and P14-1000. Similarities were also found for B19-800 and P14-1000 when BLASTX was used. The SCAR fragments which contained repeated motives in their sequences were analysed with and without the repeated motives and did not show significant similarities when repeated motives were excluded.

B12-500 exhibited a high similarity with ribosomal RNA genes of *Zea mays* or *Lupinus luteus*, B19-800 with trehalose-phosphate synthase of *Saccharomyces cerevisiae* and P14-1000 with a transposable element of *Antirrhinum majus* or *Glycine max*.

Table 3 Results of Southern blot hybridization experiments between SCAR fragments and RAPD or restricted genomic DNA of oaks

SCAR fragment	RAPD†	Genomic DNA number of repeats‡
A1 500	TD	LR
A17 700	S	MR–HR
B11 1500	I	N
B12 500	D	LR–MR
B12 750	I	LR
B12 800	I	N
B19 800	N	MR
E6 1600	I	LR
F1 1000	TD	LR
F14 700	I	LR–MR
174 400	S	LR
I13 300	S	LR
I14 250	I	LR
I14 780	I	LR
I16 500	D	LR
P14 450	I	LR
P14 1000	I	MR
P17 1400	S	LR
R12 500	S	N
U1 500	D	LR
U7 790	S	LR

†I: identical patterns between RAPD and hybridization scores. S: hybridization with several bands including the selected band. D: discrepancy between RAPD and hybridization profiles; at least one of the 43 samples exhibited a discrepancy. TD: total discrepancy; hybridized fragments of all samples were of the same size but different from the selected RAPD fragment. N: no results obtained.
‡HR: highly repeated sequences (>100). MR: moderately repeated sequences (10–100). LR: low copy number sequences (<10).

Discussion

Genomic organization of discriminant regions

The screening procedure of RAPD fragments and the subsequent conversion of informative RAPD fragments into SCARs resulted in the identification of 23 informative regions in the genome. In order to evaluate the degree of repetition of these regions (repeated or single copy sequences), Southern blots of restricted genomic DNA were hybridized with SCAR probes derived from cloned RAPD fragments. In 14 (74 per cent) cases, we observed weak signals of low copy sequences and in 5 (26 per cent) cases, moderately repeated sequences (Fig. 5, Table 3). The low copy number of informative SCAR fragments that we obtained appears therefore to be an exception, when compared to other results obtained with RAPD fragments. Except for *Arabidopsis*, where low copy numbers were found for RAPD fragments (Reiter *et al.*, 1992), in all other cases RAPD fragments appear to be highly repeated in the genome: for example, 89 per cent of RAPD probes from *Pinus pinaster* (Plomion *et al.*, 1995) hybridized with highly repeated sequences whereas 50 per cent or less of RAPD probes from *Eucalyptus* (Grattapaglia & Sederoff, 1994), cocoa (N'Goran *et al.*, 1994), lettuce (Paran & Michelmore, 1993) and *Petunia* (Peltier *et al.*, 1994) corresponded to low copy number sequences.

These results seem to indicate that the proportion of repeated sequences is correlated to the DNA content (Flavell *et al.*, 1974). *Quercus robur* and *Q. petraea* are known to have small genomes (1.8 pg/2C, Favre & Brown, 1996). However, the number of copies of RAPD fragments is still lower than those observed in other organisms having similar genome sizes to *Eucalyptus* (Grattapaglia & Sederoff, 1994). Other causes may therefore be advocated to explain the low copy number. One of them may be related to the technique that was used for the hybridization experiments. Here, we used cloned RAPD fragments as probes instead of fragments excised from gels. Consequently, we avoided the possible presence of several comigrating RAPD fragments in the probe. Finally, the results we obtained may just reveal that discriminant genomic regions between the two species are mostly represented in single copies, e.g. in coding regions of the genome. Three sequences among the 23 SCAR fragments present a significant similarity with identified sequences in data libraries. The P14–1000 sequence exhibited a high similarity with a transposable element. It has been shown that transposition could promote perturbations in different regions of the genome and could give rise to abnormalities in the mating system, for example (Rose & Doolittle, 1983; Krieber & Rose, 1986). The maintenance of the two species in the complex of two interfertile species, *Q. petraea* and *Q. robur*, has often been interpreted as the result of selection pressures towards different contrasting ecotypes (Kleinschmit *et al.*, 1995). It is possible that the discriminant regions identified during the RAPD screening procedure are involved in the adaptation to the ecological niches corresponding to the two species.

