ELSEVIER

Forest Ecology and Management 156 (2002) 5–26

Forest Ecology and Management

www.elsevier.com/locate/foreco

Chloroplast DNA variation in European white oaks Phylogeography and patterns of diversity based on data from over 2600 populations

Rémy J. Petit[a,*], Ulrike M. Csaikl[b], Sándor Bordács[b,c], Kornel Burg[b], Els Coart[d], Joan Cottrell[e], Barbara van Dam[f], John D. Deans[g], Sylvie Dumolin-Lapègue[a,1], Silvia Fineschi[h], Reiner Finkeldey[i], Amanda Gillies[f,2], Izabela Glaz[a,j], Pablo G. Goicoechea[k], Jan S. Jensen[l], Armin O. König[m], Andrew J. Lowe[g], Søren F. Madsen[n], Gabor Mátyás[j], Robert C. Munro[g], Maria Olalde[k], Marie-Hélène Pemonge[a], Flaviu Popescu[a,o], Danko Slade[a,p], Helen Tabbener[e], Daniela Taurchini[i], Sven G.M. de Vries[f], Birgit Ziegenhagen[m], Antoine Kremer[a]

[a]*Institut National de la Recherche Agronomique (INRA), Station de Recherches Forestières, BP 45, F-33611 Gazinet Cedex, France*
[b]*Austrian Research Centre Seibersdorf (ARCS), A-2444 Seibersdorf, Austria*
[c]*Department of Forestry, National Institute for Agricultural Quality Control (NIAQC), Budapest Keleti K. u. 24, H-1024 Budapest, Hungary*
[d]*CLO, Department of Plant Genetics and Breeding, Caritasstraat 21, B-9090 Melle, Belgium*
[e]*Forestry Commission (FC), Forest Research, Northern Research Station, Roslin, Midlothian EH25 9SY, Scotland, UK*
[f]*ALTERRA Green World Research, P.O. Box 47, NL-6700 AA Wageningen, The Netherlands*
[g]*Institute of Terrestrial Ecology (ITE), Bush Estate, Penicuik, Midlothian EH26 0QB, Scotland, UK*
[h]*CNR Istituto per l'Agroselvicoltura, via Marconi 2, I-05010 Porano TR, Italy*
[i]*WSL, Zürcherstrasse 111, CH Birmensdorf, Switzerland*
[j]*Department of Genetics and Physiology of Woody Plants, Forest Research Institute, St. 3 Bitwy Warszawskiej 1920r, PL 00973 Warsaw, Poland*
[k]*NEIKER A.B., Granja Modelo-Arkaute, 01080 Vitoria-Gasteiz, Spain*
[l]*Danish Forest and Landscape Research Institute (DFLRI), Hørsholm Kongevej, DK-2970 Hørsholm, Denmark*
[m]*Bundesforschungsanstalt für Forst- und Holzwirtschaft (BFH), Institut für Forstgenetik und Forstpflanzenzüchtung, Sieker, Landstrasse 2, D-22927 Grosshansdorf, Germany*
[n]*The Arboretum, The Royal Agricultural and Veterinary University, Kirkegaardsvej 3a, DK-2970 Hørsholm, Denmark*
[o]*ICAS, Arboretum of Simeria, Str. Biscaria 1, 2625 Simeria, Romania*
[p]*Forest Research Institute, 10450 Jastrebarsko, Croatia*

Abstract

A consortium of 16 laboratories have studied chloroplast DNA (cpDNA) variation in European white oaks. A common strategy for molecular screening, based on restriction analysis of four PCR-amplified cpDNA fragments, was used to allow

[*] Corresponding author. Tel.: +33-5-57-97-90-87; fax: +33-5-57-97-90-88.
E-mail address: remy@pierroton.inra.fr (R.J. Petit).
[1] Present address: IFREMER, Laboratoire de Génétique et Pathologie, BP 133, F-17390 La Tremblade, France.
[2] Present address: Sir Harold Mitchell Building, School of Environmental and Evolutionary Biology, University of St Andrews, St Andrews, Fife KY16 9TH, Scotland, UK.

0378-1127/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0378-1127(01)00645-4

comparison among the different laboratories. A total of 2613 oak populations (12,214 individual trees from eight species) weresampled from 37 countries, and analysed with the four fragments. They belong to eight related oak species: *Quercus robur*, *Q. petraea*, *Q. pubescens*, *Q. frainetto*, *Q. faginea*, *Q. pyrenaica*, *Q. canariensis* and *Q. macranthera*. During this survey, 45 chloroplast variants were detected and are described together with their phylogenetic relationships, but several of these haplotypes were pooled when there were some risks of confusion across laboratories during the survey, and finally 32 remained that were mapped and used in diversity analyses. A strong phylogeographic structure is apparent from the data, where related haplotypes have broadly similar geographic distributions. In total, six cpDNA lineages are identified, which have distinct geographic distributions, mainly along a longitudinal gradient. Most haplotypes found in northern Europe are also present in the south, whereas the converse is not true, suggesting that the majority of mutations observed were generated prior to postglacial recolonisation, corroborating the conclusions of earlier studies. The description of a new western European lineage constitutes a major finding, compared to earlier phylogenetic treatments. Although the eight oak species studied systematically share cpDNA variants when in sympatry, they partition cpDNA diversity differently, as a consequence of their different ecology and life history attributes. Regional differences in levels of differentiation also exist (either species-specific or general); these seem to be related to the intensity of past and present management of the forests across Europe but also to the level of fragmentation of the range within these regions. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Admixture; Genetic diversity; Genetic differentiation; Introgression; Phylogeny; Phylogeography; Refugia

1. Introduction

Oaks were one of the first plant taxa for which cytoplasmic DNA markers were used to examine geographic variation within and among populations (Kremer et al., 1991; Whittemore and Schaal, 1991; Ferris et al., 1993; Petit et al., 1993a). More recent studies have extended these early results, yielding a detailed picture of chloroplast DNA (cpDNA) variation both range-wide across Europe (Dumolin-Lapègue et al., 1997) and at finer scales within particular regions (Ferris et al., 1995, 1998; Johnk and Siegismund, 1997; Petit et al., 1997; Dumolin-Lapègue et al., 1998, 1999a). The chloroplast genome is known to be maternally inherited in oaks (i.e. transmitted by seeds only; Dumolin et al., 1995) and, as a consequence, displays higher levels of among population variation compared with nuclear markers (Petit et al., 1993b). The absence of paternal leakage has been indirectly inferred by demonstrating the interdependent assorting of mitochondrial and chloroplast genomes (Dumolin-Lapègue et al., 1998). Thus a phylogeny based on the chloroplast (or mitochondrial) genome for oak species is not complicated by recombination, although homoplasy due to convergent or reverse mutations may still arise. Earlier studies have demonstrated that cpDNA variation is geographically structured in oaks and that related haplotypes often have similar distribution (Dumolin-Lapègue et al., 1997). Because of the intrinsic historical information content of these phylogenies, the spatio-temporal dynamics of the maternal lineage can be uncovered by first studying the phylogenetic relationships of cpDNA variants and then mapping their geographic distribution.

Most previous oak phylogeographic studies included several species and demonstrated extensive sharing of cytoplasmic variants amongst species. However, this does not imply that oak species will partition cpDNA diversity in an identical way (Dumolin-Lapègue et al., 1999a). Moreover, the partitioning of cpDNA diversity may also differ according to region; as a consequence, the influence of geography and taxonomy on the partitioning of cpDNA variation should be investigated jointly.

Over the last 4 years, a team of 16 laboratories has cooperated to extend previous studies in order to conduct a detailed European-wide survey of cpDNA variation in oaks (Kremer, 2001). In this paper, the methods used are presented along with a description and phylogenetic analyses of the haplotypes discovered during this survey. This provides a basis for the nine regional papers included in this issue, and for the companion paper dealing with postglacial recolonisation routes (Petit et al., 2002a). Moreover, an analysis of species and regional effects on cpDNA diversity in Europe is presented. Throughout, comparisons are made between measures of diversity that take account of phylogenetic information and those that do not. This allows comparisons to be made between patterns of diversity caused by the mixing of haplotypes

belonging to the same cpDNA lineages and those due to the mixture of two or more lineages during recolonisation.

2. Material and methods

2.1. Plant material

A total of 2613 oak populations (12,214 individual trees) were sampled throughout Europe by the laboratories involved in that study. Some of this material has been analysed previously, notably 345 European populations studied by Dumolin-Lapègue et al. (1997) and 378 populations originating from southern France (Dumolin-Lapègue et al., 1998). The initial goal was to concentrate the study on the two most widespread and economically valuable species *Quercus robur* and *Q. petraea*. However, to achieve a better coverage of Europe, especially in regions of inferred refugia, it proved necessary to enlarge this study to other related white oak species. Some bias in favour of the economically important species may have persisted, and this should be borne in mind when considering the distribution maps of population samples per species. Unfortunately, it has not yet been possible to obtain samples from the following countries: Albania, Bulgaria, Greece, and Macedonia. Other countries insufficiently covered for meaningful analyses of cpDNA diversity include Ireland (where a study is in progress; G.C. Douglas, pers. comm.), as well as White Russia, Bosnia and Herzegovina, Russia, Turkey, Ukraine, and Yugoslavia. Furthermore, species of the white oak complex are also present in other neighbouring countries such as Algeria, Morocco, Tunisia and the Caucasian region, where there has been little or no sampling so far.

The following species are included in this survey: *Q. robur* L. (present in 1469 populations), *Q. petraea* (Matt.) Liebl. (1054 populations), *Q. pubescens* Willd. (508 populations), *Q. pyrenaica* Willd. (97 populations), *Q. faginea* Lam. (78 populations), *Q. frainetto* Ten. (27 populations), *Q. canariensis* Willd. (11 populations), and *Q. macranthera* Fisch. and May (two populations) (Table 1). Taxonomic identification was based on morphological characters using the nomenclature of Bussoti and Grossoni (1997). Only those taxa that are well described and easily distinguishable were included, so that the above species should be considered sensu lato. In particular, trees labelled as *Q. pedunculiflora* (Rumania) were treated as *Q. robur*, *Q. dalechampii* and *Q. polycarpa* were considered with *Q. petraea* and those named *Q. virgiliana* were combined with *Q. pubescens* (see Bordács et al., 2002 for more information on these taxa). The oak material (acorns, twigs or leaves) was collected in forests or in provenance tests. When collections were made in the forest, sample points were distributed according to a systematic grid system (every 50 km). However, during the course of the project, additional sample points were selected. Whenever possible, material of inferred autochthonous origin was sampled, but in many cases it was not possible to be conclusive, so the origin was considered unknown or even dubious when there were indications of plantation. As a rule, within each forest, five trees were sampled at a distance of at least 50 m, and at most 500 m, to one another. A summary of the number of individuals sampled according to country of origin and laboratory where analysis were performed is provided in Table 2. The following information was recorded for each population: population code number, name of the forest, longitude, latitude and altitude, country, nature of the stand (e.g. hedges, forest pasture, forest, provenance test), origin (autochthonous, dubious, unknown, introduced), provenance name (for populations sampled in provenance tests) and name of the laboratory who did the collection.

