Theor Appl Genet (1993) 87:122–128

Geographic structure of chloroplast DNA polymorphisms in European oaks

R. J. Petit [1], **A. Kremer** [1], **D. B. Wagner** [2]

[1] INRA, Laboratoire de Génétique et Amélioration des Arbres Forestiers, BP 45, 33611 Gazinet Cedex, France
[2] Department of Forestry, University of Kentucky, Lexington, Kentucky 40546-0073, USA

Received: 25 October 1992 / Accepted: 11 February 1993

Abstract. Chloroplast DNA polymorphisms have been detected by the conventional Southern-blotting hybridization method in four species of European oaks (*Quercus petraea, Q. robur, Q. pubescens* and *Q. pyrenaica*). Three polymorphisms, shared by at least three of these species, can be scored directly in ethidium bromide-stained gels and were used in a broad survey of the level of differentiation of the oak species and of their pattern of genetic structure in western Europe. The highly significant geographic variation and the high genetic differentiation ($Gst = 0.895$, $s_{Gst} = 0.025$) indicate a low level of cytoplasmic gene flow. We conclude that cytoplasmic genomes are well suited for the reconstruction of past migrational routes of such a complex of species.

Key words: *Quercus petraea – Q. robur – Q. pubescens* – Diversity – Bootstrap – Autocorrelation

Introduction

The description of the geographic structure of forest trees in temperate zones is of interest for several reasons. First, in these zones there is often a limited number of species that are common and which are characterized by a wide geographical range. By focusing on these taxa much may be learned about the events which shaped their genetic architecture on a continental scale. Second, patterns of recolonization after the last glaciation have been investigated successfully by methods such as the study of fossil pollen. Since we are dealing with organisms having a long mean-generation time, events dating from the last glaciation are relatively recent in the biological histories of these species and may still be traced by studying their genomes. Third, a knowledge of the genetic structure of forest tree species, especially its geographic component, is required for the management of tree improvement and gene conservation programs.

Unfortunately, numerous studies using nuclear markers have shown that, despite a high genetic diversity at the species level, only a small fraction of the total diversity in forest trees accounts for population differentiation (Hamrick and Godt 1990). The use of genes present in cytoplasmic organelles (mitochondria or chloroplasts) may be appropriate to overcome this problem, because they are usually maternally inherited in angiosperm species and hence transmitted only by seeds.

We chose four European white oaks (*Quercus* section *Lepidobalanus* Endl., *Fagaceae*) as model species to test the utility of organellar polymorphisms for analyses of population differentiation. They are diploid, wind-pollinated species which occupy large continuous ranges. Previous studies using isozymes markers have revealed little geographic differentiation for the loci analysed (Kremer et al. 1991; Müller-Starck et al. 1993). Here we identify cpDNA polymorphisms in these species and study the geographic structure of cpDNA diversity using a hierarchical decomposition of the total diversity (Nei 1977) and spatial auto-correlation analyses (Sokal and Oden 1978). A preliminary report of our studies, focusing on nuclear and organelle variation between *Q. petraea* and *Q. robur*, and which incorporated less than half of the populations studied here, has been published elsewhere (Kremer et al. 1991).

Materials and methods

DNA preparation

Total cellular DNA is extracted from young elongating leaves of 3–4 week-old seedlings grown in a greenhouse under optimum

Communicated by P. M. A. Tigerstedt
Correspondence to: R. J. Petit

conditions. The extraction protocol has been adapted from Doyle and Doyle (1990): we use dichloromethane in place of chloroform and octanol and two extractions instead of one. DNA is precipitated first in isopropanol for half an hour at −20 °C. It is then washed in ethanol overnight, dried, dissolved in water and reprecipitated in ethanol and washed several times with 76% ethanol/10 m*M* ammonium acetate.

