Copyright © 2001 by the Genetics Society of America

# Diverging Trends Between Heterozygosity and Allelic Richness During Postglacial Colonization in the European Beech

**B. Comps,* D. Gömöry,† J. Letouzey,* B. Thiébaut‡ and R. J. Petit§**

**Laboratoire d'Ecologie Génétique, Université Bordeaux I, F-33405 Talence, France, †Technical University in Zvolen, Faculty of Forestry, SK-96053 Zvolen, Slovakia, ‡Institut de Botanique, F-34090 Montpellier, France and §Laboratoire de Génétique et Amélioration des Arbres Forestiers, F-33611 Gazinet Cedex, France*

Manuscript received July 7, 2000
Accepted for publication October 13, 2000

ABSTRACT

Variation at 12 polymorphic isozyme loci was studied in the European beech on the basis of an extensive sample of 389 populations distributed throughout the species range. Special emphasis was given to the analysis of the pattern of geographic variation on the basis of two contrasting measures of genetic diversity, gene diversity ($H$) and allelic richness, and to their relationship. Measures of allelic richness were corrected for variation in sample size by using the rarefaction method. As expected, maximum allelic richness was found in the southeastern part of the range (southern Italy and the Balkans), where beech was confined during the last ice age. Surprisingly, $H$ was lower in refugia than in recently colonized regions, resulting in a negative correlation between the two diversity measures. The decrease of allelic richness and the simultaneous increase of $H$ during postglacial recolonization was attributed to several processes that differentially affect the two diversity parameters, such as bottlenecks due to long-distance founding events, selection during population establishment, and increased gene flow at low population densities.

IN temperate regions, areas that have remained occupied through long periods, during the last ice age up to the present (*i.e.*, refugia), are expected to harbor higher levels of genetic diversity compared to those that have been recolonized after the last ice age (Hewitt 1996). However, there is only limited evidence to support this hypothesis, partly because it is often difficult to identify precisely these refugia with independent (nongenetic) data.

The theory of bottlenecks or transitory reductions in the effective population size predicts a strong decrease of allelic richness and a more limited decrease of $H$ at neutral loci, since rare alleles will be more readily affected by drift than the more frequent ones (Nei *et al.* 1975). But this theory may not be strictly applicable to the founding events that have taken place during postglacial colonization. In the context of a metapopulation where the component populations are subjected to extinction and recolonization, Wade and McCauley (1988) suggested that a decrease in $H$ and an increase in levels of differentiation was to be expected in newly founded populations, but Slatkin (1977) insisted that colonization could constitute a form of gene flow, when the number of founders is large enough. While population bottlenecks may initially reduce levels of genetic diversity (Cornuet and Luikart 1996; Nei *et al.* 1975), admixture effects, which could take place when immigrating populations meet, may result in a secondary increase of diversity. Finally, selection during the early phases of establishment could help maintain more genetic diversity than expected using strictly neutral models.

Explicit empirical tests of historical factors that have shaped levels of diversity within and among populations remain scarce, especially in plant populations. An early exception is the work of Cwynar and MacDonald (1987). New opportunities to interpret patterns of diversity of plants (and especially trees) emerge as more becomes known of the past distributions and demographic processes during the postglacial colonization (MacDonald 1993).

The European beech, *Fagus sylvatica* L. (Fagaceae), is particularly suited for a detailed investigation of the factors that have influenced levels of diversity across its geographic range. Patterns of diversity of beech in Europe are not complicated by the existence of interfertile species, except for the presence of an allopatric related beech species in Asia minor (*F. orientalis*). In addition, a very detailed pollen record has been obtained for this anemophilous species, because it has colonized western and northern Europe relatively late during the Holocene. This has allowed the identification of the most likely refugia and colonization routes (Huntley and Birks 1983; European Pollen Database, S. Brewer, R. Cheddadi and J.-L. de Beaulieu, unpublished data), which have been confirmed by chloroplast DNA investigations (Demesure *et al.* 1996). Finally, the European beech is one of the species for which the largest surveys based on isozyme markers exist. For example, within Germany, a recent compilation reports

*Corresponding author:* R. J. Petit, INRA, Laboratoire de Génétique et d'Amélioration des Arbres Forestiers, B.P. 45, F-33611 Gazinet Cedex, France. E-mail: remy@pierroton.inra.fr

multilocus data from >40,000 beech trees distributed in 246 populations (Konnert *et al.* 2000), and the species has been used as a model to investigate viability differences among isozyme genotypes (*e.g.*, Müller-Starck 1993). The large isozyme data set that is the basis of the present analysis covers most of the range of the European beech and has been assembled during the last decade. Only a limited part of the data has been published so far (Comps *et al.* 1990, 1991, 1998; Hazler *et al.* 1997).

Many empirical investigations consider only *H*, the probability that two alleles sampled at random are different, a parameter called *gene diversity* by Nei (1973). This approach is restrictive, since allelic richness has quite different dynamics and may be more useful to identify historical processes such as bottlenecks (Luikart *et al.* 1998) and population admixture (Chakraborty *et al.* 1988). Here we use both diversity statistics to investigate the patterns of distribution of genetic variation in the European beech; the rarefaction method is employed to correct for variation in sample size in the measure of allelic richness (Hurlbert 1971; Petit *et al.* 1998).

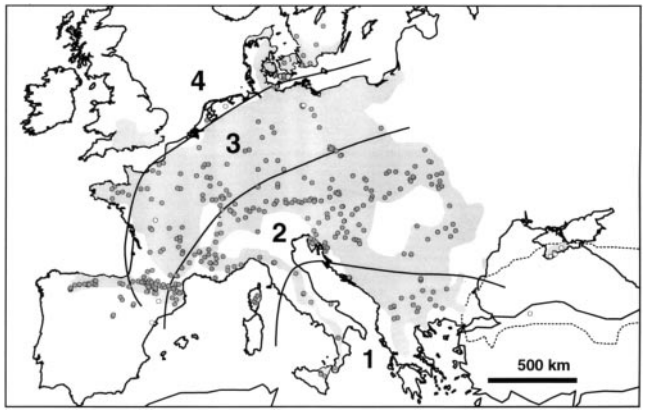



Figure 1.—Sampled beech populations. The shaded areas indicate the range of the European beech (*Fagus sylvatica*). The dotted line outlines the western distribution of *F. orientalis.* Four zones have been determined, corresponding to colonization history. Zone 1 corresponds to the range of beech during the last ice age, as inferred from fossil pollen. Zones 2 and 3 correspond to beech forests of intermediate age (older and younger than ~4000 years, respectively). Zone 4 is the area colonized most recently (<2000 years). Circles indicate the position of each sampled population (389 in total). The open circles correspond to populations that were not included in any group.

