Molecular Ecology (2008) **17**, 4597–4607 doi: 10.1111/j.1365-294X.2008.03941.x

Geostatistical genetic analysis for inferring the dispersal pattern of a partially clonal species: example of the chestnut blight fungus

C. DUTECH, J.-P. ROSSI, O. FABREGUETTES and C. ROBIN
INRA, UMR1202 BIOGECO, Domaine de Pierroton, F-33612 Cestas, France

Abstract

Spatial genetic analyses can be used to infer dispersal processes in natural populations. For partially clonal species with alternating sexual and asexual reproduction, the repetition of genotypes must be taken into account in analyses. The methods currently employed to evaluate the relevance of the spatial scale used for the estimation of gene flow are not suitable for these species. We investigated recently developed methods for taking into account repeated genotypes and for determining whether the sampling scale is large enough to capture all the spatial genetic structure existing within a population. We applied these methods to a fungal plant pathogen species, *Cryphonectria parasitica*, which has caused the death of many American and European chestnut trees since its introduction from Asia at the beginning of the 20th century. These methods were found to be useful for unravelling the effects of clonality and historical gene flow on the spatial genetic structure, and indicated that dispersal processes have probably occurred over a larger spatial scale than previously assumed.

Keywords: clonal lineages, *Cryphonectria parasitica*, forest pathogenic fungus, intrahaploid mating, isolation by distance, spatial genetic structure, spore dispersal

Received 15 April 2008; revisions received 7 August 2008; accepted 20 August 2008

Introduction

Estimates of gene dispersal are essential if we are to understand local adaptation and speciation in natural populations (Slatkin 1985). If gene dispersal decreases with spatial distance, then, in the absence of selection, the genetic similarity between pairs of individuals should decrease with geographical distance [the isolation by distance (IBD) model; Wright 1943]. Spatial autocorrelation methods are useful approaches to testing whether dispersal of this type occurs in natural populations (e.g. Hardy & Vekemans 1999; Smouse & Peakall 1999). Under the hypothesis of drift–dispersal equilibrium, historical gene flow can be inferred from this method, provided that data are collected at the relevant spatial scale (Rousset 1997; Rousset 2000). This approach has been successfully applied to various plant species (see Vekemans & Hardy 2004 for a review), providing estimates of gene dispersal distances consistent with direct estimates of propagule dispersal (e.g. Hardy *et al*. 2006). These methods have been used less frequently for partially clonal organisms, which alternate sexual and asexual reproduction during their life cycle (see Arnaud-Haond *et al*. 2007 for a review). For these organisms, the effect of repeated genotypes on spatial genetic structure (SGS) must be specifically considered for correct inference of historical gene flow.

Several methods have been proposed for separating out the respective contributions to SGS of the spatial distribution of clones and genotype relatedness (see for example Hämmerli & Reusch 2003). Retaining one isolate per genotype may be a useful method for assessing the spatial distance over which clonal repetitions affect SGS (i.e. the average clonal subrange; Alberto *et al*. 2005). Furthermore, removal of the effects of clonal structure in this way makes it possible to infer historical gene flow correctly (Reusch *et al*. 1999; Alberto *et al*. 2005). However, the spatial location of the single copy per genotype is critical and may bias the spatial genetic analysis (Wagner *et al*. 2005). For example, use of the central position of each clonal subrange may be associated with unrealistic biological assumptions (e.g. isotropic dispersion), whereas the choice of a random location may generate a large variance on estimates when clones are widely dispersed throughout the population.

Correspondence: C. Dutech, Fax: +33 5 57 12 26 21;
E-mail: cdutech@bordeaux.inra.fr.

Spatial genetic analyses also aim to determine whether the relevant spatial scale has been investigated. Rousset (1997) showed that estimates of gene flow based on the slope of the regression line of the genetic correlogram were correct only if the distances separating individuals were between about σ (the square root of the second moment of parent–progeny axial distance, Rousset 1997) and 20 σ. Outside this spatial scale, the inference of gene flow may be severely biased. For situations in which σ is not known, Vekemans & Hardy (2004) developed an iterative procedure for estimating the correct spatial range over which this regression should be carried out, provided that population effective density is known. However, for partially clonal organisms, it is difficult to estimate population effective density (Yonezawa *et al*. 2004).