We also investigated the similarity between fragments showing the same electrophoretic mobility in a gel. Twenty-one SCAR fragments were used as probes to hybridize Southern blots of RAPD

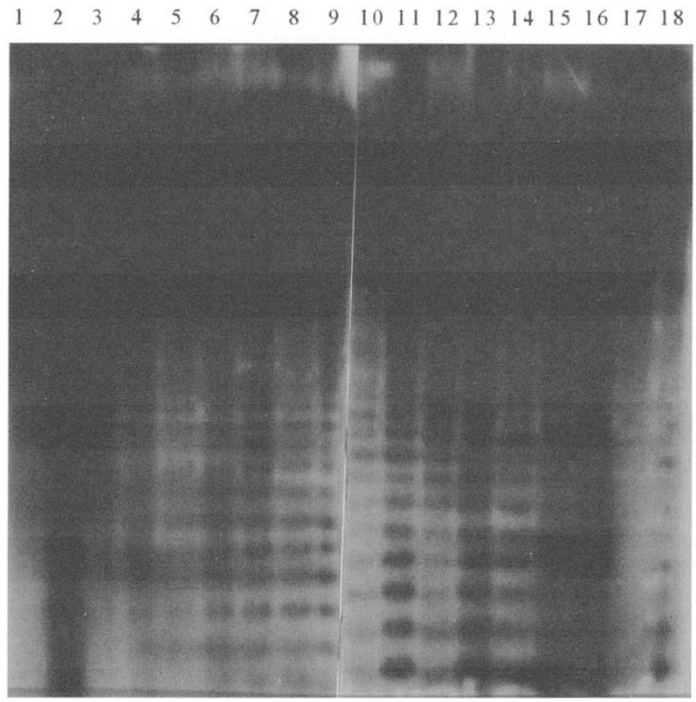

Fig. 5 Hybridization patterns of SCAR P14-1000 with restricted genomic DNA. Lanes 1–9: pedunculate oaks. Lanes 10–18: sessile oaks.

patterns from which they originated (the two SCARs for which the PCR gave complex patterns were discarded). In 1 per cent of the cases, the RAPD fragment, not visible after ethidium bromide staining, was revealed after hybridization. In 2.1 per cent of the cases, the RAPD fragment of the expected size did not hybridize with the appropriate SCAR fragment. The former situation corresponds to a greater sensitivity of the hybridization as compared to ethidium bromide staining; the latter indicates that the same primers can amplify different sequences of the same size (Fig. 6). Similar results have already been observed for *Allium* (Wilkie *et al.*, 1993), *Brassica* (Thormann *et al.*, 1994), *Gliricidia* (Chalmers *et al.*, 1992), *Petunia* (Peltier *et al.*, 1994) and *Vitis* (Xu *et al.*, 1995).

Differences of polymorphisms between RAPD and the corresponding SCAR

RAPD assay allows the detection of single base changes in genomic DNA (Klein-Lankhorst *et al.*, 1991; Williams *et al.*, 1993). Because we did not know the cause of the polymorphism observed in the RAPD profiles (mismatches in the priming sites or molecular rearrangments within the RAPD fragment), we could not predict whether simple extension of the 10 bp RAPD primer to longer specific primers would retain this polymorphism. Thus we did not add systematically 10 bp to the 10 bp RAPD primer and on several occasions we chose a new primer pair within the sequence. Conversion of RAPDs to SCARs may have resulted in the modification of the original polymorphism.