## MATERIALS AND METHODS

**The study species:** The European beech is a monoecious diploid ($2n = 24$) late-successional forest tree; forests dominated by this species cover ~17 million ha in Europe. As the beech generally grows in a temperate and rather wet climate, its populations occur at low altitudes in the north and at higher altitudes in the south of its range. It is a highly outcrossing wind-pollinated tree species characterized by a self-fertilization rate of 0–0.10 (Merzeau *et al.* 1994). Beech does not reproduce until it is 40–50 years old. The production of beech nuts is characterized by irregular mast years. The European jay (*Garrulus glandarius*) is the main long-distance seed disperser (Nilsson 1985), but the bulk of the seeds are disseminated over small distances.

**Material:** A total of 389 populations of the European beech (*F. sylvatica* L.) were sampled through Europe over the natural range of the species (Figure 1). The sampled populations are considered to be of natural origin and were growing at altitudes ranging from sea level to 1700 m. Regions not sampled included Great Britain and mainland Denmark, where it was difficult to ascertain the status of the beech forests due to the presence of numerous artificial plantations, and Greece and Bosnia. In each population beech twigs with dormant buds were sampled from 50 nonadjacent individuals (to avoid the sampling of related trees) chosen at random over a 3–4-ha area in a homogeneous environment.

**Electrophoresis analysis:** Enzymes were extracted from buds and cortical tissue of each individual and were separated on starch and acrylamide electrophoresis gels following the procedures of Thiébaut *et al.* (1982), Merzeau *et al.* (1989), and Müller-Starck and Starke (1993). Ten isozyme systems, encoding 15 loci, were analyzed: *Acp* (acid phosphatase, EC 3.1.3.2), *Got* (glutamate oxaloacetate transaminase, EC 2.6.1.1), *Idh* (isocitrate dehydrogenase, EC 1.1.1.42), *Mdh* (malate dehydrogenase, EC 1.1.1.37), *Mnr* (menadione reductase, EC 1.6.99.2), 6-*Pgd* (6-phosphogluconate dehydrogenase, EC 1.1.1.44), *Pgi* (phosphoglucose isomerase, EC 5.3.1.9), *Pgm* (phosphoglucomutase, EC 2.7.5.1), *Per* (peroxidases, EC 1.11.1.7), and *Sod* (superoxide dismutase, EC 1.15.1.1).

**Data analysis:** To identify the broad-scale trends in genetic differentiation throughout the species' range, the populations were combined into 55 groups of 5–9 populations (mean of 7), according to their geographical proximity. Six populations were not included in any group: 2 were too isolated, and 4 had strikingly unusual allelic frequencies. These 4 populations may have experienced particularly strong founding events (Spain) or may be of artificial origin (western France, where chloroplast DNA haplotypes of nonlocal origin have been detected; Demesure *et al.* 1996). The groups were further combined into four zones (excluding Crimea, due to its peripheral situation and ambiguous taxonomic position; Gömöry *et al.* 1998), on the basis of the time since beech first colonized each of these regions (Figure 1). Zone 1 consists of 5 groups originating from Bulgaria, Macedonia, and southern Italy, and corresponds to the most ancient beech forests. In these regions, beech has been detected continuously in palynological sequences spanning the last ice age (Bulgaria, Bozhilova *et al.* 1989; Greece, Tzedakis 1993; Italy, Leonardi and Menozzi 1995; Watts *et al.* 1996; see also Bennett *et al.* 1991; Berglund *et al.* 1996). Zone 2 includes 26 groups that are relatively close to the inferred glacial refugia, whereas zone 3 includes 16 more peripheral groups of populations. The limit between zones 2 and 3 was based on published (Huntley and Birks 1983) and unpublished (S. Brewer, R. Cheddadi and J.-L. de Beaulieu, personal communication) isopollen maps representing the inferred limit of distribution of beech ~4000 years ago. Zone 4 was made up of the 7 westernmost and northernmost groups of populations, at the extreme periphery of the species' range, which were founded 2000 years ago or less: western Spain, Brittany, Denmark, and Sweden (Huntley and Birks 1983; Berglund *et al.* 1996).

For each population, the observed heterozygosity ($H_O$), the gene diversity ($H$), and the heterozygote deficit ($F_{IS}$) were computed for each locus. For each group, mean within-population gene diversity ($H_S$), total gene diversity ($H_T$), and the coefficient of genetic differentiation ($F_{ST}$) were estimated along with their variances following Pons and Chaouche

(1995). Computation of allelic richness for specified sample sizes was based on the rarefaction method developed in ecology by Hurlbert (1971). Under this method, allelic richness is denoted as $A[g]$ and corresponds to the number of different alleles found when $g$ gene copies—the specified sample size—are sampled at the locus in question. If a total of $N$ ($N > g$) gene copies are examined at this locus, the expected number of different alleles in a sample of size $g$ may be obtained by the formula

$$A[g] = \sum_i \left[1 - \binom{N - N_i}{g} \Big/ \binom{N}{g}\right] = \sum_i \left[1 - \prod_{k=0}^{g-1} \frac{N - N_i - k}{N - k}\right],$$

where $N_i$ represents the number of occurrences of the $i$th allele among the $N$ sampled gene copies.

Average sample size was 47.4 individuals per population, with sample sizes of most populations exceeding 40 diploid individuals (that is, 80 gene copies). We therefore chose to compare allelic richness after rarefaction to a common sample size of $g = 80$. The mean within-population allelic richness was named $A_S[g]$ and the total allelic richness $A_T[g]$. The coefficient $A_{ST}[g] = 1 - (A_S[g] - 1)/(A_T[g] - 1)$ was used to measure the partitioning of allelic richness among populations (Petit *et al.* 1998). It largely depends on the distribution of rare alleles, notably whether they tend to be clustered in some populations (high $A_{ST}$) or are distributed more evenly so that one population is representative of the whole species (low $A_{ST}$). For statistics at the group level, we estimated allelic richness at a common sample size of 400 gene copies (the smallest group comprising 225 individuals, *i.e.*, 450 gene copies). For comparisons among zones, which varied between 1573 trees in zone 1 and 8638 in zone 2, we compared allelic richness at a specified sample size of 1000 gene copies. A simplified version of the program (CONTRIB) used to make these computations is available at www.pierroton.inra.fr/genetics/labo/Software/.