2.2. PCR-RFLP procedure

A technical workshop was organised at INRA Bordeaux in 1997 to share the molecular techniques and ensure from the outset that comparable results would be obtained across the laboratories involved in the project. Methods follow those of Dumolin-Lapègue et al. (1997), with some, usually minor, modifications across the 12 laboratories carrying out the molecular analyses, as reported in the regional papers included in this issue (Bordács et al., 2002; Cottrell et al., 2002; Csaikl et al., 2001, 2002; Fineschi et al., 2002; Olalde et al., 2002; Jensen et al., 2002; König et al., 2002; Petit et al., 2002b). For DNA isolation, the methods either followed that of Dumolin et al. (1995) based on CTAB, or used the QIAGEN plant DNA kit. In the majority of laboratories, four largely non-coding cpDNA

Table 1
Number of populations sampled per species

Haplotype	Lineage	*Q. robur*	*Q. pubescens*	*Q. petraea*	*Q. pyrenaica*	*Q. faginea*	*Q. frainetto*	*Q. macranthera*	*Q. canariensis*	Total
1	C	272	104	169		8	1			522
2	C	31	42	24						92
3	C		1							1
4	A	29	4	27						55
5	A	109	70	75			13			252
6	A	21	4	15			4			41
7	A	431	169	304		7	2			743
8	F	1								1
9	F	1						2		3
10	B	455	58	260	53	19			1	683
11	B	181	34	104	15	6			2	295
12	B	312	35	214	25	14				506
13	D	3		5			1			9
14	D	3								3
15	D	25		14			1			36
16	D	8		2			2			12
17	D	14	41	13			9			67
18	D		1							1
19	D	1					1			2
20	D	2		10						10
21	E	2	3							4
22	E								1	1
24	B	6		2	3	1				12
25	B	2			2				4	7
26	A	1								1
27	E				5	24			3	31
28	E		1			2			1	2
29	E					1				1
30	A			1				1		2
31	A	3		3						6
32	B			1						1
33	E					1				1
Total	6	1469	508	1054	97	78	27	2	11	2613

fragments were studied, each with one restriction enzyme: *trn*D/*trn*T (DT) with *Taq*I, *psa*A/*trn*S (AS) with *Hin*fI, *psb*C/*trn*D (CD) with *Taq*I, and *trn*T/*trn*F (TF) with either *Hin*fI or *Alu*I. The three first primer pairs (AS, CD, and DT) are described in Demesure et al. (1995), the last one (TF) in Taberlet et al. (1991). For the construction of the phylogenetic trees, the haplotypes identified during the survey were further characterised for two additional point mutations involving two more primers–enzyme combinations: DT–*Alu*I and TF–*Cfo*I. This latter polymorphism had been used in other cpDNA surveys in Europe (Ferris et al., 1993, 1998). The restriction fragments were separated by electrophoresis on 8% polyacrylamide gels as described in Dumolin et al. (1995), or in higher resolution systems as described in Csaikl et al. (2001).

Some participants did not analyse all fragments systematically, but only those that had been shown to discriminate among the haplotypes discovered in a preliminary screening within a given region. In addition, two alternative restriction enzymes were used in the analysis of fragment TF: initially, *Alu*I was used in all laboratories, but then it was found that *Hin*fI allowed the identification of several new cpDNA types in addition to providing good resolution of existing variation; however, for some laboratories, there was no time to include this new combination and they used the

Table 2
Number of populations sampled per country and per laboratory

Country	Laboratory[a]												
	ARCS	CLO	CNR	DFLRI	FC	ALTERRA	IFG	INRA	ITE	NEIKER	NIAQC	WSL	Total
Algeria	0	0	0	0	0	0	0	1	0	0	0	0	1
Armenia	0	0	0	0	0	0	0	2	0	0	0	0	2
Austria	115	0	0	0	0	0	2	3	1	0	0	0	121
Belgium	0	9	0	0	0	0	2	11	0	0	0	0	22
White Russia	0	0	0	0	0	0	0	4	0	0	0	0	4
Bosnia and Herzegovina	0	0	0	0	0	0	3	0	0	0	0	0	3
Croatia	0	0	0	0	0	0	0	38	0	0	0	0	38
Czech Republic	0	0	0	0	0	0	3	5	0	0	0	0	8
Denmark	0	0	0	19	0	0	2	2	27	0	0	0	50
England	0	0	0	5	162	0	3	5	0	0	0	0	175
Estonia	0	0	0	7	0	0	0	0	0	0	0	0	7
Finland	0	0	0	10	0	0	0	0	0	0	0	0	10
France	0	0	0	0	0	0	4	896	0	0	0	0	900
Georgia	0	0	0	0	0	0	0	2	0	0	0	0	2
Germany	0	0	0	1	0	0	316	20	0	0	0	0	337
Hungary	0	0	0	0	0	0	0	3	0	0	53	0	56
Ireland	0	0	0	0	0	0	0	4	0	0	0	0	4
Italy	0	0	130	0	0	0	0	42	0	0	0	24	172
Latvia	0	0	0	8	0	0	0	2	0	0	0	0	10
Lithuania	0	0	0	6	0	0	0	0	0	0	0	0	6
Luxembourg	0	0	0	0	0	0	2	0	0	0	0	0	2
Morocco	0	0	0	0	0	0	0	3	0	0	0	0	3
Netherlands	0	0	0	2	0	13	0	2	0	0	0	0	17
Norway	0	0	0	7	0	0	0	2	0	0	0	0	9
Poland	0	0	0	7	0	0	2	48	0	0	0	0	57
Portugal	0	0	0	0	0	0	0	2	0	19	0	0	21
Rumania	0	0	0	0	0	0	0	74	0	0	0	0	74
Russia	0	0	0	0	0	0	0	18	0	0	0	0	18
Scotland	0	0	0	1	60	0	0	4	0	0	0	0	65
Yugoslavia	0	0	0	0	0	0	0	5	0	0	0	0	5
Slovenia	0	0	0	0	0	0	0	2	0	0	0	0	2
Slovakia	0	0	0	0	0	0	13	3	0	0	0	0	16
Spain	0	0	0	0	0	0	0	9	0	165	0	0	174
Sweden	0	0	0	19	0	0	0	3	1	0	0	0	23
Switzerland	0	0	15	0	0	0	0	3	0	0	0	148	190
Turkey	0	0	0	0	0	0	0	1	0	0	0	0	1
Ukraine	0	0	0	0	0	0	0	8	0	0	0	0	8
Yugoslavia	0	0	0	0	0	0	0	5	0	0	0	0	5
Total	115	9	145	92	222	13	352	1227	29	184	53	172	2613
References[b]	1, 2, 3	3	2, 4	3, 5, 6, unpublished	7	3	1, 3, 5, unpublished	1, 3, 4, 5, 8, 9, unpublished	6, unpublished	10	1	2, 11	

[a] For signification of abbreviations and full address of institutions, see above (authors list).
[b] References for the original data: 1—Bordács et al., 2002; 2—Csaikl et al., 2001; 3—König et al., 2002; 4—Fineschi et al., 2002; 5—Csaikl et al., 2002; 6—Jensen et al., 2002; 7—Cottrell et al., 2002; 8—Dumolin-Lapègue et al., 1997; 9—Petit et al., 2002b; 10—Olalde et al., 2002; 11—Mátyás, 1999.

*Alu*I enzyme throughout. As a consequence, some haplotypes may not have been systematically distinguished by some participants; in this case, haplotypes were described using subscripts (e.g., haplotypes 5a, 5b, and 5c) solely to indicate this possibility. Although all haplotypes described here were included in the phylogenetic analyses, all other analyses are restricted to those haplotypes scored systematically.

2.3. Phylogenetic analysis

The data were scored as multistate unordered characters: each polymorphic restriction fragment was a character and the states were different sizes of this fragment. Nomenclature was as in Dumolin-Lapègue et al. (1997): the length variants were noted from 1 to 6, 9 being reserved for restriction site

mutations (Appendix B). The Fitch or the Kitsch algorithms of the PHYLIP 3.5 package (Felsenstein, 1993) was used to produce the phylogenetic trees. Fitch estimates phylogenies from distance matrix data under the 'additive tree model' according to which the distances are expected to equal the sums of branch lengths between the species. This method does not assume an evolutionary clock. Kitsch assumes that all tip haplotypes are contemporaneous and that there is an evolutionary clock. This means that branches of the tree cannot be of arbitrary length, but are constrained so that the total length from the root of the tree to any species is equal.

2.4. Genetic diversity analysis

The frequencies of the haplotypes as well as the distances between them (number of different restriction fragments) were used to compute diversity and differentiation measures, following Pons and Petit (1996), using the software HAPERMUT available at http://www.pierroton.inra.fr/genetics/labo/Software/. All measures of diversity as well as their standard errors were computed both by taking the distance between haplotypes into account (in the case of v) and by ignoring genetic distance (in the case of h). The diversity parameter v is defined as follows:

$$v = \sum_{ij} \pi_{ij} x_i x_j$$

where π_{ij} is the distance between haplotypes i and j, and x_i and x_j are their observed frequencies. When all distances π_{ij} are equal to 1, v becomes identical to the classical measure of diversity h, which is the probability that two haplotypes selected at random in the population will differ. The resulting coefficients of differentiation are called N_{ST} (when based on v) and G_{ST} (when using h), and can be directly compared. On the other hand, the measure of diversity v is scale-dependant (as it depends on the choice of the metrics used to compute the distance, for instance, whether absolute or relative number of different restriction sites were used). To allow a comparison between v-type and h-type measures, we computed the weighted mean distance between all haplotypes, taking into account their frequencies, and excluding the comparisons between identical haplotypes (i.e., the mean number of different restriction fragments between two different haplotypes selected at random in the total population): $\overline{D} = \sum_{ij} x_i x_j \pi_{ij} / \sum_{i,j \neq i} x_i x_j$. Each distance π_{ij} was divided by this value, so that when all distances are equivalent, v should be equal to h (i.e., $E(v) = h$). For all analyses (per species, and/or per region), the same value of $\overline{D}$ was used computed on the basis of the complete data set (involving all haplotypes and their frequencies in the total population); in this way, all values of v can be directly compared. To compare N_{ST} and G_{ST}, Burban et al. (1999) used a permutation approach that provides a way to evaluate the existence of a phylogeographic structure. The test is designed to answer the following question: when two different haplotypes are found within a population, will they differ less (i.e., have fewer distinct restriction fragments) than would be expected by chance? Here, we extend this test to the comparisons of v_S versus h_S and of v_T versus h_T. Negative values for the difference $(v - h)$ are expected when the haplotypes of the subpopulation considered represent only a small part of the total phylogenetic diversity. A total of 500 permutations of haplotypes identities were carried out for each data set. The 500 permutated values of v_S, v_T and N_{ST} were ranked and the distribution compared to the observed value without permutation to infer significance levels. Levels of diversity and differentiation were computed for each species and for each of eight European regions (the same than in the nine regional papers), except that the data for Scandinavia (Jensen et al., 2002) was pooled with that of the Baltic countries and Poland (Csaikl et al., 2002), to have sufficient sample sizes for meaningful comparisons. These regions are (1) the Alps (parts of Austria, France, Germany, and Italy; see Csaikl et al., 2001; Mátyás, 1999), (2) 'northern Balkans' (part of Austria, Bosnia and Herzegovina, Croatia, Hungary, Rumania, Slovakia, Yugoslavia; see Bordács et al., 2002), (3) 'central Europe' (part of Austria, Belgium, Czech Republic, Germany, the Netherlands; see König et al., 2002), (4) 'northern Europe' (the Baltic countries, Poland and Scandinavia; see Csaikl et al., 2002; Jensen et al., 2002), (5) France (Petit et al., 2002b), (6) Great Britain (Cottrell et al., 2002), (7) the Iberian peninsula (Olalde et al., 2002), and (8) the Italian peninsula (including Corsica; see Fineschi et al., 2002). Note that some of these regions are partly overlapping, especially France, central Europe and Italy with the Alpine region, so that the results are

not fully independent. Analyses were also made for species within these regions, provided that the number of populations was large enough (>25 populations). All populations considered to be introduced or of dubious origin were eliminated from these analyses, as well as those for which sample size was lower than three individuals.