For 11 SCARs out of 21 (52.4 per cent), PCR amplification resulted in the loss of the original RAPD polymorphism indicating that the polymorphism resulted from mismatches at the original priming sites. For seven SCARs (33.3 per cent), amplification occurred only for individuals which exhibited the corresponding RAPD fragments. In these cases, the polymorphism between individuals originated from sequence divergence or rearrangements that either altered the orientation of the primers or resulted in the separation of the primers by too great a distance to allow amplification (Paran & Michelmore, 1993). For three SCARs (14.3 per cent), variant length polymorphism was directly observed after PCR with, in two cases, the appearance of a second amplified fragment. For A17-700, the presence of a subrepetition of 250 bp within the sequence of the RAPD fragment explains the PCR amplification of two fragments of 700 and 250 bp, respectively. For B11-1500 and I13-300, sequence analysis did not afford any explanation for this phenomenon but hybridization experiments indicated that the two fragments produced by B11-1500 harboured similar sequences. In these cases, PCR

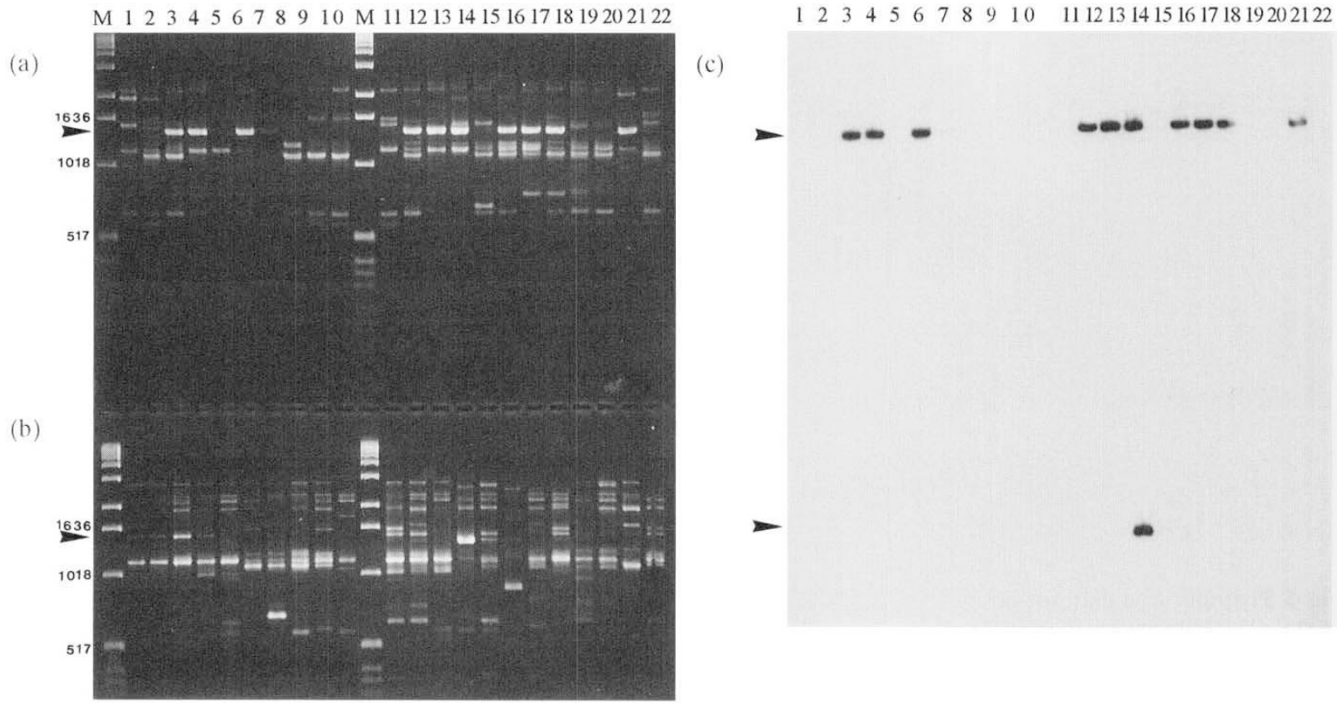

Fig. 6 Comparison of RAPD profiles with hybridization patterns. SCAR B11-1500 is used as probe and hybridized with RAPD products of B11. (a) RAPD profiles. Lanes 1 and 11: molecular weight ladder (1 kb). Lanes 2–22: pedunculate oak samples. (b) RAPD profiles. Lanes 1 and 11: molecular weight ladder (1 kb). Lanes 2–22: sessile oak samples. (c) Hybridization patterns with SCAR B11-1500 used as probe. Samples are arranged as in (a) and (b).

conditions could not be optimized to obtain a single amplified fragment. Converting informative RAPD fragments into SCARs did not lead to the same conclusions as the RAPD fragments themselves. Whereas RAPD polymorphism was mostly caused by mismatches on the primer sites or molecular rearrangements, SCAR polymorphism originated from variations within the amplified fragment. This change does not affect the identification of loci involved in differentiation but may modify the assignment of alleles as different between the two species.

Nucleotide diversity and differentiation derived from RAPD data