Identification of polymorphisms

A total of six individuals belonging to three species of oaks, *Q. robur*, *Q. petraea* and *Q. pubescens*, were used to identify polymorphisms. They were selected from populations scattered over Europe, including *Q. petraea* from Denmark, Wales and Armenia, *Q. robur* from northern and north-eastern France, and *Q. pubescens* from central France. We used in mixture two large *Petunia hybrida* chloroplast DNA probes (Palmer et al. 1983) in Southern (1975)-blot experiments: P3 (21.0 kilobases) located in the large single-copy region and P4 (19.0 kb) at one junction of the inverted repeat and the small single-copy region [see Sytsma and Schaal (1985) for the map positions of the probe fragments]. A total of 33 different restriction endonucleases were assayed, including ten with 4-base recognition sites, one with a 4.5-base site (*Nci*I recognizes a pentanucleotide sequence but the third position can be either a Guanine or a Cytosine), and 22 with 6-base sites. The estimate of the nucleotidic diversity, H, and its variance were calculated as in Engels (1981, formulas 11 and 21).

Rapid screening of selected polymorphisms

Total cellular DNA was digested overnight with 10 units of a single restriction endonuclease [*Hin*dIII, *Cfo*I or *Ava*I, (Bethesda Research Laboratories, Inc.)]. Restriction fragments were separated by electrophoresis in 0.9% agarose and 100 m*M* Tris acetate/12.5 m*M* sodium acetate/1 m*M* EDTA, pH 8.1, at 2 V/cm for 36 h. The gel was stained with ethidium bromide and photographed under UV light (at 254 nm) with Kodak TMY 4″ × 5″ ISO 400 film using an MP4 Polaroid camera. This rapid method was possible because (1) cpDNA is present in many copies in leaves, (2) the nuclear genome is of relatively small size in oaks (Ohri and Ahuja 1990), (3) these restriction endonucleases do not cut cpDNA and nuclear DNA identically, due in part to differential degrees of methylation (Palmer 1986), and (4) the cpDNA polymorphisms involved relatively high-molecular-weight fragments.

Population samples

There were 62 populations of *W. petraea* (Matt.) Liebl, 25 of *Q. robur* L. and 14 of *Q. pubescens* Willd.. A few individuals from seven populations of *Q. pyrenaica* Willd. were also included. In the majority of cases, populations correspond to forest management 'regeneration units' and cover 15–20 hectares. Such populations were selected on the basis of taxonomic purity (as inferred from leaf and fruit morphology), but limited mixture cannot be ruled out. Acorns were taken at random from large bulked seed collections (50 collection points, ensuring that many seed trees will be represented in the bulk). In other cases, including most *Q. pubescens* and *Q. pyrenaica* populations, open-pollinated progenies were collected instead from a few fertile trees. Each adult tree could then be checked for taxonomic purity.

Hierarchical analysis

Diversity analysis followed the procedures recommended by Nei (1987) with corrections for small sample sizes. The individual within-population diversities were calculated using the formula 8.5 of Nei (1987). The average intrapopulation diversity h_S and the interpopulation diversity h_T are estimated using formulas 7.39 and 7.40 of Nei (1987) without the term for heterozygosity. The diversity h_G for an intermediate level (constituted by species in our study) is the average of the gene diversities of the species, made with weights proportional to the number of populations within species (Chakraborty and Leimar 1987). The total diversity can then be apportioned into its different components: proportion of this total diversity that resides within populations, h_S/h_T, among populations within species, $G_{SG}=(h_G-h_S)/h_T$, among species within a complex of species, $G_{GT}=(h_T-h_G)/h_T$, and among populations within a complex, $G_{ST}=G_{SG}+G_{GT}=(h_T-h_S)/h_T$.

Since all polymorphisms are supposed to be completely linked on the chloroplast genome (Chiu and Sears 1985), the study was done at the multi-restriction site level, i.e., each cytotype was considered as one allele. Since we never found within-individual diversity, i.e. heteroplasmy, the within-individual level of diversity was not incorporated into the analyses. The few populations of *Q. pyrenaica* were not included.

Resampling procedures

We used the bootstrap method (Efron 1979) in order to estimate the standard errors of the principal measures of diversity and differentiation described above. One-thousand artificial data sets (called bootstrap samples) were obtained from the original data set after sampling with replacement both the populations and the individuals. The parameters of diversity were then calculated for each of these bootstrap samples and the empirical standard deviations of the 1000 estimates were used as standard errors of these parameters.