The significance of differences in genetic diversity parameters between the four geographic zones was tested using a Wilcoxon signed-ranks test in the Splus statistical package. This nonparametric test for paired comparisons was appropriate because homologous loci were used in the different populations. The test does not assume that the data are normally distributed (Luikart *et al.* 1998). To further examine the trend across loci for gene diversity in refugia *vs.* peripheral populations, we included data from four additional loci (*Got*-2, *Lap*-1, *Mdh*-1, and *Skdh*-1) that have been studied in a separate investigation of 186 distinct populations distributed mostly in eastern and central Europe (Gömöry *et al.* 1999).

Bottlenecks were inferred for each population using the program BOTTLENECK (Cornuet and Luikart 1996), which contrasts $H$ with the heterozygosity predicted on the basis of the observed number of alleles under the assumption of mutation-drift equilibrium (Ewens 1972). This distribution is obtained through simulating the coalescent process under the infinite alleles model. For comparison, the corresponding standardized deficit of heterozygosity ($dH/SD$; Cornuet and Luikart 1996) and its associated $P$ value were used.

The correlations between the geographic variables (longitude, latitude, and altitude) and the diversity parameter estimates were studied at the group level (54 groups, excluding Crimea), and at the population level (376 populations included in the 54 groups).

## RESULTS

**Overall diversity:** At each of 15 allozyme loci, diploid genotypes were obtained from an average of 18,440 trees distributed over 389 populations, resulting in the scoring of a total of 497,146 allelic products. Three loci were monomorphic in all populations (*Acp*-2, *Got*-3, and *Pgm*-2), although they are known to be polymorphic in other species of Fagus (B. Comps, unpublished data). The 12 remaining loci have 2–5 alleles per locus (total of 41; Table 1). Thirteen of the 41 alleles studied were present in all populations. Among the 28 remaining alleles, on which all measures of allelic richness were based, 8 were present in <5% of the 389 populations, and 8 in >95% of the populations.

There were on average 2.26 alleles per polymorphic locus and per population (standardized samples of 80 gene copies). At the *species* level, the equivalent number, *i.e.*, based on a sample of 80 gene copies, is 2.45 alleles per locus, so that the overall coefficient of differentiation for allelic richness $A_{ST}[80]$ is 0.13.

The mean observed heterozygosity $H_O$ (0.275) was slightly lower than the value expected for random mating populations ($H_S = 0.282$), indicating a limited heterozygote deficit ($F_{IS} = 0.025$). One of the 12 loci has a significant heterozygote excess (*Acp*-1), and 4 have a significant deficit, particularly strong in the case of locus *Mdh*-2 ($F_{IS} = 0.247$). The overall value for $F_{ST}$ is 0.059, and all 12 $F_{ST}$ estimates are highly significantly different from zero, varying from 0.031 to 0.114.

**Correlation between geographic variables and diversity parameters:** Significant correlations were found between latitude, longitude, altitude, and several diversity parameters (Table 2). Longitudinal trends were usually stronger than latitudinal ones, especially with allelic richness measures. Altitudinal trends exist, but altitude is strongly correlated with latitude (−0.82 at the group level, −0.70 at the population level). In general, correlations are higher at the group level. The strongest correlations with geography are those between longitude and allelic richness, especially for total allelic richness ($A_T[400]$; $r = 0.73$, $P < 0.001$). The positive correlation between latitude and $H_S$ is also noteworthy ($r = 0.43$ at the group level, $P < 0.001$). The negative correlation observed between the two measures of genetic variation is remarkable ($r = -0.45$ between $H_T$ and $A_T[400]$, $P < 0.001$).

**Geographic patterns for genetic diversity estimates:** All genetic diversity parameters and fixation indices were estimated for each of the 55 groups. Only the maps representing the results at the group level are included, since they provide a good overall synthesis. Figure 2A illustrates the results for the mean within-population gene diversity ($H_S$) [results for total group diversity ($H_T$) are very similar and are not shown]. Southeastern populations are characterized by below average values of $H_S$. Above average values are found in the remainder of the range. Strikingly, the pattern is the reverse for allelic richness (Figure 2B): the highest values, at the group level ($A_T[400]$), are found in Italy and in southeastern Europe. The mean within-population allelic richness