Nucleotide diversity estimates from RAPD fragments are dependent on strict assumptions indicated by Clark & Lanigan (1993). The sequencing of the informative RAPD fragments and the hybridization experiments performed here allowed us to verify some of them. The stronger assumption is that RAPD polymorphism is caused mainly by point mutations in the primer sequence and is allele specific. Here, seven of 21 polymorphic RAPD markers were attributed to insertion/deletion, whereas the remaining 14 were caused by mutations in the primer sites. Hybridization experiments also showed that informative RAPD fragments were allele specific in both species as also shown by inheritance studies on a subset of fragments in controlled crosses (Moreau *et al.*, 1994; Bodénès *et al.*, unpublished results).

The estimates of nucleotide diversity and differentiation obtained here may therefore be considered as 'rough' figures depicting the heterogeneity of genomic nucleotide divergence between the two species. Although the overall divergence among the two species is extremely low (0.5 per cent), there are a few 'hot spots' where the divergence increases to 3 per cent. These results are in the range of nucleotide differentiation percentages obtained with the same method among populations of a given species (0.02–0.08 per cent in trembling aspen, Chong *et al.*, 1994) or among species within a given genus (9–13 per cent in *Shorea* spp., tropical forest species from Malaysia, Harada *et al.*, 1994). The results obtained on nucleotide diversity and divergence clearly confirmed earlier analysis conducted with allozymes (Zanetto *et al.*, 1994) or RAPD data (Moreau *et al.*,

1994) which indicated that (1) the two species were poorly differentiated and (2) *Q. petraea* was more variable than *Q. robur*. As these authors advocated, asymmetric hybridization may be responsible for the discrepancy in the level of diversity in the two species.

Acknowledgements

This work was supported by an EU grant, BIO2-CT93-0373, 'Development and adaptation of rapid molecular screening techniques for assessing genetic diversity in forest trees'. We thank Gregor Aas, Gerard Burzinsky, Luis Gil, Jochen Kleinschmit, Jörg Kleinschmit, Alexander Tucovic, Guiseppe Vendramin and the French Forest Service (O.N.F.) for collecting the material.

References

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. AND LIPMAN, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

BACILIERI, R., DUCOUSSO, A., PETIT, R. J. AND KREMER, A. 1996. Mating system and asymmetric hybridization flow in a mixed stand of European oaks. *Evolution*, **50**, 900–908.

BODÉNÈS, C., LAIGRET, F. AND KREMER, A. 1996. Inheritance and molecular variations of PCR-SSCP fragments in Pedunculate oak (*Quercus robur* L.). *Theor. Appl. Genet.* (in press).

CHALMERS, K. J., WAUGH, R., SPRENT, J. I., SIMONS, A. J. AND POWELL, W. 1992. Detection of genetic variation between and within populations of *Gliricidia sepmium* and *G. maculata* using RAPD markers. *Heredity*, **69**, 465–472.