Auto-correlation analysis

Data from all species were pooled since the among-species component of the genetic differentiation was low (see below). Longitudes and latitudes (expressed in degrees) were transformed into "extended Lambert 2 coordinates", which enabled the computation of the distance in kilometers between the two members of any pair of populations. All possible pairs of populations were then assigned to distance classes, with class intervals of 20 kilometers. Moran's (1950) index and Standard Normal Deviates (SNDs) were computed and plotted versus distance in kilometers on correlograms (Sokal and Oden 1978) to summarize the spatial distribution of each of the three common cytotypes. When the number of pairs in a distance class is large enough, a significant spatial pattern is indicated at the 5% probability level when $|\mathrm{SND}| \geq 1.96$. However, the validity of the tests of excess or lack of "like-pairs" depends on the expected number of like- (or unlike-) pairs within each class. As a consequence, we selected only the classes where the expected number of like- (or unlike-) pairs is higher than 1 (Wagner et al. 1991). Moreover, since the interpretation of the significance of individual SNDs in a correlogram is complicated by the fact that the individual values are correlated, we first tested the overall significance of the correlogram using Šidák's (1967) test. Following the conservative approach of Wagner et al. (1991), a single test was used for the three correlograms.

Results

Polymorphisms in the oak chloroplast genome

A total of 32 endonucleases yielded reliable patterns, i.e., there were no indications of undigested DNA. Twenty-six of these produced identical restriction patterns for all six individuals. The six remaining endonucleases which

yielded polymorphic patterns were: *Hin*dIII, *Ava*I, *Dde*I, *Nci*I, *Hpa*II, and *Cfo*I. There were two different patterns for five of these and three patterns for *Dde*I. The *Hin*dIII polymorphism was also detected (Fig. 1) with a probe P6, adjacent to P3 and which is also in the large single-copy region (Sytsma and Schaal 1985). The *Ava*I polymorphism was detected with P4, whereas the *Cfo*I polymorphism was revealed with P3 in additional experiments where the probes were used separately.

The *Ava*I, *Cfo*I and *Hin*dIII polymorphisms could be detected directly in ethidium bromide-stained agarose gels (Fig. 2). For *Hin*dIII, however, only the larger fragment of the polymorphism can be seen.

All three of these polymorphisms can be interpreted as single restriction site gains or losses since in all cases a large fragment is alternative to two smaller fragments of complementary sizes. Since large deletions or rearrangements would have been detected by several endonucleases, *Dde*I, *Hpa*II and *Nci*I polymorphisms (which represent separate mutational events, otherwise they would not give independent information) probably also result from point mutations or very small deletions/insertions. If each polymorphism is due to a point mutation, and accounting for the variable lengths of the endonuclease recognition sites, the expected nucleotidic diversity (Engels 1981) is: $H=0.0006$, $s_H=0.0002$.

The six individuals co-varied with *Cfo*I and *Ava*I, i.e., for a given individual, the restriction pattern from *Ava*I was always associated with a specific pattern from *Cfo*I. Considering all polymorphisms simultaneously, there were a total of five different chloroplast genotypes (Table 1). However, a sixth genotype found in the population survey was characterized by a third *Ava*I-*Cfo*I combination. In order to root the phylogenetic tree, an individual of the American white oak species *Q. alba* L. was characterized for the three endonucleases *Ava*I/*Cfo*I/*Hin*dIII. The observed multi-restriction site genotype of *Q. alba* (*Ava*I: 12.3 kb, *Cfo*I: 8.8 kb, *Hin*dIII: 5.4 kb) was considered ancestral (*cf.* Table 2). A cladogram (Fig. 3), which is free from homoplasy, was constructed by hand.