In order to compare allelic richness across regions, we standardised to a common sample size of 100 trees using the software CONTRIB (available at http://www.pierroton.inra.fr/genetics/labo/Software/).

3. Results

3.1. cpDNA polymorphisms

The haplotypes have been detected on the basis of the information provided by four PCR fragments, each digested by one restriction enzyme: 25 of these haplotypes have been previously described in white oaks (22 in Dumolin-Lapègue et al., 1997 (haplotype 23 was eventually identified as a mislabelled individual of an oak belonging to the Cerris section), and three more in Dumolin-Lapègue et al., 1998). Twenty haplotypes are described for the first time, along with the patterns corresponding to three other European oak species not included in the survey (*Q. cerris* (=ex haplotype 23), *Q. suber*, *Q. ilex*). Four restriction diagrams summarising the polymorphisms observed in acrylamide gels are provided for the 45 types (DT–*Taq*I, AS–*Hin*fI, CD–*Taq*I, and TF–*Hin*fI, see Appendix A). Other (usually rare) haplotypes discovered during the course of the project are not described here because they could not be verified by the INRA laboratory in time to be incorporated into the present synthesis.

Although 45 types were detected, the number of haplotypes that will be considered in the diversity analyses and for the maps is 32. Haplotypes likely to have been confused with each other during the survey (because they involve fragments characterised by many length variants or because they were not scored systematically by all participants) were grouped at the lowest level of resolution available to all partners and were labelled with subscripts (e.g. haplotype 5 includes 5a, 5b, etc.). There were six composite haplotypes (i.e., haplotypes for which two or more sub-types could be identified): 4, 5, 10, 12, 17 and 24.

3.2. Phylogenetic relationships between the haplotypes

Among the 45 haplotypes, 63 length variants were detected in 20 polymorphic restriction fragments corresponding to the four primers–enzyme combinations, i.e., a mean of 3.1 length variants per polymorphic restriction fragment. Eleven additional fragments were monomorphic. The inclusion of three other species used as outgroups (*Q. cerris*, *Q. ilex* and *Q. suber*) raised the number of length variants to 71 and that of the polymorphic fragments to 22. The matrix that forms the basis of the phylogenetic analyses is provided in Appendix B. Partial sequence information of the fragment DT allowed the point mutation in fragment DT3 to be distinguished from length variants and this was therefore coded as a separate character (Dumolin-Lapègue et al., 1999b). In total, including the two additional point mutations, there were 57 phylogenetically informative characters (shared by at least two haplotypes) and 10 autapomorphies (mutations scored in a single haplotype) within the white oak group. There were two more autapomorphies present in *Q. ilex* and one for *Q. suber*, and five characters were unique to both *Q. suber* and *Q. ilex*. The phylogenetic trees obtained using the Fitch and Kitsch algorithms were largely consistent (compare Figs. 1 and 2). Lineage C (haplotypes 1, 2 and 3, located mainly from Italy to Scandinavia) includes two divergent groups of haplotypes (1, and 2 and 3) and is therefore poorly resolved. This lineage was slightly better resolved in the neighbour-joining analysis obtained by Dumolin-Lapègue et al. (1997), where other fragments had also been studied. On the other hand, lineage B (haplotypes located mostly in western Europe: 10–12, 24, 25 and 32; see Fig. 3) and lineage A (haplotypes 4–7, 26, and 30–31) are clearly distinguished. However, the definition of lineage A in the analysis presented here was still not as good as that in the previous analysis of Dumolin-Lapègue et al. (1997). The two algorithms (Fitch and Kitsch) give a somewhat different picture for the remaining haplotypes. Among them, a first group of haplotypes can be distinguished, particularly when using the Kitsch procedure. It includes haplotypes found around the western part of the Mediterranean region: from eastern Spain (27–29, 33) to south-western France

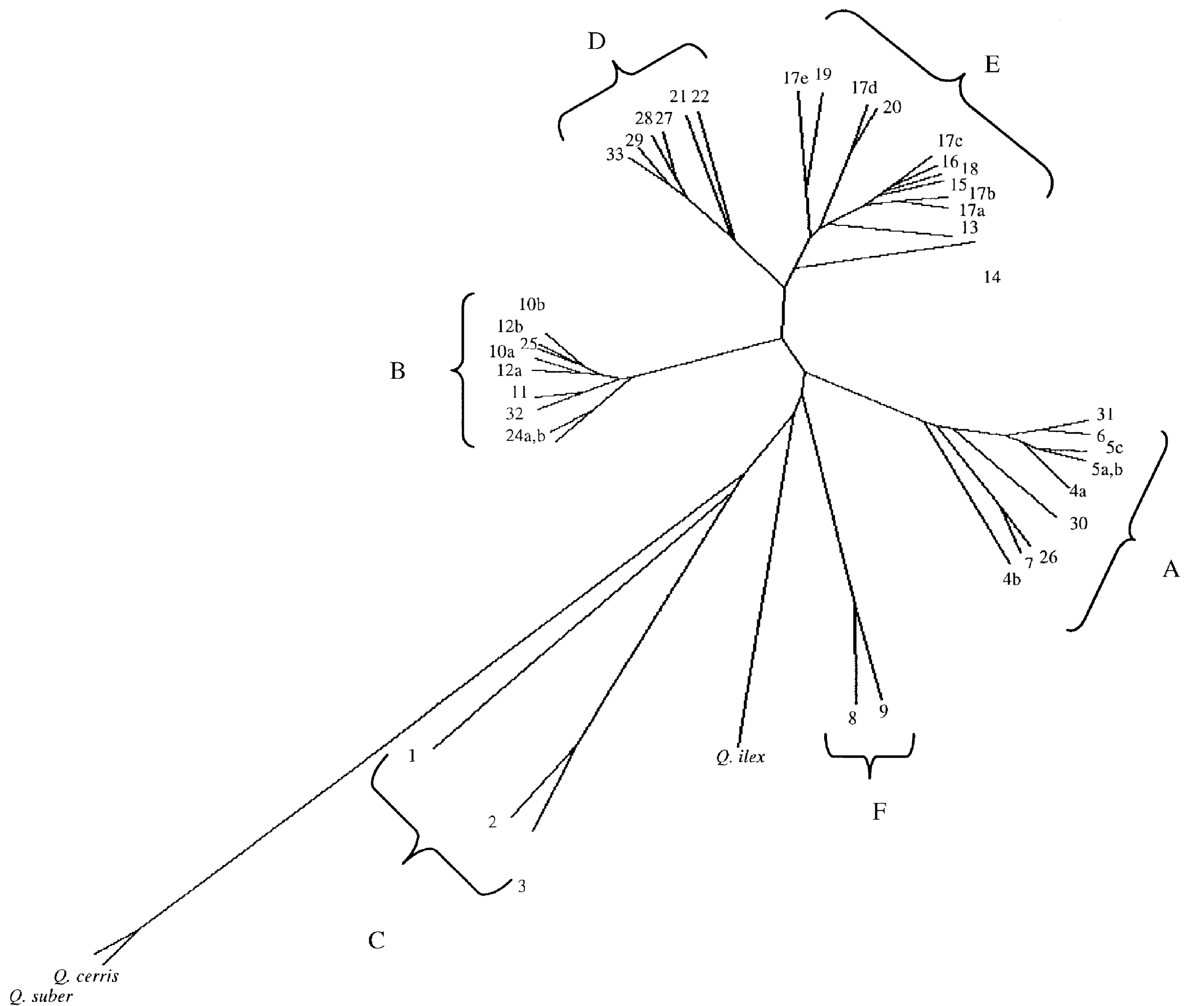

Fig. 1. Phylogenetic tree of the 45 cpDNA haplotypes obtained using the Fitch algorithm. All 42 haplotypes identified in the ROBUR complex and those corresponding to one individual from each of three outgroup species (*Q. ilex*, *Q. suber* and *Q. cerris*) are represented. The lineages defined in the text (A–F) are shown.

(21) and Algeria (22). This group is labelled lineage D here (Fig. 3). Of the remaining haplotypes, haplotypes 13–20 have a relatively basal location in the Fitch tree, between lineages B and D. However, they are grouped by the Kitsch method, although haplotype 14 is somewhat more divergent (Figs. 1 and 2) and haplotypes 17d and 20 are in an intermediate position with lineage D. This lineage (labelled E) is therefore likely to be provisory. However, all haplotypes belonging to this group have a more eastern distribution (from Italy to Rumania, up to Turkey; Fig. 3). The last two remaining haplotypes, found near the Black Sea (8 and 9), form a separate group that was also well supported by bootstrap values in the analysis of Dumolin-Lapègue et al. (1997), and is labelled lineage F.