CHONG, D. K. X., YANG, R. Y. AND YEH, F. C. 1994. Nucleotide divergence between populations of trembling aspen (*Populus tremuloides*) estimated with RAPDs. *Curr. Genet.*, **26**, 374–376.

CLARK, A. G. AND LANIGAN, C. M. S. 1993. Prospects for estimating nucleotide divergence with RAPDs. *Mol. Biol. Evol.*, **10**, 1096–1111.

CRAWFORD, D. J. 1985. Electrophoretic data and plant speciation. *Syst. Bot.*, **10**, 405–416.

DENHART, D. T. 1966. A membrane filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Commun.*, **23**, 641–646.

DUPOUEY, J. L. AND BADEAU, V. 1993. Morphological variability of oaks (*Quercus robur* L., *Quercus petraea* (Matt.) Liebl., *Quercus pubescens* Willd.) in northeastern France: preliminary results. *Ann. Sci. For.*, **50**, 35s–40 s.

FAVRE, J. M. AND BROWN, S. 1996. A flow cytometric evaluation of the nuclear DNA content and GC per cent in genomes of European oak species. *Ann. Sci. For.*, **53**, 915–917.

FLAVELL, R. B., BENETT, M. D., SMITH, J. B. AND SMITH, D. B. 1974. Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem. Genet.*, **12**, 257–269.

GRATTAPAGLIA, D. AND SEDEROFF, R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics*, **137**, 1121–1137.

HARADA, K., KINOSHITA, A., SHUKOR, N. A. A., TACHIDA, H. AND YAMAYKI, T. 1995. Genetic variation estimated in three Shorea species by the RAPD analysis. *Jap. J. Genet.*, **69**, 713–718.

KENDALL, M. G. AND STUART, A. 1977. *The Advanced Theory of Statistics*, Vol. I, 4th edn. Griffin, London.

KLEIN-LANKHORST, R. M., VERMUNT, A., WEIDE, R., LIHARSKA, T. AND ZABEL, P. 1991. Isolation of molecular markers for tomato (*L. esculentum*) using random amplified polymorphic DNA (RAPD). *Theor. Appl. Genet.*, **83**, 108–114.

KLEINSCHMIT, J. R. G., BACILIERI, R., KREMER, A. AND ROLOFF, A. 1995. Comparison of morphological traits of pedunculate oak (*Q. robur* L.) and sessile oak (*Q. petraea* (Matt.) Liebl.). *Silvae Genet.*, **44**, 256–269.

KREMER, A. AND PETIT, R. J. 1993. Gene diversity in natural populations of oak species. *Ann. Sci. For.*, **50**, 186s–202s.

KRIEBER, M. AND ROSE, M. R. 1986. Molecular aspects of the species barrier. *Ann. Rev. Ecol. Syst.*, **17**, 465–485.

LEBART, L., MORINEAU, A. AND TABART, N. 1984. *Multivariate Descriptive Stastitical Methods: Correspondance Analysis and Related Techniques for Large Matrices.* J. Wiley and Sons, New York.

LISITSYN, N., LISITSYN, N. AND WIGLER, M. 1993. Cloning the differences between two complex genomes. *Science*, **259**, 946–951.

MOREAU, F., KLEINSCHMIT, J. AND KREMER, A. 1994. Molecular differentiation between *Q. petraea* and *Q. robur* assessed by random amplified DNA fragments. *Forest Genet.*, **1**, 51–64.

MÜLLER-STARCK, G., HERZOG, S. AND HATTEMER, H. H. 1993. Intra and inter population genetic variation in juvenile populations of *Quercus robur* L. and *Quercus petraea* Liebl. *Ann. Sci. For.*, **50**, 233s–244s.

NEI, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

NEI, M. AND LI, W. H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 5269–5273.

NEI, M. AND MILLER, J. C. 1990. A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics*, **125**, 873–879.