Hierarchical analysis

A total of 834 individuals from 107 populations were analysed. Most individuals were studied with *Hin*dIII, *Ava*I and/or *Cfo*I. With all three enzymes, only four of the eight possible multi-restriction site combinations were detected. The strong association found among sites led us in the remainder of the analyses to consider the cytotypes as alleles at a single haploid locus. Individuals analysed with two enzymes only (*Ava*I or *Cfo*I and *Hin*dIII) were attributed to one of the three cytotypes: 1:1:1, 1:1:2 and 2:2:1 (ranked in the following order: *Ava*I:*Cfo*I:*Hin*dIII), considering the scarcity of the 1:2:1 cytotype (present in only two populations) and the ab-

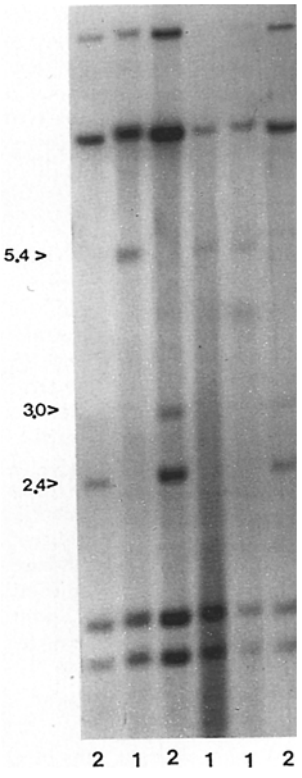

Fig. 1. Polymorphism detected by hybridization with the chloroplast DNA probe P6. Samples have been digested with endonuclease *Hin*dIII

Table 1. Genotypes of the six individuals studied at each of the six endonucleases which yielded polymorphic patterns

Restriction sites	Individuals					
	1	2	3	4	5	6
*Hin*dIII	2	1	1	1	1	2
*Ava*I	1	1	1	2	1	1
*Dde*I	2	1	1	3	3	2
*Nci*I	1	1	1	1	1	2
*Hpa*II	2	2	2	2	1	2
*Cfo*I	1	1	1	2	1	1

Table 2. Molecular weights (in kilobases) of the polymorphic fragments and inferred relationships

Enzyme	Fragment size, kb		
	Primitive	Intermediate	Derived
*Hin*dIII	5.4 (1)		3.0+2.4 (2)
*Ava*I	12.3 (1)		7.3+5.0 (2)
*Dde*I	2.5 (3)	3.8/2.5 (2)	3.8/2.8 (1)
*Nci*I	2.2 (1)		1.8 (2)
*Hpa*II	3.4 (2)		3.6 (1)
*Cfo*I	8.8 (1)		4.5+4.3 (2)