3.3. Distribution of haplotypes across species

Eight species were sampled in this survey (Fig. 4). The number of populations of each species in which the 32 haplotypes were found is presented in Table 1. Haplotypes are shared by up to six of the eight species. For instance, the two most frequent species, *Q. robur*

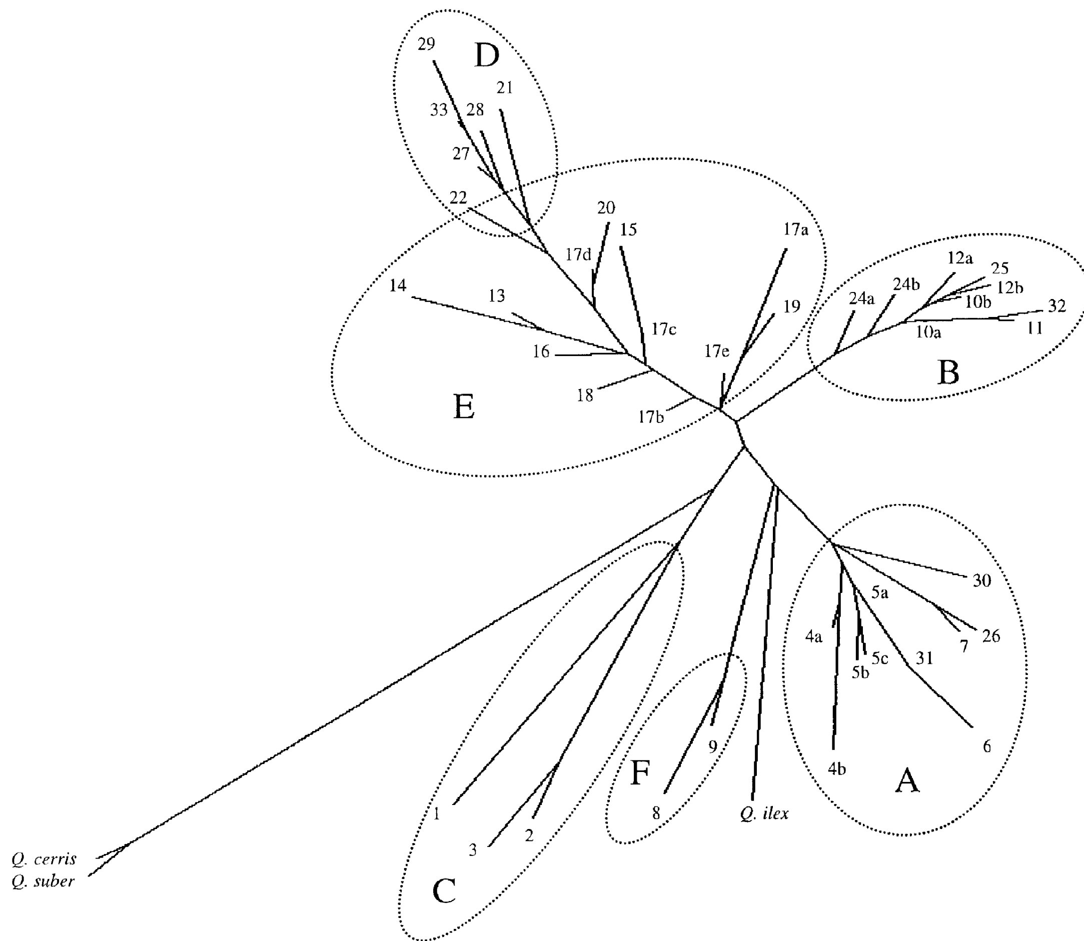

Fig. 2. Phylogenetic tree of the 45 cpDNA haplotypes obtained using the Kitsch algorithm. Same legend as in Fig. 1.

(present in 1469 populations) and *Q. petraea* (present in 1054 populations), share 15 haplotypes (from lineages A, B, C and E). Only the very rare haplotypes are restricted to one species. Compared to the previous study (Dumolin-Lapègue et al., 1997), it is now clear that *Q. canariensis* (formerly represented by only one population) belongs to the white oak cpDNA complex, as it shares five haplotypes belonging to two of the white oak cpDNA lineages (Olalde et al., 2002). On the other hand, no intertaxon exchanges have been reported so far between white oaks and *Q. ilex* and with *Q. cerris* or *Q. suber*. These species are characterised by very distinct lineages (Figs. 1 and 2), so that in principle such exchanges would have been identified readily.

3.4. Analysis of cpDNA diversity

In total, 2264 populations samples contained three or more individuals. The total diversity (h_T) was 0.848, and the coefficient G_{ST} was 0.828, significantly lower than N_{ST} (0.876) (Table 3).

3.4.1. Variation across species

Low sample sizes precluded analyses for *Q. canariensis* and *Q. macranthera*; so the analyses were restricted to the six remaining species.

3.4.1.1. Measures of diversity. As expected, the three most abundant and widespread oak species in Europe (*Q. robur*, *Q. petraea*, and *Q. pubescens*) have the

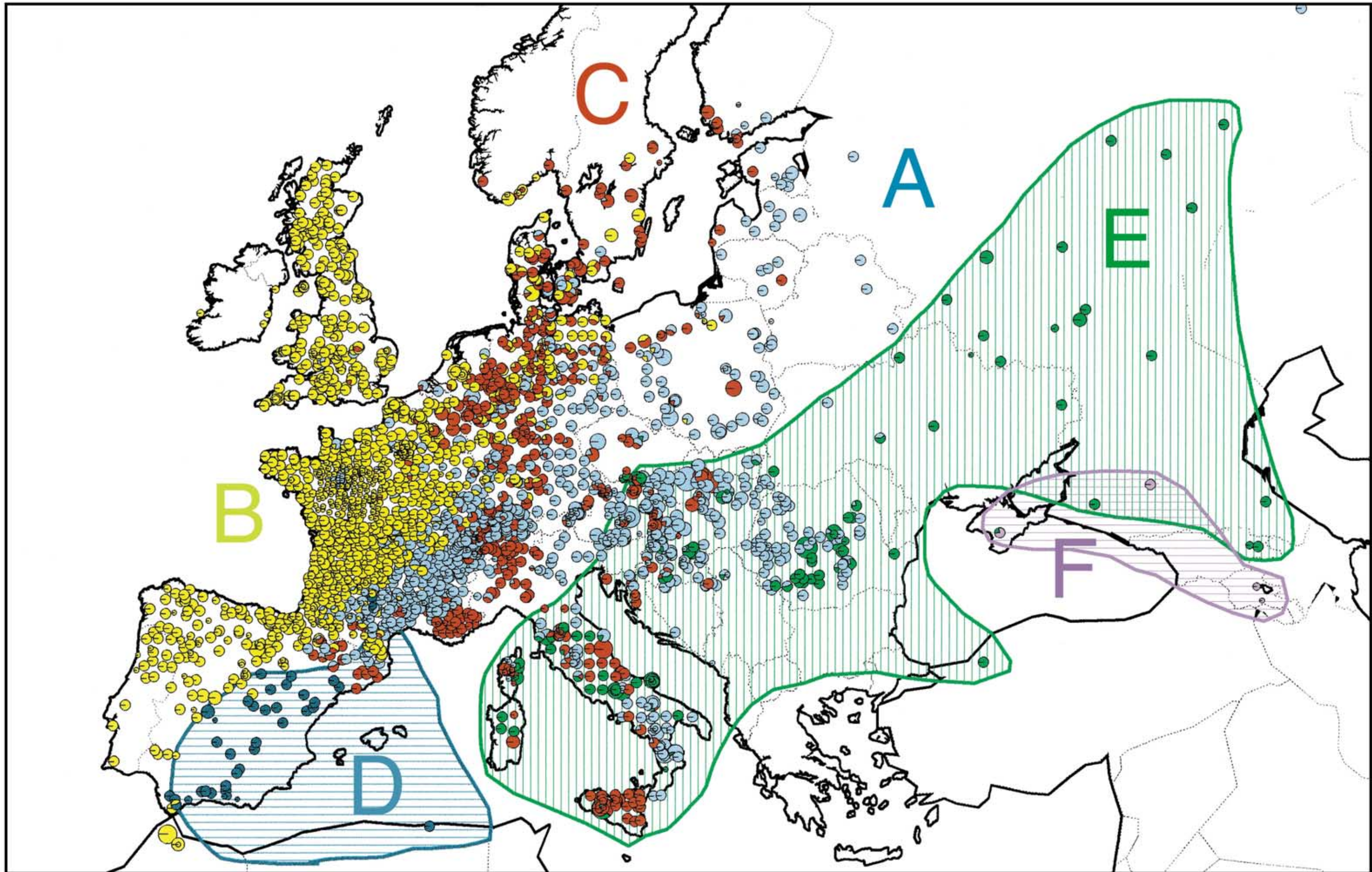

Fig. 3. Map of cpDNA lineages in Europe. The geographic distribution of the six cpDNA lineages identified (A–F) is provided. Different haplotypes belonging to the same lineage were pooled and are represented by the same colour. The limits of distribution for three of the six lineages (D–F) are shown.

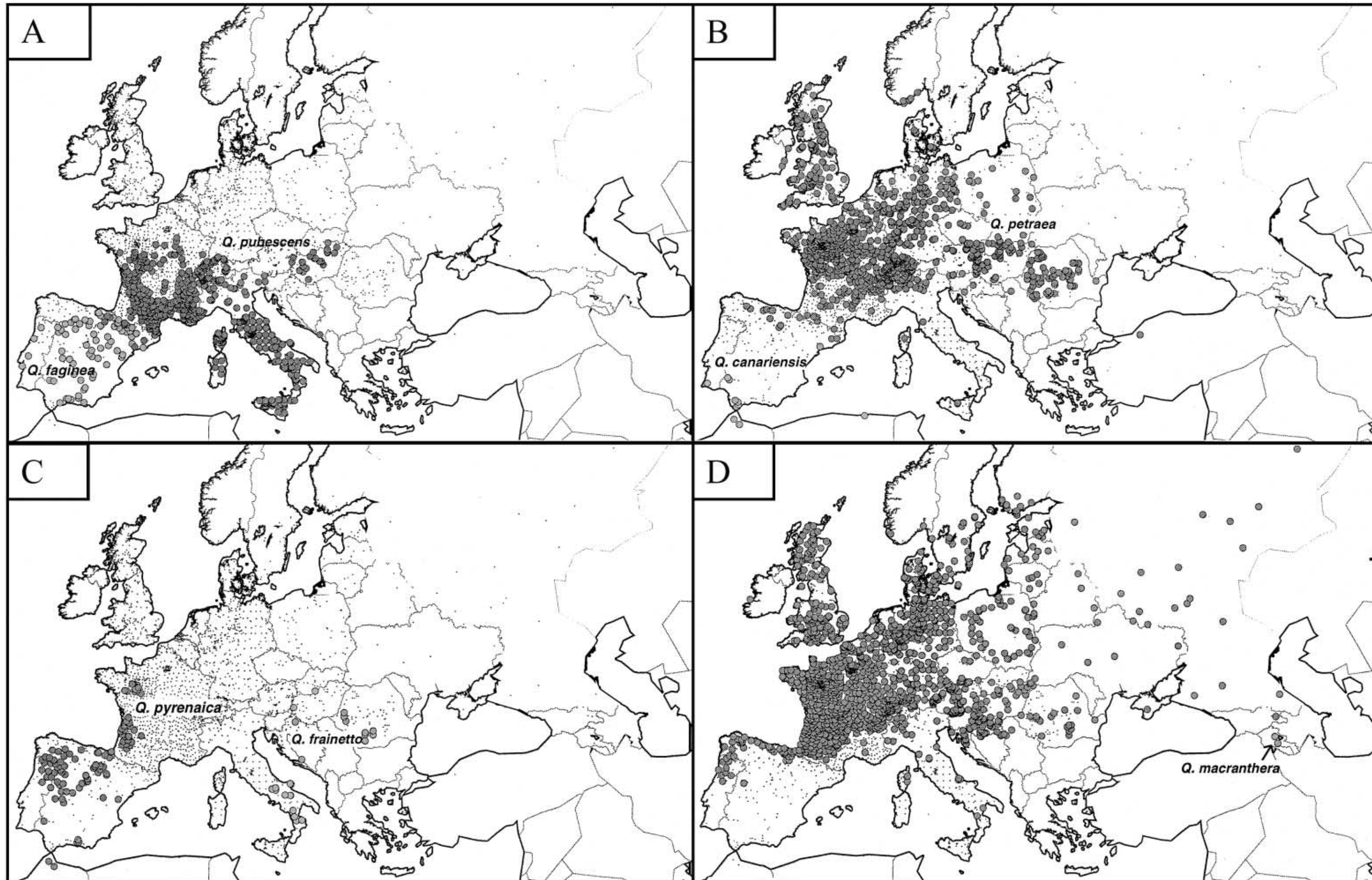

Fig. 4. Geographic distribution of the eight oak species sampled: (A) *Q. pubescens* and *Q. faginea*; (B) *Q. canariensis* and *Q. petraea*; (C) *Q. frainetto* and *Q. pyrenaica*; (D) *Q. macranthera* and *Q. robur*.