Wagner *et al*. (2005) developed new methods providing a better description of SGS, for inferring gene flow for these species. They took repeated genotypes into account without arbitrarily focusing on a single spatial location, by developing a weighting procedure retaining the spatial positions of all individuals but applying a weighting to each genotype inversely proportional to its frequency. This method is potentially useful, but has been used only sporadically to date (but see Werth *et al*. 2006). Wagner *et al*. (2005) also introduced variograms as a tool for investigating SGS within populations, and as a useful alternative to correlogram analysis. Variograms are structure functions similar to correlograms but expressing the strength of spatial autocorrelation. Unlike correlograms, empirical (i.e. observed) variograms are used to fit a theoretical model, the parameters of which express different aspects of the spatial pattern studied (Goovaerts 1997). Interestingly, Wagner *et al*. (2005) highlighted the relationships between these model parameters and gene flow. They also showed that variograms were more efficient than correlograms for estimating the distance at which the genetic similarity between two individuals becomes independent (i.e. the range of the variogram). This distance cannot be accurately defined with correlograms because it depends on the sampling scheme (Vekemans & Hardy 2004). We made use of both correlogram and variogram analyses in this study and compared the utility of these two approaches for inferring dispersal processes for a fungal species pathogenic on plants. Many fungal plant pathogens display an alternation of sexual and asexual reproduction, but the spatial genetic structure of only a few such species has been investigated at the population level (Arnaud-Haond *et al*. 2007). However, these species deserve greater attention because they may constitute an important component of natural and cultivated ecosystems and may cause considerable economic loss (Pimentel *et al*. 2001; Gilbert 2002).

Cryphonectria parasitica (Murril) Barr, the chestnut blight fungus, is a typical plant pathogen species with a major impact on forest ecosystems and chestnut orchards. This species was introduced into North America and Europe from Asia at the beginning of the 20th century and has greatly changed the composition of the forest flora, particularly in North America, where it almost caused the extinction of the American chestnut (*Castanea dentata*; Anagnostakis 1987). *C. parasitica* grows into the bark up to the cambium, causing typical diffuse cankers and the death of the distal part of the trunk or branches. Its life cycle includes asexual and sexual stages with different dispersal mechanisms. The asexually produced pycnidiospores are thought to be dispersed over short distances (to the same or a neighbouring tree) through water splash and, possibly, over longer distances with the assistance of animal vectors (Russin *et al*. 1984). The sexual spores — ascopores — are formed in perithecia, the sexual bodies. Once mature, they are ejected forcibly into the air and dispersed over a few hundred metres by the wind (Heald *et al*. 1915). *C. parasitica* is a heterothallic ascomycete. Thus, sexual reproduction can occur only between two haploid mycelia of different mating types. However, detailed genetic analyses have shown that some crosses nonetheless occur between identical haploid mycelia (i.e. clones) because mycelia may have nuclei differing only at their mating type locus (McGuire *et al*. 2004). This kind of mating has been referred to as 'selfing' (Marra *et al*. 2004; McGuire *et al*. 2004), but we prefer to use the term 'intrahaploid mating', to distinguish this process from typical selfing, in which gametes from the same diploid individual fuse (see Giraud *et al*. 2008 for a detailed discussion of terms). This particular kind of intrahaploid mating facilitates the dispersal of a given identical haploid genotype via sexual spores in *C. parasitica*.