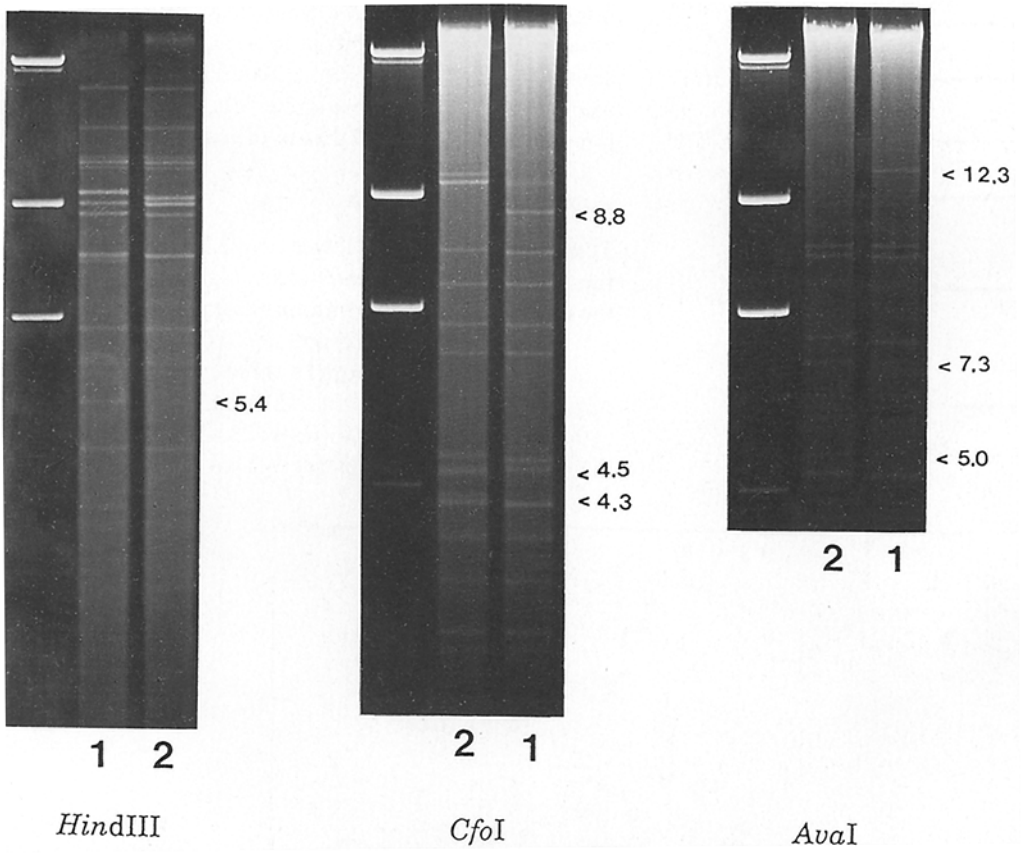

Fig. 2. Polymorphisms detected directly by ethidium bromide staining of the gels, after digestion of total cellular DNA by three different endonucleases

Table 3. Sample sizes for the hierarchical analysis of diversity, and diversity measures at the species, and complex of species, level

Item	Frequency			Total
	Q. pubescens	*Q. robur*	*Q. petraea*	
Cytotype				
111	0.284	0.073	0.273	0.226
221 + 121	0.303	0.373	0.442	0.409
112	0.413	0.554	0.285	0.365
Individuals	81	171	470	722
Populations	11	22	58	91
Mean	7.36	7.77	8.10	7.93
h_S (s_{h_S})	0.157 (0.074)	0.041 (0.026)	0.063 (0.020)	h_S = 0.069 (0.017)
h_T (s_{h_T})	0.678 (0.067)	0.544 (0.065)	0.666 (0.017)	h_G = 0.638
				h_T = 0.660 (0.015)
G_{ST} ($s_{G_{ST}}$)	0.769 (0.119)	0.925 (0.038)	0.905 (0.030)	G_{GS} = 0.863
				G_{ST} = 0.895 (0.025)

sence of the four other three-sites cytotypes. The rare 1:2:1 cytotype was pooled with the 2:2:1 genotype in the hierarchical decomposition of total diversity. Eventually, a total of 722 individuals could be typed. The genetic composition of all populations studied are illustrated in Fig. 4, with the exception of two populations in Turkey and Armenia. Considering only the populations where four or more individuals could be typed (91 populations), there were $16/91 \cong 18\%$ polymorphic populations.

The results of the hierarchical analyses (Table 3) confirm this low within-population diversity since $h_S/h_T = 10.5\%$ of the total diversity resides within popula-

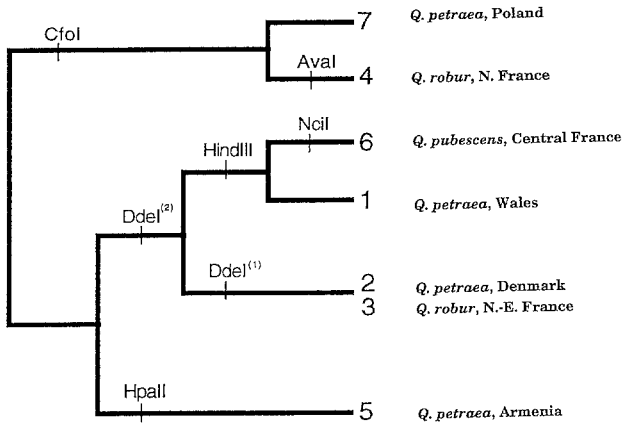

Fig. 3. Phylogenetic tree of seven individuals belonging to three different species of European white oaks

tions, $G_{SG}=86.3\%$ of the total diversity is distributed among populations within species, and $G_{GT}=3.2\%$ among species ($G_{ST}=89.5\%$). Sampling-error estimates using the bootstrap were quite low, except when few populations were sampled as in the case of *Q. pubescens*.