Table 3
Levels of diversity and differentiation by species

	No. of populations (≥3 individuals)	Harmonic mean number of individuals per population	No. of haplotypes	h_S (standard error)[a,b,c]	v_S (standard error)	h_T (standard error)	v_T (standard error)	G_{ST} (standard error)	N_{ST} (standard error)
Q. robur	984	4.29	23	0.183 (0.009)	0.837 (0.004)	0.781 (0.010)	0.197 (0.010)[ns]	1.034 (0.011)[ns]	0.809 (0.010)[*]
Q. frainetto	23	4.77	9	0.125 (0.050)	0.745 (0.062)	0.832 (0.065)	0.071 (0.030)[*]	0.457 (0.067)[*]	0.844 (0.066)[ns]
Q. petraea	650	4.23	17	0.122 (0.009)	0.847 (0.005)	0.856 (0.011)	0.091 (0.008)[ns]	0.835 (0.011)[ns]	0.891 (0.010)[**]
Q. pubescens	331	4.37	13	0.081 (0.011)	0.824 (0.010)	0.902 (0.013)	0.072 (0.011)[ns]	0.854 (0.019)[ns]	0.915 (0.013)[ns]
Q. faginea	58	4.18	10	0.043 (0.021)	0.785 (0.036)	0.946 (0.026)	0.055 (0.027)[ns]	0.843 (0.051)[ns]	0.934 (0.029)[ns]
Q. pyrenaica	44	3.96	6	0.025 (0.018)	0.635 (0.061)	0.961 (0.028)	0.008 (0.006)[ns]	0.257 (0.073)[ns]	0.968 (0.024)[ns]
All species	2353	4.62	32	0.146 (0.005)	0.848 (0.003)	0.828 (0.006)	0.105 (0.005)[**]	0.848 (0.007)[ns]	0.876 (0.006)[**]

[a] Test of the hypothesis that v_S differs from h_S.
[b] Test of the hypothesis that v_T differs from h_T.
[c] Test of the hypothesis that N_{ST} differs from G_{ST}.
[ns] Not significant ($p > 0.05$), [*]: significant ($0.01 < p < 0.05$), [**]: highly significant ($p < 0.01$).

highest levels of total chloroplast diversity, as measured by h_T, with values above 0.8. *Q. faginea* and *Q. frainetto* have intermediate values, whereas *Q. pyrenaica* is characterised by a lower diversity ($h_T = 0.635$). If the genetic distance measures between haplotypes are taken into account, the diversity values become more heterogeneous across species: the phylogenetic diversity v_T is only 0.26 in *Q. pyrenaica*, 0.457 in *Q. frainetto*, around 0.80 for *Q. faginea*, *Q. petraea* and *Q. pubescens*, and above 1 for *Q. robur*. Two species (*Q. robur* and *Q. faginea*) have higher v_T than h_T values, whereas the contrary is true for the other four species (the difference between the two measures being significant in the case of *Q. frainetto*). *Q. robur* is composed of 22 haplotypes of all major lineages, in a somewhat more balanced way than in the complete data set, hence its v_T values higher than 1. In contrast, v_T in *Q. pyrenaica* is much lower than h_T: this species is composed of six haplotypes, five of which belong to the same lineage (B), whereas *Q. frainetto* is composed of 10 haplotypes that belong to only two lineages (A and E).

3.4.1.2. Coefficients of differentiation. G_{ST} varies from 0.781 in *Q. robur* to 0.961 in *Q. pyrenaica*, whereas N_{ST} values are usually higher than the corresponding G_{ST} values (except for *Q. faginea*; see Table 3). N_{ST} values are therefore more similar among species, from 0.809 in *Q. robur* to 0.968 in *Q. pyrenaica*. Seven of the 15 pairwise tests between species are significant for G_{ST}, and only five for N_{ST} (results not shown). In two cases, the difference between the two coefficients of differentiation is significant: for *Q. robur* and for *Q. petraea*, the largest difference being observed in *Q. petraea*. The ranking across species is the same for G_{ST} and N_{ST}: *Q. robur* < *Q. frainetto* < *Q. petraea* < *Q. pubescens* < *Q. faginea* < *Q. pyrenaica*. In a previous study of French material (Dumolin-Lapègue et al., 1999a), the same ranking was found for *Q. robur*, *Q. petraea*, and *Q. pubescens*, but only the difference between the two extremes (*Q. robur* and *Q. pubescens*) was significant. Here, in addition to confirming the trend for these three species (all pairwise comparisons being highly significant), we demonstrate that the two western European species (*Q. faginea* and *Q. pyrenaica*) are characterised by even higher G_{ST} values.

3.4.2. Variation across regions

3.4.2.1. Allelic richness. Despite the considerably larger sampling effort in northern Europe (a total of 9128 individuals had been studied north of the 45°N parallel, compared to 2803 south of it), only three haplotypes (8, 14, and 26) were found north of this limit that were not detected south of it. In contrast, 12 haplotypes were found south of this limit but did not occur in the northern area, and 17 were common to both regions. Among the eight geographic regions compared, the observed number of haplotypes varies from 5 in Great Britain to 12 in the Balkans or in Spain (Table 4). However, some of these haplotypes are very rare or of uncertain origin. For instance, the number of autochthonous haplotypes in Great Britain

is more likely to be 3 than 5 (see Cottrell et al., 2002). Moreover, sample sizes vary across regions (from 855 trees in Scandinavia to 3605 in France; see Table 4). To obtain a more representative idea of the haplotype richness, sample sizes were standardised to 100 trees selected at ran-dom from the entire data sets of each region. Standardised estimates of haplotype richness vary from 4.2 in Great Britain, 5.5 in the Alpine region and in France, to 9.7 in Spain and 10.3 in the Balkans. The effect of the correction is particularly clear for France, where an extensive survey revealed as many as 11 haplotypes, but where only 5.5 are expected in a random sample of 100 trees. Despite its southern position and its role as ice-age refugium, Italy has a lower level of allelic richness (10 haplotypes but 6.7 after rarefaction) than regions further north such as central Europe or northern Europe (11 and 9 haplotypes observed, and 8.7 and 8.2 expected among 100 trees, respectively).

3.4.2.2. Genetic diversity. The lowest levels of total diversity (h_T) are again observed in Great Britain and in the Alpine region. However, contrary to the results for allelic richness, the lowest total diversity is observed for the Alpine region ($h_T = 0.52$), where only two haplotypes (1 and 7) predominate (compared to 3 in Great Britain where $h_T = 0.64$). The most diverse regions based on this parameter are again the Iberian peninsula and the northern Balkans (about 0.80), this time followed by Italy ($h_T = 0.78$). However, when the phylogenetic information is taken into account, the rankings completely change: based on v_T, central Europe is the most diverse region (0.86), followed by northern Europe and by Italy (0.78). Great Britain had by far the lowest level of diversity (0.21), the second lowest value was for northern Balkans (0.62). The v-type measure is significantly lower than its h-type equivalent in the case of Great Britain (Table 5). France, and especially the Iberian and Balkan peninsulas also have a deficit in v compared to h, though not significantly so.

3.4.2.3. Coefficients of genetic differentiation. In the Iberian and the Italian peninsula, as well as in northern Europe and in the Alpine region, a higher proportion of diversity is partitioned among populations (as measured by both G_{ST} and N_{ST}), whereas the lowest values are for central Europe and Great Britain. The values vary from 0.73 to 0.89 in the case of G_{ST} and from 0.74 to as much as 0.93 for N_{ST}. The coefficient of differentiation (G_{ST}) among the eight regions is much lower but non-negligible: 0.18; furthermore, N_{ST} among regions is significantly higher ($N_{ST} = 0.286$, $P < 0.01$). This means that the differentiation among regions is due in part to the high phylogenetic divergence between haplotypes, and that the overall differentiation among populations is due for about one-third to a differentiation between regions, when based on N_{ST}. The overall phylogeographic effect is still discernible at a within region level. Indeed, in six regions (i.e., all but the Alpine region and Great Britain), N_{ST} was larger than G_{ST}, and the differences are significant in three cases (France, northern Balkans and northern Europe). In France, the two estimates had the largest difference (0.08); relatively large differences are also present in the Balkans, northern Europe and Iberia; however, for the Alpine region and Great Britain there was no indication of a phylogeographic component to the observed genetic structure. This is expected given their genetic composition, which consists of two dominant haplotypes in the Alpine region and of three equally divergent haplotypes all from lineage B in Great Britain (and hence no way to detect phylogeographic structure). In the other regions, the potential to detect a difference between N_{ST} and G_{ST} is higher. In France, for instance, the three closely related haplotypes (10–12) of the B lineage have a similar distribution and are commonly found in the same populations; similarly, in the Balkans, haplotypes of lineages E are dominating east of the Carpathians. Hence, mixtures of related haplotypes are more frequent there than mixtures of divergent haplotypes, which accounts for the observed difference between N_{ST} and G_{ST}.