Milgroom & Lipari (1995) analysed the SGS within five American populations, using fingerprints and restriction fragment length polymorphism markers, and allele shared distance methods. They found an aggregation of identical genotypes at distances of a few metres (the same or neighbouring trees), consistent with limited asexual dispersal. However, it remained unclear whether IBD occurred over a larger scale in these American populations. An absence of drift–dispersal equilibrium might account for this result, although the small size of the sampling window (a few tens of metres) and/or the limited number of samplings of individuals may have made it difficult to detect significant IBD. A larger sampling scheme, with isolates separated by several hundreds of metres, is required to test these hypotheses. At this spatial scale, using the methods developed by Vekemans & Hardy (2004) and Wagner *et al*. (2005), we were able to describe in detail the SGS and the relationships between genetic similarity and spatial distance. In Europe, this pathogen spread more slowly and caused less severe epidemics than in North America, partly because of the occurrence of a mycovirus (*Cryphonectria* hypovirus 1, CHV-1, Heiniger & Rigling 1994). This double-stranded RNA virus affects the reproductive biology of the fungus, causing a

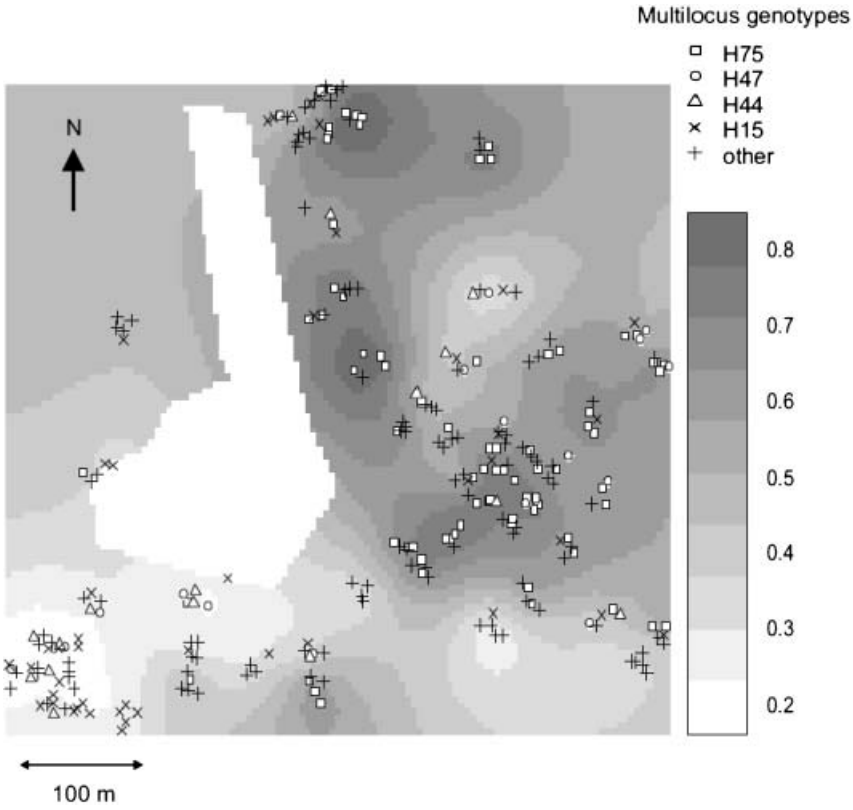

Fig. 1 Location and distribution of sampled *Cryphonectria parasitica* isolates and of the four most repeated multilocus genotypes in the French population studied. The greyscale level gives the estimated degree of genetic similarity of isolates (in number of shared alleles) to the most dominant multilocus genotype (H75) obtained by a kriging procedure (see text for details). The white area indicates an absence of sampling due to the recent logging of a chestnut plot.

loss of female fertility, and the attenuation of virulence (Nuss 2005). Most European chestnut populations (*C. sativa*) which are not located at the epidemic front, harbour the virus-infected fungus (Heiniger & Rigling 1994; Robin & Heiniger 2001). The frequency of clonal reproduction may therefore be generally higher in Europe, resulting in a more marked SGS, as suggested in a previous study (Bissegger *et al*. 1997).

We addressed three questions for a *C. parasitica* population sampled in southern France, which may be considered a study case for other partially clonal organisms: (i) Can correlogram and variogram analyses disentangle the effects of clonal distribution and gene flow on SGS? (ii) Can an IBD pattern be detected once the effect of repeated genotypes has been eliminated? (iii) Can we assess the relevance of the spatial range investigated for the inference of dispersal processes?