**TABLE 1**

**Diversity parameters across loci and across zones**

| | $N_p$ | $N$ | $m$ | $A$ | $A_T[1000]^a$ | $A_S[80]$ | $A_T[80]$ | $H_O$ | $H_S$ | $H_T$ | $F_{ST}{}^a$ | $F_{IS}{}^a$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Acp*-1 | | 17,544 | 45.1 | 3 | 3.00 | 2.92 | 3.00 | 0.52 | 0.50 | 0.54 | 0.06*** | −0.03*** |
| *Got*-1 | | 18,200 | 46.8 | 4 | 2.82 | 2.04 | 2.10 | 0.25 | 0.27 | 0.30 | 0.11*** | 0.06*** |
| *Idh*-1 | | 18,987 | 48.8 | 3 | 3.00 | 2.44 | 2.69 | 0.36 | 0.36 | 0.38 | 0.05*** | 0.01 |
| *Mdh*-3 | | 18,810 | 48.4 | 2 | 2.00 | 1.97 | 2.00 | 0.36 | 0.36 | 0.38 | 0.04*** | 0.01 |
| *Mdh*-2 | | 18,488 | 47.5 | 5 | 3.49 | 2.76 | 3.02 | 0.15 | 0.20 | 0.21 | 0.03*** | 0.25*** |
| *Mnr*-1 | | 19,186 | 49.3 | 5 | 4.38 | 2.32 | 2.85 | 0.12 | 0.12 | 0.13 | 0.03*** | 0.00 |
| *Per*-1 | | 18,279 | 47.0 | 2 | 2.00 | 2.00 | 2.00 | 0.39 | 0.41 | 0.43 | 0.04*** | 0.05*** |
| *Per*-2 | | 18,377 | 47.2 | 3 | 3.00 | 2.93 | 3.00 | 0.39 | 0.39 | 0.40 | 0.04*** | 0.01 |
| *6Pgd*-1 | | 17,760 | 45.7 | 4 | 3.11 | 2.15 | 2.68 | 0.15 | 0.16 | 0.17 | 0.09*** | 0.03* |
| *Pgi*-1 | | 19,094 | 49.1 | 4 | 3.10 | 1.60 | 1.97 | 0.04 | 0.04 | 0.04 | 0.03*** | 0.02 |
| *Pgm*-1 | | 18,834 | 48.4 | 4 | 2.92 | 2.04 | 2.14 | 0.33 | 0.33 | 0.36 | 0.08*** | −0.01 |
| *Sod*-1 | | 17,725 | 45.6 | 2 | 2.00 | 1.99 | 2.00 | 0.25 | 0.25 | 0.27 | 0.09*** | 0.01 |
| All loci | | 18,440 | 47.4 | 41 | 2.90 | 2.26 | 2.45 | 0.28 | 0.28 | 0.30 | 0.07 | 0.02 |
| Zone 1 | 33 | 1,573 | 47.7 | 40 | 3.10 (a) | 2.28 | 2.53 | 0.24 | 0.25 | 0.26 | 0.03 (a) | 0.06 (a) |
| Zone 2 | 180 | 8,638 | 48.0 | 40 | 2.84 (b) | 2.26 | 2.37 | 0.27 | 0.28 | 0.29 | 0.04 (a) | 0.05 (a) |
| Zone 3 | 116 | 5,539 | 47.8 | 34 | 2.51 (c) | 2.25 | 2.31 | 0.31 | 0.32 | 0.33 | 0.04 (a) | 0.03 (a) |
| Zone 4 | 47 | 2,138 | 45.5 | 31 | 2.43 (c) | 2.20 | 2.31 | 0.31 | 0.31 | 0.33 | 0.07 (b) | 0.00 (b) |

$N_p$, number of populations included in each zone; $N$, total number of trees analyzed at each locus or in each of the four zones; $m$, mean number of trees analyzed per population; $A$, number of alleles observed at the locus or within the zone; $A_T[1000]$, total allelic richness after rarefaction to 1000; $A_S[80]$ and $A_T[80]$, respectively, mean within- and total allelic richness after rarefaction to 80; $H_O$, $H_S$, and $H_T$, observed, within-population, and total heterozygosities; $F_{ST}$ and $F_{IS}$, coefficient of genetic differentiation and heterozygote deficit; *, **, ***, levels of significance based on 5, 1, and 0.1% confidence levels, respectively.

[a] Among zones, parameters ($A_T$, $F_{ST}$, $F_{IS}$) with (a), (b), and (c) are significantly different based on the Wilcoxon signed-ranks test.

has a similar pattern, except that Italian populations are below average (results not shown).

Geographic trends for fixation indices ($F_{IS}$ and $F_{ST}$) are very different from those observed for $H$ and for allelic richness. The coefficient of differentiation is above average in northern Europe (Sweden and Denmark), in Brittany (western France), northwestern Spain, and Rumania, as well as in Corsica, in southeastern France, and in the Pyrenees (Figure 2C). Except for Corsica and southeastern France, the same regions are characterized by negative $F_{IS}$ values (*i.e.*, heterozygote excess; see Figure 2D). Geographic patterns of differentiation for allelic richness ($A_{ST}$; see Figure 2E) resemble those observed for $F_{ST}$ (Figure 2C) and for allelic richness (Figure 2B), as populations from both zones 1 and 4 present higher levels of differentiation. Finally, the analysis made to infer bottlenecks differentiates southeastern European populations from the remaining ones, closely matching the results obtained for $H_S$; out of 383 tests, 71 (19%) were significant at $P < 0.05$, none of them located in zone 1 (corresponding to the oldest populations; Figure 2F).

**Evolution of genetic diversity during colonization:** Diversity parameters were computed in each of the four zones corresponding to increasingly recent beech forests (Figure 1 and Table 1). Despite the comparatively small size of zone 1 (9% of the populations), all but 1 allele (*Got*-1, allele 1) were detected in this zone. Similarly, in zone 2 (48% of the populations), 40 of the 41 alleles are present. In zone 3, 7 alleles are lacking, and 10 are lacking in zone 4. After standardization to a common sample size of 1000 gene copies, the highest value was observed in zone 1 (3.10 alleles per polymorphic locus), followed by zone 2 (2.84), and zones 3 and 4 (2.51 and 2.43, respectively). Wilcoxon signed-ranks tests across the 12 loci showed that all differences, except between zones 3 and 4, were significant, indicating a steady decrease in allelic richness during postglacial colonization. Within-population and within-group allelic richness, on the basis of a sample of 80 gene copies, follow the same trend. For $H_T$ and $H_S$, the trend is opposite, with larger values in newly colonized areas. Total diversity $H_T$ is 30% higher in newly colonized regions than in refugia (0.33 in zone 4 *vs.* 0.26 in zone 1), and within-population diversity $H_S$ is 24% larger (0.31 in zone 4 *vs.* 0.25 in zone 1). The coefficient of genetic differentiation among populations ($F_{ST}$) also increases away from refugia, especially in zone 4, where it reaches 0.07, compared to only 0.03 in zone 1. Finally, the heterozygote deficit $F_{IS}$, on the basis of the 12 loci, decreases with time since colonization, from 0.06 in zone 1 to close to 0 in zone 4.

Although the mean within-population gene diversity $H_S$ increases from zone 1 to zones 3 and 4, the pattern is not consistent over loci. For each of the 12 loci and for 4 additional ones taken from a study by Gömöry *et al.* (1999 and unpublished results), the average of $H$ for peripheral populations (zones 3 and 4) was plotted