3.4.3. Variation across regions within species

Differences between regions for G_{ST} and/or N_{ST} are still observed for several areas when analyses are restricted to the intraspecific level (Table 5). Indeed, in general, the ranking of overall genetic differentiation across regions also applies at the species level; for instance, for both *Q. robur* and *Q. petraea*, the highest values of G_{ST} and N_{ST} are in northern Europe, whereas central Europe has low values for both species. The Alpine region has the highest G_{ST} value for both *Q. pubescens* and *Q. petraea*, and the third highest for

Table 4
Sample sizes and number of haplotypes detected in each region

Region	No. of trees	No. of populations	No. of haplotypes	No. of haplotypes/100 individuals	% Mixed populations	% Allochthonous populations	*Q. robur* (%)	*Q. petraea* (%)	*Q. pubescens* (%)	*Q. faginea* (%)	*Q. frainetto* (%)	*Q. pyrenaica* (%)	*Q. canariensis* (%)	*Q. sp.* (%)
Alpine region	2091	451	8	5.5	29.7	2.0	41.2	30.9	26.6					1.4
France	3605	878	11	5.5	41.5	0.0	46.6	29.3	22.2			1.3		0.6
Central Europe	2155	426	11	8.7	9.2	5.2	56.3	40.8						2.9
Great Britain	1148	240	5	4.2	45.4	5.4	41.9	37.9						20.2
North Balkans	1113	222	12	10.3	16.2	5.9	39.0	47.9	6.6		4.9			1.6
Northern Europe	855	172	9	8.2	5.2	12.8	83.6	16.1						0.2
Iberian peninsula	858	198	12	9.7	20.7	0.0	6.4	3.2	33.0	5.1		6.8	24.5	4.8
Italian peninsula	922	194	10	6.7	8.2	1.5	8.8	4.1	79.6		7.5			
Total Europe	11937	2613	32	12.6	27.3	2.8	43.9	29.5	15.1	2.3	1.0	2.1	0.4	3.3

Table 5
Levels of diversity and differentiation by region

	No. of populations	Harmonic mean number of individuals per population	No. of haplotypes	h_S (standard error)	v_S (standard error)	h_T (standard error)	v_T (standard error)	G_{ST} (standard error)	N_{ST} (standard error)
Q. robur									
Alpine region	168	4.27	8	0.120 (0.018)	0.150 (0.023)^{ns}	0.478 (0.035)	0.561 (0.042)^{ns}	0.749 (0.036)	0.733 (0.037)^{ns}
France	305	3.95	9	0.178 (0.016)	0.177 (0.017)^{ns}	0.724 (0.012)	0.740 (0.019)^{ns}	0.755 (0.022)	0.761 (0.023)^{ns}
Central Europe	213	4.79	10	0.247 (0.020)	0.260 (0.023)^{ns}	0.749 (0.016)	0.853 (0.011)^{ns}	0.671 (0.026)	0.696 (0.027)^{ns}
Great Britain	93	4.19	5	0.232 (0.030)	0.063 (0.013)*	0.665 (0.014)	0.166 (0.016)*	0.651 (0.045)	0.622 (0.054)^{ns}
North Balkans	82	4.18	10	0.222 (0.033)	0.169 (0.031)^{ns}	0.788 (0.026)	0.618 (0.055)^{ns}	0.718 (0.042)	0.727 (0.052)^{ns}
Northern Europe	116	4.79	9	0.099 (0.019)	0.079 (0.019)^{ns}	0.733 (0.022)	0.763 (0.025)^{ns}	0.864 (0.026)	0.896 (0.025)*
Iberian peninsula	42	3.84	5	0.179 (0.043)	0.035 (0.009)^{ns}	0.649 (0.036)	0.158 (0.022)^{ns}	0.725 (0.067)	0.780 (0.056)*
Q. petraea									
Alpine region	121	4.20	6	0.062 (0.016)	0.075 (0.020)^{ns}	0.574 (0.038)	0.642 (0.048)^{ns}	0.892 (0.029)	0.884 (0.032)^{ns}
France	170	3.73	6	0.118 (0.019)	0.061 (0.014)^{ns}	0.712 (0.018)	0.539 (0.033)^{ns}	0.835 (0.026)	0.886 (0.025)^{ns}
Central Europe	169	4.62	9	0.154 (0.019)	0.174 (0.023)^{ns}	0.765 (0.021)	0.857 (0.017)^{ns}	0.799 (0.026)	0.797 (0.027)^{ns}
Great Britain	84	4.06	4	0.094 (0.025)	0.034 (0.014)^{ns}	0.605 (0.042)	0.239 (0.047)^{ns}	0.845 (0.040)	0.857 (0.055)^{ns}
North Balkans	97	4.60	11	0.139 (0.025)	0.070 (0.014)**	0.792 (0.029)	0.593 (0.051)^{ns}	0.825 (0.033)	0.883 (0.025)**
Northern Europe	27	4.44	8	0.107 (0.036)	0.068 (0.029)^{ns}	0.842 (0.032)	0.892 (0.042)^{ns}	0.873 (0.042)	0.924 (0.033)*
Q. pubescens									
Alpine region	106	4.43	5	0.037 (0.015)	0.043 (0.017)^{ns}	0.516 (0.026)	0.669 (0.033)^{ns}	0.929 (0.029)	0.936 (0.026)^{ns}
France	152	4.32	6	0.089 (0.018)	0.068 (0.017)*	0.687 (0.028)	0.799 (0.030)^{ns}	0.870 (0.026)	0.915 (0.021)**
Italian peninsula	141	4.60	8	0.094 (0.017)	0.088 (0.017)^{ns}	0.774 (0.015)	0.775 (0.028)^{ns}	0.879 (0.022)	0.888 (0.022)^{ns}
All species									
Alpine region	409	4.78	8	0.076 (0.010)	0.094 (0.012)^{ns}	0.517 (0.020)	0.618 (0.025)^{ns}	0.853 (0.018)	0.848 (0.019)^{ns}
France	733	4.32	11	0.159 (0.010)	0.086 (0.007)*	0.743 (0.006)	0.642 (0.015)^{ns}	0.786 (0.013)	0.867 (0.012)**
Central Europe	389	4.95	10	0.208 (0.014)	0.227 (0.016)^{ns}	0.756 (0.013)	0.860 (0.010)^{ns}	0.726 (0.019)	0.736 (0.019)^{ns}
Great Britain	222	4.81	5	0.159 (0.017)	0.052 (0.009)*	0.641 (0.018)	0.207 (0.022)*	0.752 (0.026)	0.750 (0.038)^{ns}
North Balkans	193	4.80	12	0.182 (0.020)	0.121 (0.016)*	0.803 (0.018)	0.615 (0.036)^{ns}	0.773 (0.025)	0.803 (0.027)*
Northern Europe	141	4.83	9	0.111 (0.018)	0.084 (0.017)^{ns}	0.750 (0.019)	0.785 (0.023)^{ns}	0.852 (0.024)	0.893 (0.022)*
Iberian peninsula	179	4.38	12	0.089 (0.016)	0.045 (0.012)*	0.799 (0.016)	0.671 (0.046)^{ns}	0.889 (0.020)	0.933 (0.017)^{ns}
Italian peninsula	171	4.78	10	0.098 (0.016)	0.089 (0.015)^{ns}	0.776 (0.013)	0.782 (0.025)^{ns}	0.873 (0.021)	0.886 (0.020)^{ns}
Among regions	8	1248	27	0.724 (0.036)	0.646 (0.071)^{ns}	0.883 (0.027)	0.904 (0.027)^{ns}	0.181 (0.036)	0.286 (0.077)**

Q. robur. However, there is some interaction; for instance, in Great Britain, *Q. robur* has very low G_{ST}/N_{ST} values but *Q. petraea* has average ones.

4. Discussion

This survey of cpDNA variation in European oaks, although building on already established methods (Dumolin-Lapègue et al., 1997), constitutes a significant step forward both quantitatively and qualitatively. First, the data set has increased by an order of magnitude, as the number of trees typed is 12,214 individual oak trees, compared to 1412 in the previous survey. Simultaneously, the number of populations has risen from 345 to 2613. Several oak species are now represented by significant numbers of populations, so that their diversity can be better compared: this is the case for *Q. faginea*, *Q. frainetto* and *Q. pyrenaica*, in addition to the three more abundant species *Q. robur*, *Q. petraea* and *Q. pubescens*. In fact, the present survey also helps to determine more accurately the geographic distribution of some of the taxa studied, and concentrates on those oak species that can be reasonably identified in the field by non-specialists. Cumbersome oak taxonomy and description of excessive number of poorly supported taxa, as well as the lack of scientific exchanges across countries, had long impeded the progress in the field of oak biogeography, ecology and genetics.

Although the present survey encompasses roughly the same range as that studied in Dumolin-Lapègue et al. (1997), several regions that had been poorly sampled have now been investigated in much more detail such as the Iberian peninsula, the north of the Balkans, Scandinavia, etc. However, further collaborative efforts and new partnerships with countries situated in North Africa, in south-eastern and eastern Europe, and around the Black Sea, will be necessary to obtain a complete overview of the biogeography of white oaks across Europe based on cpDNA.

In total, 45 haplotypes were described and, after pooling some that could not be screened systematically, 32 were mapped across Europe. Haplotypes described for the first time here originate from mid- to low-latitudes, and in regions that had previously only been sparsely sampled, specifically the Iberian peninsula and Rumania. The discovery of these additional haplotypes, together with the careful verification of the cpDNA patterns and the addition of the TF–*Hin*fI data, provide new insights into the intraspecific phylogeny of this group.

Most notable is the detection of a western Mediterranean lineage (called D), which consists of several new Iberian haplotypes (those found in the eastern part of Spain), and of haplotypes 21 and 22 from southern France and from Algeria, respectively. The distribution of this lineage is centred in the western part of the Mediterranean region, west of the Italian peninsula. Two other haplotypes from Corsica, Sardinia and the Italian continent (17d and 20) appear to be intermediate between this lineage D and lineage E. The opening of the Gulf of Lions and the counter-clockwise rotation of Corsica and Sardinia away from the Iberian peninsula took place during the early Miocene (i.e., about 20 million years ago, de Jong, 1998), and is probably too early to account for this phylogeographic affinity. Further phylogenetic and biogeographic studies are therefore needed to clarify this relationship. On the basis of the information currently available, these two haplotypes have been grouped in lineage E. The provisory recognition of this rather basal group of haplotypes centred in the Balkans (labelled group E), which also extend into Italy, is also noteworthy. Some of the haplotypes comprising this group originate from the eastern part of the Balkan region, south of the Carpathian mountains (13–16 and 17c). In contrast, lineage A, which is also predominant in the Balkans, may have had a more western origin. The identification of a small but quite divergent lineage (labelled lineage F) centred in the eastern part of the Black Sea and composed of haplotypes 8 and 9 (Dumolin-Lapègue et al., 1997) highlights the need for further studies in this region. Finally, as judged by the comparison amongst the three outgroups, it appears that the divergence between some lineages defined here may be as important as that between different sections of the genus. For example, although *Q. cerris* and *Q. suber* (which belong both to the Cerris section, Manos et al., 1999) are clearly divergent from the white oak complex, this is not the case of *Q. ilex*. At low taxonomical level, cpDNA sequence divergence may therefore constitute a poor predictor of the ability of oak species to exchange genes. The high level of sequence divergence between *Q. suber* and *Q. ilex* has not precluded repeated exchanges between these species, both in Morocco and

in Spain (Belahbib et al., 2001, P. Jimenez, L. Gil and R.J. Petit, unpublished data). However, by contrast, the relatively moderate divergence of the cpDNA genomes of *Q. ilex* and the white oak complex is apparently accompanied by a rather strict gene exchange barrier. Indeed, during this survey, the *Q. ilex* haplotype has never been detected in any white oak individual. Similarly, the more divergent *Q. cerris/Q. suber* haplotypes were not found in white oaks.