Materials and methods

Sampling strategy

We studied a population of *Cryphonectria parasitica* collected from the Dordogne (France), a geographical area in which *Castanea sativa* forests cover several hundred hectares as fragmented coppice stands. The survey plot (44°40′32″N, 01°02′16″E) showed two continuous large patches of *C. sativa* separated by a recently logged area formerly planted with chestnut (Fig. 1). The age of the stand was not precisely known, but had been estimated at more than 20 years (F. Dercq, personal communication). Cankers caused by *C. parasitica* were randomly distributed within the stand (personal observation). However, we chose to use a hierarchical sampling scheme to capture spatial genetic structure over scales of a few metres or of hundreds of metres. Fifty-nine clusters separated by between 25 and 600 m were located within the study plot. (Fig. 1). In each of these clusters, five trees with at least one canker at human head height separated by 5 to 20 m were selected, and one canker per tree was marked for further analyses. Different sprout clusters were sampled for each canker.

In February 2005, four bark samples per selected canker were removed with a cork borer (8 mm diameter) from the active canker area. Mass isolation was carried out as described by Robin *et al*. (2000) on PDA (potato dextrose agar medium, Difco Laboratories) plates in the laboratory. For each canker collected, we chose one of the four mycelia obtained at random for subsequent genetic analysis. After the transfer of the studied isolates onto fresh plates, we recorded the culture morphology of all the isolates studied after 4 days in the dark at 25 °C and 4 days in the light at room temperature as 'white' (W, no pigmentation and few or no asexual spores) or 'orange' (O, orange mycelium and abundant sporulation). The white phenotype of *C. parasitica*

is associated with CHV-1 infection (Nuss 1992) and is used to identify hypovirulent isolates. We explored the spatial structure of *Cryphonectria* hypovirus infections using a variogram calculated on binary data (i.e. presence/absence). We calculated 95% confidence envelopes for empirical variograms by permutation of the data values on the spatial locations (999 permutations) using the R package geoR (Ribeiro & Diggle 2001). The strains used for this study are maintained in the INRA (Institut National de la Recherche Agronomique, Bordeaux) collection.

Finally, we looked for sexual structures (perithecia) of *C. parasitica* in a subsample of 90 cankers evenly distributed in the studied site in April 2007. Bark samples (2 × 2 cm) were then removed in all of the cankers carrying visible stromata (mycelial structures in which asexual and sexual spores are formed). These samples were examined under a dissecting microscope to confirm the presence of perithecia in the cankers.

DNA extraction and genotyping of isolates

Total DNA was extracted from lyophilized mycelium obtained from cultures on PDA plates overlaid with cellophane and ground to a fine powder, as described by Hoegger *et al.* (2000). DNA was extracted with a slightly modified version of the procedure described by Breuillin *et al*. (2006), in which the extraction buffer was replaced by a buffer containing sodium dodecyl sulphate (SDS), as described by Milgroom *et al*. (1992).