**TABLE 2**

**Matrix of correlation between geographic variables and diversity parameters**

| | Long | Lat | Alt | $H_S$ | $H_T$ | $F_{IS}$ | $F_{ST}$ | $A_S$[80] | $A_T$[80] | $A_T$[400] |
|---|---|---|---|---|---|---|---|---|---|---|
| Long | — | 0.27* | −0.21 | −0.32* | −0.37** | 0.21 | −0.26* | 0.54*** | 0.60*** | 0.73*** |
| Lat | 0.28*** | — | −0.82*** | 0.43*** | 0.34** | −0.21 | −0.30* | −0.09 | −0.16 | −0.13 |
| Alt | −0.18** | −0.70*** | — | −0.43*** | −0.33* | 0.13 | 0.34** | −0.18 | 0.01 | −0.02 |
| $H_S$ | −0.23*** | 0.29*** | −0.27*** | — | 0.97*** | −0.30* | 0.01 | −0.17 | −0.27* | −0.44*** |
| $H_T$ | — | — | — | — | — | −0.34** | 0.25 | −0.20 | −0.24 | −0.45*** |
| $F_{IS}$ | 0.12* | −0.14* | 0.07 | −0.06 | — | — | −0.21 | 0.32* | 0.16 | 0.22 |
| $F_{ST}$ | — | — | — | — | — | — | — | −0.16 | 0.09 | −0.12 |
| $A_S$[80] | 0.29*** | −0.04 | −0.13* | 0.15* | — | 0.15* | — | — | 0.85*** | 0.74*** |
| $A_T$[80] | — | — | — | — | — | — | — | — | — | 0.85*** |

Long, longitude; Lat, latitude; Alt, altitude. Group level estimates are given above the diagonal, and population level estimates below (except for $H_T$, $F_{ST}$, and $A_T$, which are defined only at the group level); $A_S$[80] stands for allelic richness at the population level. *, **, ***, significant correlation coefficients at the 5, 1, and 0.1% levels, respectively.

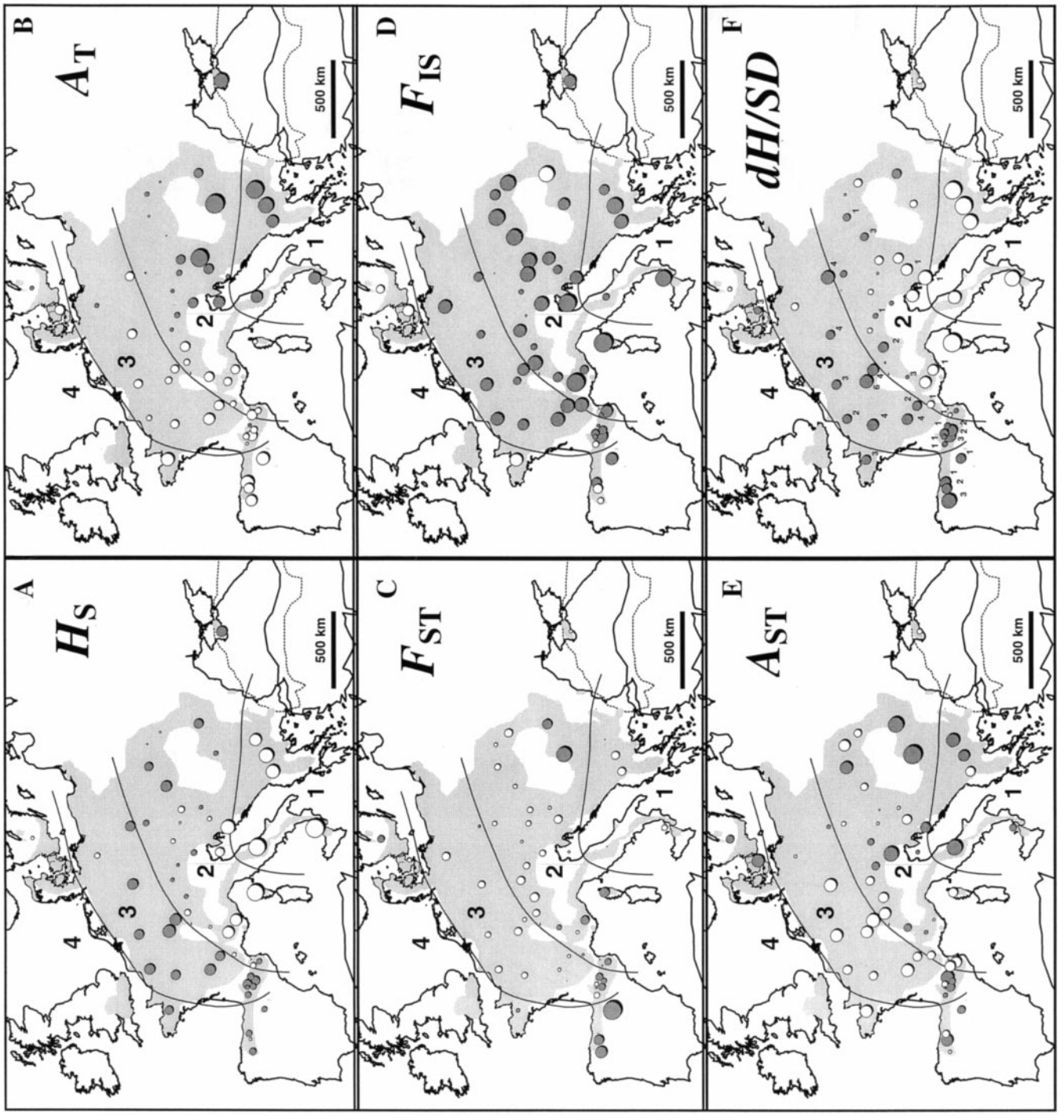
against the corresponding estimates for refugia (zone 1; Figure 3). Confidence intervals ($P = 0.95$) are provided. Six loci have higher gene diversity in refugia, and 10 have higher gene diversity in peripheral zones. However, 4 of the 6 loci for which gene diversity had decreased during postglacial colonization had already low levels of gene diversity in refugia.

## DISCUSSION

**Changes in genetic diversity during colonization:** In population genetics theory, the distribution of the number of different alleles in a sample, at mutation-drift equilibrium, is given by Ewens' (1972) sampling formula; this formula remains central to many recent developments in population genetics (*e.g.*, Cornuet and Luikart 1996; see below). In empirical studies, the rarefaction method developed in ecology by Hurlbert (1971) provides an efficient way to standardize allelic richness. Using this approach, we could demonstrate the existence of a significant and steady decline in allelic richness during postglacial recolonization of Europe by beech. This result was expected, because cumulative founding events are predicted to lead to allele losses. Because the mutation rates at isozyme loci are low (*e.g.*, Neel *et al.* 1988), much more time must elapse before new alleles can restore the initial allelic diversity.