Auto-correlation analysis

The Šiďak's p value was lower than 0.003, indicating that the correlograms are statistically highly significant (i.e. the cytotypes are non-randomly distributed).

There were 34 significant ($P<0.05$) SNDs out of a total of 150 computed for the three cytotypes (25-excess of like-pairs and 9-excess of unlike-pairs). Within the first 300 kilometers, out of 45 tests, 23 were significant (all involving excess of like-pairs). Within the first 40 kilome-

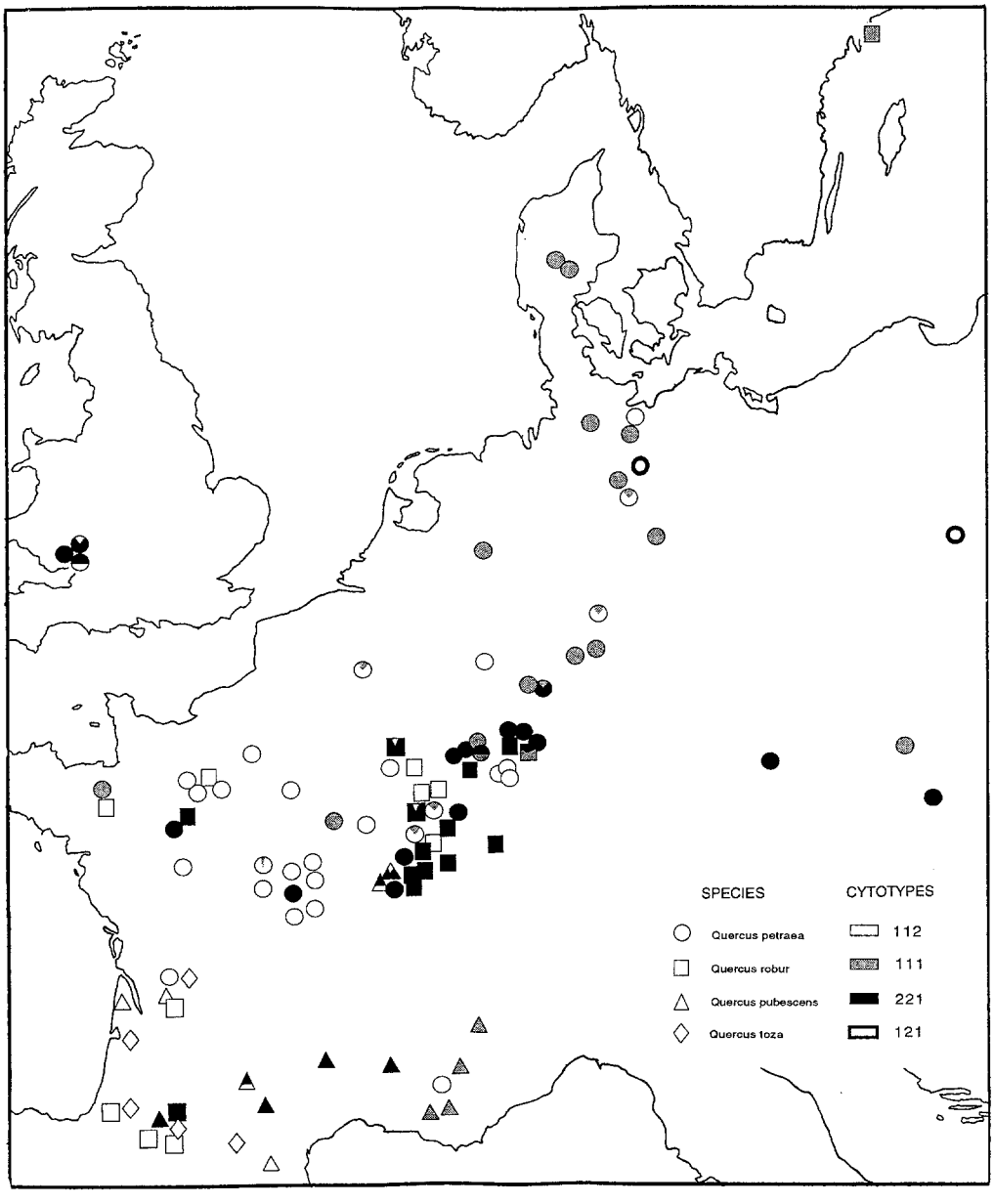

Fig. 4. Map of the cpDNA genotypes in western and central Europe. Each symbol represents one sampling point with an average of 7–8 individuals sampled. *Shapes* of symbols indicate the taxonomic status of the population (inferred from the morphology of adult trees) and *colors* indicate the cytotype. The *sectors* indicate the proportion of each genotype when the population is polymorphic

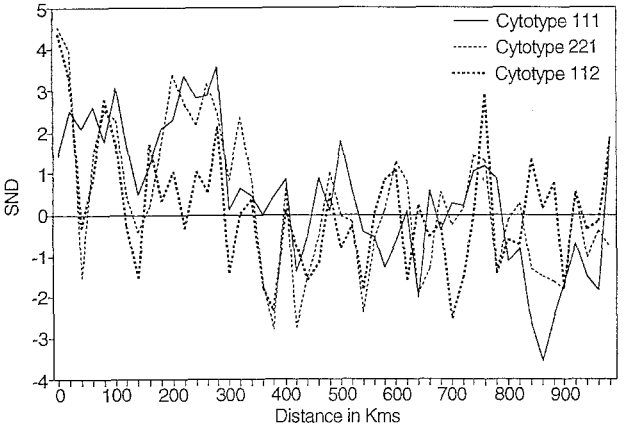

Fig. 5. Correlogram of SNDs as a function of distance for the three common cpDNA genotypes: cytotypes (——) 111, (– – –) 221, and (· · · ·) 112. Values above +1.96 or below −1.96 indicate significant excess or deficit of frequency of "like-pairs" in a given distance class with a 95% confidence level

ters, highly significant auto-correlations (SNDs > 4) were detected for cytotypes 221 (+121) and 112. For pairs of populations located more than 300 kilometers away, non-significant and significant negative associations were predominant. Clearly, there is a very strong pattern of geographic structure.

Discussion

Based on comparisons made within taxonomic families (Smith and Doyle 1986, Wilson et al. 1990) it has already been suggested that woody species have a low cpDNA substitution rate as compared with other land plants. Though only a limited number of polymorphisms were scored, our study