Across the whole sample set, areas where refugia have been postulated (Huntley and Birks, 1983; Brewer et al., 2002; Petit et al., 2002a) such as the Iberian and the Italian peninsula, generally displayed higher levels of chloroplast diversity. Superimposed on this 'refugium' effect, a 'mixing' effect was found for regions situated at higher latitudes, in central and in northern Europe. Those regions located at the junction between several colonisation routes (e.g. Petit et al., 2002a) were also found to display very high levels of diversity. In contrast, and as a consequence of the strong phylogeographic structure, regions such as southern Italy or southern Iberia contain only a subset of the lineages described and exhibit therefore lower levels of phylogenetic diversity. By focusing on just two criteria, the level of allelic richness (standardised to 100 trees) (high, intermediate, and low) and the relative values of the two diversity parameters h_T and v_T (e.g. $h_T \leq v_T$ or $h_T > v_T$), six different cases can be distinguished, each corresponding to a particular situation from an historical standpoint:

1. Allelic richness high, $h_T > v_T$. This case is indicative of refugia, since many variants are present (high allelic richness), but the variation is mainly confined within lineages. This situation characterises the Balkans and Iberia.
2. Allelic richness high, $h_T \leq v_T$. This situation is indicative of the mixing of haplotypes from several different lineages, resulting in many variants (high allelic richness) from different lineages ($h_T \leq v_T$). Three regions, northern and central Europe and the Italian peninsula, fit well with this pattern.
3. Allelic richness intermediate, $h_T > v_T$. This would indicates unbalanced recolonisation, i.e. uneven contribution of refugia to sampled populations. This corresponds to the situation in France, where recolonisation with haplotypes from lineage B from Iberia is predominating.
4. Allelic richness intermediate, $h_T \leq v_T$. This would indicate balanced recolonisation from few refugia, and corresponds to the situation in the Alpine region.
5. Allelic richness low, $h_T > v_T$. This would suggest recolonisation from a single refugium, so that variation, when it occurs, is restricted to differences within lineage. This case corresponds to the situation found in Great Britain.
6. Allelic richness low, $h_T \leq v_T$. Not observed and not to be expected, since $h_T \approx v_T$ indicates mixing of lineages that should be accompanied by an increase in allelic richness.

Patterns of partitioning of diversity across regions are best examined at the species level. *Q. robur*, *Q. petraea*, and even *Q. pubescens* usually displayed the same trends in the partitioning of diversity across regions. This finding suggests that there is a common underlying reason for this partitioning that is independent of species. The intensity of human management (through seed transfers and plantations) could account for these regional tendencies. Plantations are akin to seed flow, and as such, would decrease differentiation. The intensity of human management may be invoked to account for the low level of differentiation of both *Q. robur* and *Q. petraea* in central Europe (König et al., 2002). In *Q. robur*, the level of differentiation is also particularly low in Great Britain, whereas the values for *Q. petraea* are average in Great Britain. Special interest in the use of *Q. robur* (called the British oak in vernacular English) in seed transfers and plantation programs in this part of Europe could be the reason for its reduced differentiation.

The high level of differentiation in the northern Europe region (Scandinavia, Poland and the Baltic countries) is somewhat surprising. Like in Great Britain, Germany, and the Netherlands, the human impact has been large in this area and this might be expected to lead to a reduction in genetic differentiation. On the other hand, oak forests are rather fragmented in this region, being located at the northern edge of the distribution range. Such isolation of the stands may have resulted in this high coefficient of differentiation. Moreover, *Q. petraea* is absent from much of this region, so that exchanges between species, which constitute one of the avenue for cytoplasmic gene flow (Dumolin-Lapègue et al., 1999a), will also be limited. In the case of the Alpine

region, increased population isolation (due here to the high relief characteristic of this region) may also be invoked to explain the comparably high G_{ST} value in all three species growing there (*Q. robur*, and especially *Q. petraea* and *Q. pubescens*).

Human impact is not just limited to northern Europe. In the Iberian peninsula, *Q. robur*, which is present from northern Portugal to Galicia and the Basque country, has long been used by man, and is often found planted at low density for acorn production to feed cattle (Olalde et al., 2002). The relatively low G_{ST} value for *Q. robur* in this region could result from intense historical utilisation. In any case, it demonstrates that the very high G_{ST} value for *Q. pyrenaica*, which has a very similar range to that of *Q. robur* in the Iberian peninsula, constitutes a true biological difference. This latter species readily colonizes dry, acidic soils, and has the ability to spread vegetatively by root suckers, which could account for the high cpDNA differentiation. Most populations of *Q. faginea* are also fixed for a single cpDNA variant. This small species is of limited economical importance and is unlikely to have been planted by man. This reasoning also applies to *Q. pubescens* to a slightly lesser degree, as discussed in Dumolin-Lapègue et al. (1999a). The reasons for higher differentiation found for *Q. petraea* compared to *Q. robur* are most likely due to the higher dispersal abilities of *Q. robur*, compared to the more successional *Q. petraea* (Dumolin-Lapègue et al., 1999a). Finally, *Q. frainetto* also exhibited a relatively low level of differentiation, although not as low as that exhibited by *Q. robur*. Also called *aesculus* by the Romans because of its sweet acorns, *Q. frainetto* may have been planted in the past, near villages, resulting in some mixing of cpDNA types.

5. Conclusions and perspectives

The study of cpDNA variation in Europe has revealed striking patterns of phylogeographic structure that are largely species-independent. Nevertheless, these species do differ in the way they partition cpDNA diversity among populations, as well as in many other biological attributes that are maintained despite the extensive introgression typical of many oak species. As a consequence of this introgression and of the large cumulative range of these species in Europe and around the Mediterranean Sea, broad biogeographic questions can be addressed using phylogeographic methods; this can provide a bridge with the biogeographic studies based on interspecific phylogenies, which give information at an even broader geographic scale (Manos et al., 1999). For the purpose of recovering the interspecific phylogeny of this genus, a correct sampling of all the main lineages within such species complexes seems at least as important as the sampling of the species, given the relatively deep intraspecific divides that are being revealed. The use of nuclear sequences in combination with chloroplast phylogenies should also prove very profitable, as it would help identify past introgression events. Sequencing studies should also have the power to reveal statistically well-supported intraspecific phylogenies, which will be necessary to confirm the lineage grouping of haplotypes provisionally proposed here.

Acknowledgements

We are grateful to the numerous colleagues who provided access to oak material, as acknowledged in the nine regional companion papers. The study has been carried out with financial support from the European Communities, Agriculture and Fisheries (FAIR) specific programme, CT-FAIR1 PL95-0297, 'Synthetic maps of gene diversity and provenance performance for utilisation and conservation of oak genetic resources in Europe'. It does not necessarily reflect its views and in no way anticipates the Commission's future policy in this area. R.J.P. thanks the colleagues from the Bundesforschungsanstalt für Forst- und Holzwirtschaft, Institute für Forst Genetik, for their hospitality during the preparation of the manuscript.

Appendix A

Restriction diagrams. (A) AS–*Hin*fI. (B) CD–*Taq*I. (C) DT–*Taq*I. (D) TF–*Hin*fI. Molecular weight markers are indicated on both sides of the diagram, with the size (in bp) of the fragments. The other lanes correspond to the different patterns observed with this particular PCR fragment/restriction enzyme combination. One pattern may correspond to one or several haplotypes, which are listed below under the corresponding lane.

Appendix B

Table of the 45 haplotypes detected in this survey. For each haplotype, length variants observed with each polymorphic restriction fragment are listed (DT1: largest fragment for the combination DT–*Taq*I), ranked according to their migration in the gel: from 1 to 6, with 6 migrating further down the gel, and 9 representing point mutations. The two last columns correspond to polymorphisms (two point mutations) detected in the largest restriction fragment of the corresponding primers–enzyme combination (not illustrated).

Haplotype	Fragment																							
	DT1	DT2	DT3	DT3′	DT4	AS1	AS2	AS3	AS4	AS5	AS6	CD1	CD2	CD3	CD4	CD6	TF1	TF3	TF5	TF10	TF11	TF12	TF14	DT–*Alu*I	TF–*Cfo*I
1	9	1	2	1	1	2	4	2	2	2	3	1	2	3	3	1	2	0	2	0	2	1	2	1	1
2	9	1	2	1	1	1	4	2	9	2	2	1	9	2	2	1	2	0	2	0	2	1	1	1	1
3	9	1	2	1	2	1	4	2	9	2	2	1	9	2	2	1	2	0	2	0	2	1	2	1	1
4a	1	1	1	1	1	1	6	2	2	2	3	1	1	2	3	1	2	0	2	0	2	1	2	1	9
4b	1	1	1	1	1	1	6	2	2	2	3	1	1	2	3	1	1	0	2	0	2	0	2	1	9
5a	1	1	2	1	1	1	6	2	2	2	3	1	1	2	3	1	2	0	2	0	2	1	2	1	9
5b	1	1	2	1	1	1	6	2	2	2	3	1	1	2	3	1	2	0	2	0	2	1	3	1	9
5c	1	1	2	1	1	1	6	2	2	2	3	1	1	2	3	1	2	0	2	0	2	1	1	1	9
6	2	1	2	1	1	1	6	2	2	2	3	1	1	1	3	1	2	0	2	0	2	1	2	1	9
7	1	1	2	9	1	1	6	2	2	2	4	1	1	2	3	1	2	0	2	0	2	1	2	1	9
8	1	1	2	1	1	1	2	2	2	2	3	1	1	2	1	1	2	1	2	1	2	1	2	1	1
9	1	1	2	1	1	1	5	2	2	2	3	1	1	2	3	1	2	1	2	1	2	1	2	1	1
10a	1	2	3	1	1	1	4	2	2	2	2	1	1	2	3	1	2	0	2	0	2	1	2	9	1
10b	1	2	3	1	1	1	4	2	2	2	1	1	1	2	3	1	2	0	2	0	2	1	2	9	1
11	1	2	3	1	1	1	4	2	2	2	2	1	1	2	3	1	2	0	2	0	1	1	2	9	1
12a	1	2	4	1	1	1	4	2	2	2	2	1	1	2	3	1	2	0	2	0	2	1	2	9	1
12b	1	2	4	1	1	1	4	2	2	2	1	1	1	2	3	1	2	0	2	0	2	1	2	9	1
13	1	1	2	1	1	1	4	2	2	2	5	2	1	2	3	1	2	0	2	0	2	0	2	1	1
14	1	1	2	1	1	1	3	2	2	2	6	2	1	2	3	1	1	0	2	0	2	0	2	1	1
15	1	1	3	1	1	1	4	2	2	2	5	1	1	2	1	1	2	0	2	0	2	0	2	1	1
16	1	1	3	1	1	1	4	2	2	2	5	3	1	2	3	1	2	0	2	0	2	0	2	1	1
17a	1	1	3	1	1	1	4	2	2	2	3	1	1	2	3	1	2	0	2	0	2	1	2	1	1
17b	1	1	3	1	1	1	4	2	2	2	3	1	1	2	3	1	2	0	2	0	2	1	2	1	1
17c	1	1	3	1	1	1	4	2	2	2	4	1	1	2	3	1	2	0	2	0	2	0	2	1	1
17d	1	1	3	1	1	1	3	2	2	2	4	1	1	2	3	1	2	0	2	0	2	0	1	1	1
17e	1	1	3	1	1	1	4	2	2	2	3	1	1	2	3	1	2	2	2	0	2	1	3	1	1
18	1	1	3	1	1	1	3	2	2	2	4	1	1	2	3	1	2	0	2	0	2	0	2	1	1
19	1	1	3	1	1	1	4	2	2	2	4	1	1	2	3	1	2	0	2	0	2	1	3	1	1