A more surprising result is the increase of $H$ during colonization, particularly if we consider that the decrease in allelic richness should have brought about a simultaneous (though less marked) decrease in $H$. For instance, for those loci that had lost all but their most frequent allele at the outset of recolonization, it is clear that no subsequent increase in $H$ could occur. The negative correlation between the two measures of diversity ($r = -0.44$ at the group level) and the positive but still very low correlation at the population level ($r = 0.15$) contrast with the values expected at mutation-drift equilibrium for these sample sizes (0.70, Chakraborty *et al.* 1980). The differences between analyses at the group and at the population level reflect the fact that the overall geographical trend is better captured by averaging the diversity estimates over several local populations. In comparisons among neighboring populations, those characterized by higher allelic richness will tend to have also higher gene diversities, whereas this does not hold for populations further apart. The correlation between the two diversity parameters is therefore scale dependent. Contrary to distant populations, nearby populations are exposed to more similar environments and are likely to originate from the same refugium and to have been founded at the same time.

**Bottlenecks and the distribution of variation among populations:** On the basis of the high number of significant bottleneck tests (71) and their location away from glacial refugia, it may be concluded that founding events were indeed associated with postglacial recolonization


A
$H_S$
B
$A_T$
C
$F_{ST}$
D
$F_{IS}$
E
$A_{ST}$
F
*dH/SD*
500 km
1
2
3
4

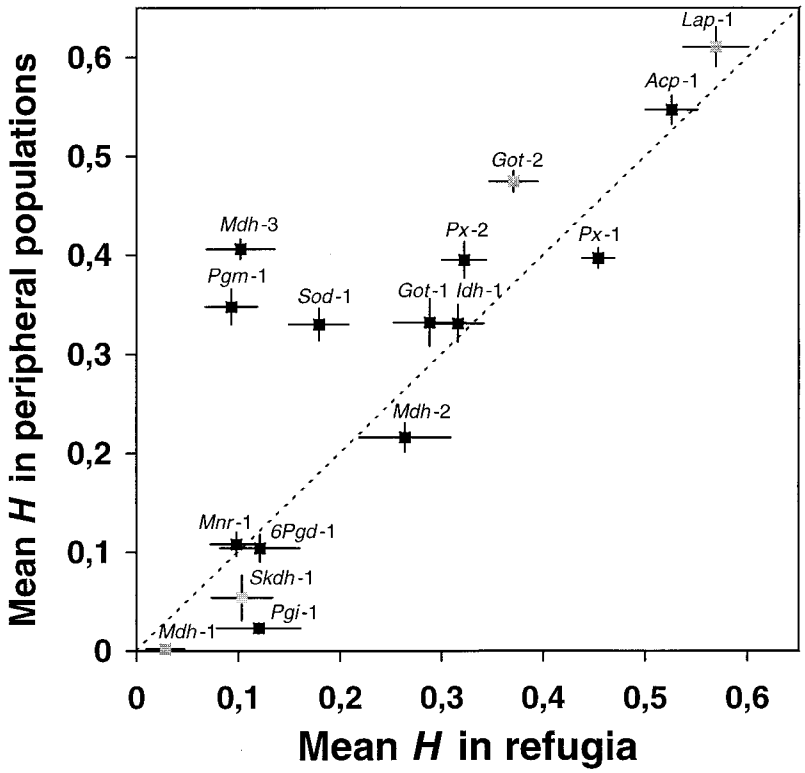



FIGURE 3.—Relationship between gene diversity at each locus in peripheral populations (zones 3 and 4) *vs.* refugia (zone 1). Standard errors are provided for each locus in both regions. Loci included in the present survey are in black, and those in gray are from GÖMÖRY *et al.* (1999).

by beech. The method of identification of bottlenecks is based on a contrast between allelic richness and $H$ (CORNUET and LUIKART 1996). Typically, bottlenecks are characterized by large losses in allelic richness and by concomitantly slight diminution of $H$, especially if population size rebounds rapidly (NEI *et al.* 1975; ELLSTRAND and ELLAM 1993). In the present example, however, the measure of the intensity of bottleneck ($dH/SD$) was strongly correlated with $H$, but less so with allelic richness (Table 2). Directional increase of $H$ (due to selection, for instance) accompanied by more limited losses of allelic richness could also result in significant tests. However, in this case, the necessary assumption of selective neutrality of the loci (CORNUET and LUIKART 1996) would not be met.

The maps for $F_{ST}$ and $F_{IS}$ estimates point to those populations colonized most recently (zone 4), which are characterized by higher differentiation and lower heterozygote deficit. These fixation indices are known to have different, more rapid dynamics compared to that of $H$ (*e.g.*, CROW and AOKI 1984). On the other hand, the measure of genetic differentiation for allelic richness ($A_{ST}$) roughly distinguishes the oldest and youngest beech populations from those of intermediate age. The different results for $F_{ST}$ and $A_{ST}$ for populations located in the refugia point again to different dynamics for rare alleles and for the frequent ones. There are few theoretical studies that consider the contrast between intra- and interpopulation allelic richness. Using the coalescent approach, WAKELEY (1998) has suggested that a measure equivalent to our $A_{ST}$ could provide better insights into levels of gene flow than the classical $F_{ST}$ approach.

**Numerical demonstration that allelic richness and $H$ can change in opposite directions:** Metapopulation models indicate the conditions under which $H$ may actually increase during colonization (WADE and MCCAULEY 1988). In the migrant pool model (colonists originating from all preexisting populations with equal probability), $H$ will increase during colonization and $F_{ST}$ will decrease if $k > 2Nm + 1$, where $k$ is the number of diploid colonists (materialized by seeds in the case of the

FIGURE 2.—Estimates of diversity, allelic richness, and fixation indices, and detection of bottlenecks in 55 groups of the European beech. Groups consist of 5–9 (mean of 7) geographically close beech populations. Values above average (or above zero for $F_{IS}$) are represented by solid circles, and values below average are indicated by open circles; circle sizes are proportional to the deviation from the mean or from zero. ($H_S$) Mean within-population gene diversity. ($A_T$) Total allelic richness per group, after rarefaction to a common sample size of 400 gene copies ($A_T[400]$). ($F_{ST}$) Coefficient of differentiation among populations within each group. ($F_{IS}$) Mean heterozygote deficit per group. ($A_{ST}$) Partitioning of allelic richness among populations within group ($A_{ST}[80]$). ($dH/SD$) Measure of the intensity of bottlenecks; the number of significant bottlenecks per group is also indicated.