also demonstrates the advantages of a large-scale population survey of cytoplasmic diversity. First, when many more individuals than polymorphisms are studied, the complete set of multi-site genotypes may be found, allowing the phylogenetic reconstruction of all lineages with accuracy. For instance, in our study, the sampling of additional populations in Poland and northern Germany led to the detection of the third *Ava*I: *Cfo*I combination. The absence of some multi-site genotypes, (i.e., only four three-sites combinations could be detected), may at first seem surprising. However, when n sites are considered, which correspond to n independent mutations, the presence of more than n+1 genotypes is unexpected, since it would involve recombination or recurrent mutation, or more generally convergence at the molecular level. If cpDNA does recombine, even rarely, new recombinant cytotypes may become fixed in some populations. Our survey, which included many populations, was particularly suited to the detection of such rare events and the results are consistent with the absence or extreme rarity of cpDNA recombination.

The high number of populations allowed quite precise estimates of G_{ST}. Differentiation among regions was demonstrated by auto-correlation analyses, with cytotypes predominating over entire parts of the range (i.e., a few hundred kilometers) and nearly fixed in smaller patches (a few tens of kilometers wide). Overall, a cline was apparent, with cytotype 1:1:1 replacing both other cytotypes eastward. Although our sampling proved adequate for the estimation of differentiation, the description of the geographic structure of oak cpDNA polymorphisms will gain from the analysis of additional polymorphisms, since each new polymorphism is expected to subdivide any one of the four mapped cytotypes into two new categories.