Appendix B (*Continued*)

Haplotype	Fragment																								
	DT1	DT2	DT3	DT3′	DT4	AS1	AS2	AS3	AS4	AS5	AS6	CD1	CD2	CD3	CD4	CD6	TF1	TF3	TF5	TF10	TF11	TF12	TF14	DT–*Alu*I	TF–*Cfo*I
20	1	1	3	1	1	1	3	2	2	2	4	2	1	2	3	1	2	0	2	0	2	0	1	1	1
21	1	1	3	1	1	1	4	2	9	1	7	1	1	2	3	1	2	0	2	0	2	0	1	1	1
22	1	1	3	1	1	1	2	2	2	1	5	2	1	2	3	1	2	0	2	0	2	0	1	1	1
24a	1	1	3	1	1	1	6	2	2	2	2	1	1	2	3	1	2	0	2	0	2	1	2	9	1
24b	1	2	3	1	1	1	6	2	2	2	2	1	1	2	3	1	2	0	2	0	2	1	2	9	1
25	1	2	2	1	1	1	4	2	2	2	2	1	1	2	3	1	2	0	2	0	2	1	2	9	1
26	1	1	3	9	1	1	6	2	2	2	4	1	1	2	3	1	2	0	2	0	2	1	2	1	9
27	1	1	3	1	1	1	4	2	2	1	7	3	1	2	3	1	2	0	2	0	2	0	1	1	1
28	1	1	3	1	1	1	4	2	2	3	7	3	1	2	3	1	2	0	2	0	2	0	1	1	1
29	1	1	2	1	1	1	1	2	2	1	7	3	1	2	3	1	2	0	2	0	2	0	1	1	1
30	1	1	2	1	1	2	6	2	2	2	2	1	1	2	3	1	2	0	2	0	2	1	2	1	9
31	2	1	2	1	1	1	6	2	2	2	3	1	1	2	3	1	2	0	2	0	2	1	2	1	9
32	1	2	2	1	1	1	4	2	2	2	2	1	1	2	3	1	2	0	2	0	1	1	2	9	1
33	1	1	2	1	1	1	4	2	2	1	7	3	1	2	3	1	2	0	2	0	2	0	1	1	1
Q. cerris	1	1	1	1	1	2	6	1	1	4	2	3	1	4	3	9	2	0	3	0	2	1	1	1	1
Q. ilex	1	1	2	1	1	1	6	2	3	2	5	1	1	2	3	1	3	0	1	0	2	1	2	1	1
Q. suber	1	1	1	1	1	2	6	1	1	4	2	4	1	4	3	9	2	0	3	0	2	1	1	1	1

References

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

Bordács, S., Popescu, F., Slade, D., Csaikl, U.M., Lesur, I., Borovics, A., Kézdy, P., König, A.O., Gömöry, D., Brewer, S., Burg, K., Petit, R.J., 2002. Chloroplast DNA variation of white oaks in northern Balkans and in the Carpathian Basin. For. Ecol. Manage. 156, 197–209.

Brewer, S., Cheddadi, R., de Beaulieu, J.-L., Reille, M., and Data Contributors, 2002. The spread of deciduous *Quercus* throughout Europe since the last glacial period. For. Ecol. Manage. 156, 27–48.

Burban, C., Petit, R.J., Carcreff, E., Jactel, H., 1999. Rangewide variation of the maritime pine bast scale *Matsucoccus feytaudi* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. Mol. Ecol. 8, 1593–1602.

Bussoti, F., Grossoni, P., 1997. European and Mediterranean oaks (*Quercus* L.; Fagaceae): SEM characterization of the micromorphology of the abaxial leaf surface. Bot. J. Linn. Soc. 124, 183–199.

Cottrell, J.E., Munro, R.C., Tabbener, H.E., Gillies, A.C.M., Forrest, G.I., Deans, J.D., Lowe, A.J., 2002. Distribution of chloroplast variation in British oaks (*Quercus robur* and *Q. petraea*): the influence of postglacial recolonization and human management. For. Ecol. Manage. 156, 181–195.

Csaikl, U.M., Burg, K., Fineschi, S., König, A.O., Mátyás, G., Petit, R.J., 2001. Chloroplast DNA variation of white oaks in the Alpine region. For. Ecol. Manage. 156, 131–145.

Csaikl, U.M., Glaz, I., Baliuckas, V., Petit, R.J., Jensen, J.S., 2002. Chloroplast DNA variation of white oaks in the Baltic countries and Poland. For. Ecol. Manage. 156, 211–222.

de Jong, H., 1998. In search of historical biogeographic patterns in the western Mediterranean terrestrial fauna. Biol. J. Linn. Soc. 65, 99–164.

Demesure, B., Sodzi, N., Petit, R.J., 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. 4, 129–131.

Dumolin, S., Demesure, B., Petit, R.J., 1995. Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. Theoret. Appl. Genet. 91, 1253–1256.

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

Dumolin-Lapègue, S., Pemonge, M.-H., Petit, R.J., 1998. Association between chloroplast and mitochondrial lineages in oaks. Mol. Biol. Evol. 15, 1321–1331.

Dumolin-Lapègue, S., Kremer, A., Petit, R.J., 1999a. Are chloroplast and mitochondrial DNA variation species-independent in oaks? Evolution 53, 1406–1413.

Dumolin-Lapègue, S., Pemonge, M.-H., Gielly, L., Taberlet, P., Petit, R.J., 1999b. Amplification of DNA from ancient and modern oak wood. Mol. Ecol. 8, 2137–2140.

Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package), Version 3.5c. Department of Genetics, University of Washington, Seattle, WA.

Ferris, C., Oliver, R.P., Davy, A.J., Hewitt, G.M., 1993. Native oak chloroplasts reveal an ancient divide across Europe. Mol. Ecol. 2, 337–344.

Ferris, C., Oliver, R.P., Davy, A.J., Hewitt, G.M., 1995. Using chloroplast DNA to trace postglacial migration routes of oaks into Britain. Mol. Ecol. 4, 731–738.

Ferris, C., King, R.A., Vainola, R., Hewitt, G.M., 1998. Chloroplast DNA recognises three refugial sources of European oaks and shows independent eastern and western immigrations to Finland. Heredity 80, 584–593.

Fineschi, S., Taurchini, D., Grossoni, P., Petit, R.J., Vendramin, G.G., 2002. Chloroplast DNA variation of white oaks in Italy. For. Ecol. Manage. 156, 103–114.

Huntley, B., Birks, H.J.B., 1983. An Atlas of Past and Present Pollen Maps for Europe: 0–13,000 Years Ago. Cambridge University Press, Cambridge, UK.

Jensen, J.S., Gillies, A., Csaikl, U.M., Munro, R., Madsen, S.F., Roulund, H., Lowe, A., 2002. Chloroplast DNA variation within the Nordic countries. For. Ecol. Manage. 156, 167–180.

Johnk, N., Siegismund, H.R., 1997. Population structure and postglacial recolonization routes of *Quercus robur* and *Q. petraea* in Denmark, based on chloroplast DNA analysis. Scand. J. For. Res. 12, 130–137.

König, A.O., Ziegenhagen, B., van Dam, B.C., Csaikl, U.M., Coart, E., Degen, B., Burg, K., de Vries, S.M.G., Petit, R.J., 2002. Chloroplast DNA variation of oaks in western Central Europe and the genetic consequences of human influences. For. Ecol. Manage. 156, 147–166.

Kremer, A., 2001. Introduction. For. Ecol. Manage. 156, 1–3.

Kremer, A., Petit, R.J., Zanetto, A., Fougère, V., Ducousso, A., Wagner, D., Chauvin, C., 1991. Nuclear and organelle diversity in *Quercus robur* and *Q. petraea*. In: Muller-Starck, G., Ziehe, M. (Eds.), Genetic Variation in European Populations of Forest Trees. Sauerländer's Verlag, Frankfurt am Main.

Manos, P.S., Doyle, J.J., Nixon, K.C., 1999. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). Mol. Phyl. Evol. 12, 333–349.

Mátyás, G., 1999. Rekonstruktion der nacheiszeitlichen Einwanderung der Eichen in der schweiz anhand ihrer chloroplasten-DNA. Thesis. Diss. Nr. 13386. ETH Zürich, Switzerland.

Olalde, M., Herrán, A., Espinel, S., Goicoechea, P.G., 2002. White oaks phylogeography in the Iberian peninsula. For. Ecol. Manage. 156, 89–102.

Petit, R.J., Kremer, A., Wagner, D.B., 1993a. Geographic structure of chloroplast DNA polymorphisms in European oaks. Theoret. Appl. Genet. 87, 122–128.

Petit, R.J., Kremer, A., Wagner, D.B., 1993b. Finite island model for organelle and nuclear genes in plants. Heredity 71, 630–641.

Petit, R.J., Pineau, E., Demesure, B., Bacilieri, R., Ducousso, A., Kremer, A., 1997. Chloroplast DNA footprints of postglacial recolonization by oaks. Proc. Natl. Acad. Sci. USA 94, 9996–10001.

Petit, R.J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., Cottrell, J., Csaikl, U.M., van Dam, B.C., Deans, J.D., Espinel, S., Fineschi, S., Finkeldey, R., Glaz, I., Goicoechea, P.G., Jensen, J.S., König, A.O., Lowe, A.J., Madsen, S.F., Mátyás, G., Munro, R.C., Popescu, F., Slade, D., Tabbener, H., de Vries, S.M.G., Ziegenhagen, B., de Beaulieu, J.-L., Kremer, A., 2002a. Identification of refugia and postglacial colonization routes of European white oaks based on chloroplast DNA and fossil pollen evidence. For. Ecol. Manage. 156, 49–74.

Petit, R.J., Latouche-Hallé, C., Pemonge, M.-H., Kremer, A., 2002b. Chloroplast DNA variation of oaks in France and the influence of forest fragmentation on genetic diversity. For. Ecol. Manage. 156, 115–129.

Pons, O., Petit, R.J., 1996. Measuring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144, 1237–1245.

Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17, 1105–1109.

Whittemore, A.T., Schaal, B.A., 1991. Interspecific gene flow in oaks. Proc. Natl. Acad. Sci. USA 88, 2540–2544.