N'GORAN, J. A., LAURENT, V., RISTERUCCI, A. M. AND LANAUD, C. 1994. Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. *Heredity*, **73**, 589–597.

PARAN, I. AND MICHELMORE, R. W. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theor. Appl. Genet.*, **85**, 985–993.

PELTIER, D., FARCY, E., DULIEU, H. AND BERVILLE, A. 1994. Origin, distribution and mapping of RAPD markers from wild *Petunia* species in *Petunia hybrida* Hort lines. *Theor. Appl. Genet.*, **88**, 637–645.

PETIT, R. J., WAGNER, D. B. AND KREMER, A. 1993 Ribosomal DNA gene and chloroplast DNA polymorphisms in a mixed stand of *Quercus robur* and *Quercus petraea*. *Ann. Sci. For.*, **50**, 41s–47s.

PLOMION, C., BARHMAN, N., DUREL, C. E. AND O'MALLEY, D. M. 1995. Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers. *Heredity*, **74**, 661–668.

REED, K. C. AND MANN, D. A. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. *Nucl. Acids Res.*, **13**, 7207–7221.

REITER, R. S., WILLIAMS, J. G. K., FELDMANN, K. A., RAFALSKI, J. A., TINGEY, S. V. AND SCOLNIK, P. A. 1992. Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 1477–1481.

ROSE, M. R. AND DOOLITTLE, W. F. 1983. Molecular biological mechanisms of speciation. *Science*, **200**, 157–162.

RYCHLIK, W., SPENCER, W. J. AND RHOADS, R. E. 1990. Optimization of the annealing temperature for DNA amplification *in vitro*. *Nucl. Acids Res.*, **18**, 6409–6412.

SAGHAI-MAROOF, M. A., SOLIMAN, K. M., JORGENSEN, R. A. AND ALLARD, R. W. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 8014–8018.

SHELDON, F. H. 1995. Advances in the theory and practice of DNA hybridization as a systematic method. In: Schierwater, B., Streit, B., Wagner, G. P. and DeSalle, R. (eds) *Molecular Ecology and Evolution: Approaches and Applications*, pp. 285–297. Birkhäuser Verlag, Basel,

SOUTHERN, E. M. 1976. Detection of specific sequences among DNA fragments separated by gel electrophoresis *J. Mol. Biol.*, **98**, 503–517.

STEINHOFF, S. 1993. Results of species hybridization with *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Ann. Sci. For.*, **50**, 137s–143s.

THORMANN, C. E., FERREIRA, M. E., CAMARGO, L. E. A., TIVANG, J. G. AND OSBORN, T. C. 1994. Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. *Theor. Appl. Genet.*, **88**, 973–980.

WILBUR, W. J. AND LIPMAN, D. J. 1983. Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 726–730.

WILKIE, S. E., ISAAC, P. G. AND SLATER, R. 1993. Random amplified polymorphic DNA (RAPD) markers for genetic analysis in *Allium*. *Theor. Appl. Genet.*, **86**, 497–504.

WILLIAMS, J. G. K., KUBELIK, A. R., LIVAK, K. J., RAFALSKI, J. A. AND TINGEY, S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.*, **18**, 6531–6535.

WILLIAMS, J. G. K., HANAFEY, M. K., RAFALSKI, J. A. AND TINGEY, S. V. 1993. Genetic analysis using random amplified polymorphic DNA markers. *Methods Enzymol.*, **318**, 705–740.

XU, H., WILSON, D. J., ARULSEKAR, S. AND BAKALINSKY, A. T. 1995. Sequence specific polymerase chain-reaction markers derived from randomly amplified polymorphic DNA markers for fingerprinting grape (*Vitis*) rootstocks. *J. Amer. Soc. Hort. Sci.*, **120**, 714–720.

ZANETTO, A., ROUSSEL, G. AND KREMER, A. 1994. Geographic variation of inter-specific differentiation between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Forest Genet.*, **1**, 111–123.