beech), and $Nm$ is the number of migrants per generation (*i.e.*, involving both seeds and pollen; Le Corre and Kremer 1998; Wade and McCauley 1988). With $F_{ST} = 0.027$ in refugia (zone 1; see Table 1), we have $Nm = \frac{1}{4}(1/F_{ST} - 1) = 9.0$, and the required condition for $H$ to increase is $k > 19$. Unfortunately, allelic richness was not included in these models, so it is not known if they yield conditions when allelic richness decreases as $H$ increases. Using the method of rarefaction, it can be shown that the newly founded populations will have, in the case of 19 diploid colonists taken from throughout the refugia, a maximal allelic richness of ~2.28 (*i.e.*, $A_T[38]$). This estimate is slightly lower than the within-population allelic richness actually observed in refugia ($A_S[95] \approx 2.34$), indicating that $H$ can increase while allelic richness decreases. However, the situation described above is not very realistic. Indeed, evidence exists that founding events involved very few seeds. Detailed analyses of chloroplast DNA structure in oaks have revealed fixation of single haplotypes within small regions, the likely consequences of rare long-distance dispersal events during colonization (Le Corre *et al.* 1997; Petit *et al.* 1997). Similar results have been found in beech (Demesure *et al.* 1996; R. J. Petit, unpublished data).

**Consequences of an increased pollen flow during colonization:** A delayed version of the multiple colonists model discussed above (by pollen rather than by seeds), associated with a temporary increase of pollen flow during colonization, could partly account for the increase in $H$. Changes in gene flow rates during population growth could have important genetic consequences (Ingvarsson 1997). For species where pollen immigration is independent of the size of the population, such as wind-pollinated trees, the rate of gene flow should be highest in the first generations, resulting in higher values of $H$ compared with the situation where growth is instantaneous. In beech, pollen flow is more widespread where trees are far apart than within closely spaced populations (Thiébaut *et al.* 1990; Merzeau *et al.* 1994). The merging of beech populations originating from separate refugia in southeast Europe during colonization could have further facilitated this trend, although genetic differentiation among populations is not very large in the refugial zone ($F_{ST} = 0.027$).

**Selection during recolonization:** In the present study, an increase of total diversity $H_T$ (and not only of $H_S$) was observed during colonization, so that the mere mixing of populations from refugia does not suffice to explain the results. This could point to selective effects acting directly or indirectly on the isozyme loci. Viability selection under stress conditions, such as those that prevail during establishment of new populations, could promote an increase in observed heterozygosity at some loci, beyond that caused by the removal of the more inbred individuals (see, *e.g.*, Müller-Starck 1993). Thiébaut *et al.* (1992) have described a particularly pronounced trend of increase of observed heterozygosity with age when beech seedlings are growing under high daylight intensity. Given the more open conditions prevailing during postglacial recolonization, this selection differential could explain why newly established beech stands maintain more diversity than expected at some loci.

**Concluding remarks:** We suggest that a combination of the factors discussed above has acted successively to limit the initial loss of gene diversity and then possibly to increase it above the initial values found in the refugia. Simultaneously, allelic richness, which had decreased as a consequence of the initial founding events, was unable to recover, given the low mutation rates at isozyme loci. Although it has became clear since 10–15 years ago that a detailed historical framework is the precondition for interpreting patterns of geographic variation in genetic diversity in many plant and animal species, the direction of change during colonization may sometimes differ from that predicted by simple models.

We thank all people who have taken part in the sampling of beech twigs, particularly Ladislav Paule and his team (Zvolen, Slovakia) for their efficient help. Thanks are also due to the technicians of the French National Forest Office and to R. M. Guilbaud for her efficient technical assistance. Discussions with Valérie Le Corre helped to clarify some of the genetic consequences of colonization, whereas Simon Brewer kindly communicated unpublished maps on postglacial colonization of beech in Europe and provided many useful references. We finally thank several anonymous referees for their thorough comments on previous versions of the manuscript. R. J. Petit was supported by a grant from the European Union dealing with genetic variation in beech (FAIR3-CT96-1464).

## LITERATURE CITED

Bennett, K. D., P. C. Tzedakis and K. J. Willis, 1991 Quaternary refugia of north European trees. J. Biogeogr. **18:** 103–115.

Berglund, B. E., H. J. B. Birks, M. Ralska-Jasiewiczowa and H. E. Wright, 1996 *Palaeoecological Events During the Last 15000 Years.* Wiley, Chichester, England.

Bozhilova, E., H. Panovska and S. Tonkov, 1989 Pollen analytical investigations in the Kupena National Reserve, West Rhodopes. Geogr. Rhodopica **1:** 186–190.

Chakraborty, R., P. A. Fuerst and M. Nei, 1980 Statistical studies on protein polymorphism in natural populations. III. Distribution of allele frequencies and the number of alleles per locus. Genetics **91:** 1039–1063.

Chakraborty, R., P. E. Smouse and J. V. Neel, 1988 Population amalgation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. Am. J. Hum. Genet. **43:** 709–725.

Comps, B., B. Thiébaut, L. Paule, D. Merzeau and J. Letouzey, 1990 Allozymic variability in beechwoods (*Fagus sylvatica* L.) over central Europe: spatial differentiation among and within populations. Heredity **65:** 407–417.

Comps, B., B. Thiébaut, I. Šugar, I. Trinajstic and M. Plazibat, 1991 Genetic variation of the Croatian beech stands (*Fagus sylvatica* L.): spatial differentiation in connection with the environment. Ann. Sci. For. **48:** 15–28.

Comps, B., C. Mátyás, J. Letouzey and T. Geburek, 1998 Genetic variation in beech populations (*Fagus sylvatica* L.) along the Alpine chain and in the Hungarian basin. For. Genet. **5:** 1–9.

Cornuet, J.-M., and G. Luikart, 1996 Description and power analy-

sis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics **144:** 2001–2014.

Crow, J. F., and K. Aoki, 1984 Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. Proc. Natl. Acad. Sci. USA **81:** 6073–6077.

Cwynar, L. C., and G. M. MacDonald, 1987 Geographical variation of lodgepole pine in relation to population history. Am. Nat. **129:** 463–469.

Demesure, B., B. Comps and R. J. Petit, 1996 Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. Evolution **50:** 2515–2520.

Ellstrand, N. C., and D. R. Elam, 1993 Population genetic consequences of small population size: implications for plant conservation. Annu. Rev. Ecol. Syst. **24:** 217–242.