The high level of differentiation was difficult to predict because acorns are transported efficiently (up to a few kilometers) by jays (*Garrulus glandarius* L.), their principal biological disperser (Bossema 1979). However, the colonizing abilities of these oaks, which are important (Huntley and Birks 1983), do not necessarily imply a high level of migration once these species have established dense forests. Hence, a scenario that would involve rapid postglacial spread from glacial refuges, followed by a "freezing" of the cytoplasmic genetic structure set during the colonization period, can be imagined. The study of tree species having wide geographic distributions and reduced seed flow may increase considerably our knowledge of past colonization routes and centres of origins of these economically and ecologically important species and, therefore, help improve the conservation of these natural resources.

Finally, a striking result is the complete absence of specificity of the cpDNA genotypes. This is clear in the phylogenetic tree (Fig. 3), and was further apparent in the hierarchical genetic diversity analysis, where differences among species accounted for only a minor proportion of the total diversity.

References

Bossema I (1979) Jays and oaks: an eco-ethological study of a symbiosis. Behaviour 70:1–117

Chakraborty R, Leimar O (1987) Genetic variation within a subdivided population. In: Ryman N, Utter F (eds) Population genetics and fishery management. Washington Sea Grant program, Seattle and London, pp 89–120

Chiu W-L, Sears BB (1985) Recombination between chloroplast DNAs does not occur in sexual crosses of *Oenothera*. Mol Gen Genet 198:525–528

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

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Stat 7:1–26

Engels WR (1981) Estimating genetic divergence and genetic variability with restriction endonucleases. Proc Natl Acad Sci USA 78:6329–6333

Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding, and genetic resources. Sinauer associates Inc., Sunderland, Massachusetts, pp 43–63

Huntley B, Birks HJB (1983) An atlas of past and present pollen maps for Europe, 0-13,000 years ago. Cambridge University Press, Cambridge, UK

Kremer A, Petit RJ, Zanetto A, Fougère V, Ducousso A, Wagner D (1991) Nuclear and organelle gene diversity in *Q. robur* and *Q. petraea*. In: Ziehe M, Müller-Starck G (eds) Genetic variation of forest tree populations in Europe. Sauerländer's Verlag, Frankfurt am Main, pp 141–172

Moran PAP (1950) Notes on continuous stochastic phenomena. Biometrika 37:17–23

Müller-Starck G, Herzog S, Hattemer HH (1993) Intra- and interpopulational genetic variation in juvenile populations of *Quercus robur* L. and *Quercus petraea* Liebl. Ann Sci For 50 (Suppl 1):233s–244s

Nei M (1977) F-statistics and analysis of gene diversity in subdivided populations. Ann Hum Genet 41:225–233

Nei M (1987) Molecular evolutionary genetics. Colombia University Press, New York

Ohri D, Ahuja MR (1990) Giemsa *C*-banded karyotype in *Quercus* L. (oak). Silvae Genet 39:216–219

Palmer JD (1986) Isolation and structural analysis of chloroplast DNA. Methods Enzymol 118:167–186

Palmer JD, Shields CR, Cohen DB, Orton TJ (1983) Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. Theor Appl Genet 65:181–189

Šidák Z (1967) Rectangular confidence regions for the means of multivariate normal distributions. J Am Stat Assoc 62:626–633

Smith JF, Doyle JJ (1986) Chloroplast DNA variation and evolution in the Juglandaceae. Am J Bot 73:730

Sokal RR, Oden NL (1978) Spatial auto-correlation in biology. 1. Methodology. Biol Jour Linnean Soc 10:199–228

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Sytsma KJ, Schaal BA (1985) Phylogenetics of the *Lisianthus skinneri* (Gentianaceae) species complex in Panama utilizing DNA restriction fragment analysis. Evolution 39:594–608

Wagner DB, Sun Z-X, Govindaraju DR, Dancik BP (1991) Spatial patterns of chloroplast DNA and cone morphology variation within populations of a *Pinus banksiana-Pinus contorta* sympatric region. Am Nat 138:156–170

Wilson MA, Gaut B, Clegg MT (1990) Chloroplast DNA evolves slowly in the palm family. Mol Biol Evol 7:303–314