Ewens, W. J., 1972 The sampling theory of selectively neutral alleles. Theor. Popul. Biol. **3:** 87–112.

Gömöry, D., I. Shvadchak, L. Paule and J. Vyšný, 1998 Genetic diversity and differentiation of beech populations in Crimea. Russ. J. Genet. **34:** 63–70.

Gömöry, D., L. Paule, R. Brus, P. Zhelev, Z. Tomovic *et al.*, 1999 Genetic differentiation and phylogeny of beech on the Balkan Peninsula. J. Evol. Biol. **12:** 746–754.

Hazler, K., B. Comps, I. Sugar, L. J. Melovski, A. Tashev *et al.*, 1997 Genetic structure of *Fagus sylvatica* L. populations in Southeastern Europe. Silv. Genet. **46:** 229–236.

Hewitt, G. M., 1996 Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. **58:** 247–276.

Huntley, B., and H. J. B. Birks, 1983 *An Atlas of Past and Present Pollen Maps for Europe, 0–13000 Years Ago.* Cambridge University Press, Cambridge, UK.

Hurlbert, S. H., 1971 The nonconcept of species diversity: a critique and alternative parameters. Ecology **52:** 577–586.

Ingvarsson, P. K., 1997 The effect of delayed population growth on the genetic differentiation of local populations subject to frequent extinctions and recolonizations. Evolution **51:** 29–35.

Konnert, M., M. Ziehe, U. Tröber, W. Maurer, A. Janßen *et al.*, 2000 Genetische Variation der Buche (*Fagus sylvatica* L.) in Deutschland: gemeinsame Auswertung genetischer Inventuren über verschiedene Bundesländer. Forst Holz **55:** 403–408.

Le Corre, V., and A. Kremer, 1998 Cumulative effects of founding events during colonisation on genetic diversity and differentiation in an island and stepping-stone model. J. Evol. Biol. **11:** 495–512.

Le Corre, V., N. Machon, R. J. Petit and A. Kremer, 1997 Colonization with long-distance seed dispersal and distribution of maternally inherited diversity in forest trees: a simulation study. Genet. Res. **69:** 117–125.

Leonardi, S., and P. Menozzi, 1995 Genetic variability of *Fagus sylvatica* L. in Italy: the role of postglacial recolonization. Heredity **75:** 35–44.

Luikart, G., W. B. Sherwin, B. M. Steele and F. W. Allendorf, 1998 Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. Mol. Ecol. **7:** 963–974.

MacDonald, G. M., 1993 Fossil pollen analysis and the reconstruction of plant invasions. Adv. Ecol. Res. **24:** 377–394.

Merzeau, D., F. Di Giusto, B. Comps, B. Thiébaut, J. Letouzey *et al.*, 1989 Genetic control of isozyme systems and heterogeneity of pollen contribution in beech (*Fagus sylvatica* L.). Silv. Genet. **38:** 195–201.

Merzeau, D., B. Comps, B. Thiébaut and J. Letouzey, 1994 Estimation of *Fagus sylvatica* L mating system parameters in natural populations. Ann. Sci. For. **51:** 163–173.

Müller-Starck, G., 1993 *Auswirkungen von Umweltbelastungen auf genetische Strukturen von Waldbeständen am Beispiel der Buche (Fagus sylvatica L.).*, Schriften aus der Forstl. Fak. d. Univ. Göttingen u. d. Niedersächs. Forstl. Versuchsanst., Vol. 112. Sauerländer's Verlag, Frankfurt a. M., Germany.

Müller-Starck, G., and R. Starke, 1993 Inheritance of isoenzymes in European beech (*Fagus sylvatica* L.). J. Hered. **84:** 291–296.

Neel, J. V., C. Satoh, P. Smouse, J.-I. Asakawa, N. Takahashi *et al.*, 1988 Protein variants in Hiroshima and Nagasaki: tales of two cities. Am. J. Hum. Genet. **43:** 870–893.

Nei, M., 1973 Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA **70:** 3321–3323.

Nei, M., T. Maruyama and R. Chakraborty, 1975 The bottleneck effect and genetic variability in populations. Evolution **29:** 1–10.

Nilsson, S. G., 1985 Ecological and evolutionary interactions between reproduction of beech *Fagus sylvatica* and seed eating animals. Oikos **44:** 157–164.

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

Petit, R. J., A. El Mousadik and O. Pons, 1998 Identifying populations for conservation on the basis of genetic markers. Conserv. Biol. **12:** 844–855.

Pons, O., and K. Chaouche, 1995 Estimation, variance and optimal sampling of gene diversity. II. Diploid locus. Theor. Appl. Genet. **91:** 122–130.

Slatkin, M., 1977 Gene flow and genetic drift in a species subject to frequent local extinctions. Theor. Popul. Biol. **12:** 253–262.

Thiébaut, B., R. Lumaret and P. Vernet, 1982 The bud enzymes of beech (*Fagus sylvatica* L.). Genetic distinction and analysis of polymorphism in several French populations. Silvae Genet. **31:** 51–60.

Thiébaut, B., J. Cuguen, B. Comps and D. Merzeau, 1990 Genetic differentiation in beech (*Fagus sylvatica* L.) during periods of invasion and regeneration, pp. 379–390 in *Biological Invasions in Europe and the Mediterranean Basin*, edited by F. di Castri, A. J. Hansen and M. Debussche. Kluwer Academic Publishers, Dordrecht/Boston/London.

Thiébaut, B., B. Comps and A. Leroux, 1992 Relation hauteur-génotype dans une régénération naturelle de hêtre (*Fagus sylvatica* L.), équienne et âgée de 18 ans. Ann. Sci. For. **49:** 321–335.

Tzedakis, P. C., 1993 Long-term tree populations in northwest Greece through multiple Quaternary climatic cycles. Nature **364:** 437–440.

Wade, M. J., and D. E. McCauley, 1988 Extinction and recolonization: their effects on the genetic differentiation of local populations. Evolution **42:** 995–1005.

Wakeley, J., 1998 Segregating sites in Wright's island model. Theor. Popul. Biol. **53:** 166–174.

Watts, W. A., J. R. M. Allen, B. Huntley and S. C. Fritz, 1996 Vegetation history and climate of the last 15,000 years at Laghi di Monticchio, southern Italy. Quat. Sci. Rev. **15:** 113–132.

Communicating editor: A. H. D. Brown