Isolates were genotyped based on 10 microsatellite loci from the *C. parasitica* genome: I07-650 (Davis *et al*. 2005), CPE1, CPE3, CPE4, CPE5, CPE8, CPG3, CPG4 (Breuillin *et al*. 2006), CPG6 and CPG14 (Kubisiak *et al*. 2007). Several loci were amplified in the same polymerase chain reaction (PCR): CPE1, CPE8 and CPG3 (PCR1), CPE4 and CPE5 (PCR2), CPG4 and CPE3 (PCR3), CPG6, CPG14 and I07–650 (PCR4). Amplification was carried out in a total volume of 15 μL, containing 2 μL of genomic DNA, 1× PCR buffer [160 mM $(NH_4)_2SO_4$, 670 mM Tris-HCl (pH 8.8 at 25 °C), 0.1% Tween-20], 5 μM dNTPs (each 1.25 μM), 2.5 mM $MgCl_2$, 2 μM of each of the six (or four) primers and 0.5 U SilverStar *Taq* polymerase (Eurogentec). PCR1 and PCR2 involved heating at 94 °C for 4 min, followed by 32 cycles of 94 °C for 45 s, 55 °C for 45 s and 72 °C for 1 min and a final extension phase at 72 °C for 7 min. PCR3 and PCR4 involved heating at 94 °C for 3 min, followed by 10 cycles of 94 °C for 45 s, an annealing temperature decreasing from 60 °C to 55 °C (0.5 °C decrease per cycle) for 45 s, and 72 °C for 1 min, followed by 25 cycles of 94 °C for 45 s, 55 °C for 45 s and 72 °C for 1 min and a final extension phase at 72 °C for 7 min. PCR products were sized on a CEQ8000 sequencer (Beckman Coulter Inc., 2002), using an internal size standard (DNA Size Standard Kit –400, Beckman Coulter) and fluorescent labels at the 5′ end of the forward primers (CPE1, CPE4, CPG4 and CPG6: green-HEX; CPE8, CPG3, I07-650 and CPE3: blue-FAM; CPE5 and CPG14: black-NED). PCR products were purified according to the manufacturer's protocol, and PCR1 and PCR2 were mixed together (0.8 and 2 μL of purified product, respectively) and sized on the sequencer. PCR3 and PCR4 were mixed in the same proportions as PCR1 and PCR2 and sized in a second run on the sequencer.

Mating type was determined for a randomly chosen subsample of the studied isolates (155 of the 276 genotyped isolates) by PCR (McGuire *et al*. 2004). The M1-GS3-rev and M1GS1n primers were used to amplify MAT-1 and primers gsl-d-l and M2-GS3 were used to amplify MAT-2. PCR was carried out under the conditions described in a previous study (Marra & Milgroom 2001).

Data analysis

Genetic diversity was initially characterized by assessing allelic diversity and its two components: allelic richness (mean number of alleles per locus) and gene diversity index (H_E; Nei 1987). Population genetic diversity was also estimated by genotypic richness (the observed number of different genotypes; g_{obs}) in the sample, and the maximum number of different genotypes as a function of the number of alleles ($g_{max} = \prod_{i=1}^{l} a_i$; with a_i the observed number of alleles for the *ith* locus). We assessed whether random sexual reproduction occurred by estimating the index of multilocus linkage disequilibrium (r_d) implemented in MultiLocus version 2.1 (Agapow & Burt 2001). This index is based on the index of association (I_A) defined by Brown *et al*. (1980), modified to remove dependence on the number of loci (Agapow & Burt 2001). I_A was originally used to test for random recombination between pairs of loci by comparing the observed and expected variance (assuming random recombination between individuals) of genetic distance between all pairs of individuals (Maynard Smith *et al*. 1993). Departure from the null hypothesis (no linkage disequilibrium; $r_d = 0$) was assessed by permuting alleles between individuals independently for each locus (1000 permutations), using MultiLocus version 2.1 (Agapow & Burt 2001).

The probability that a haploid multilocus genotype (MLG) was obtained by chance through a sexual event (P_{gen}) was calculated as described by Wang *et al*. (1997). We also estimated the probability of the second occurrence of each MLG observed at least twice in the population studied, assuming that the isolates with the same MLG were derived from different random sexual reproductive events (P_{sex}). P_{gen} and P_{sex} were estimated using GenClone version 2.0 (Arnaud-Haond & Belkhir 2007). Finally, the genetic distance for each MLG was calculated as the number of alleles common to the MLG considered and the two most frequent MLGs observed in the population.

We investigated SGS with both correlograms and variograms. Both methods are based on estimates of a coefficient measuring the degree to which individuals are similar (correlogram for the kinship coefficient, Vekemans & Hardy 2004) or dissimilar (variogram for genetic and genotypic diversity, Wagner *et al*. 2005), at various spatial scales (see below). We used the kinship coefficient (F_{ij}) proposed by Loiselle *et al*. (1995) as a means of measuring the relatedness between each pair of individuals. In haploid species like *C. parasitica*, F_{ij} equals Moran's index (I), the classical measure of autocorrelation (Legendre & Legendre 1998). A 95% confidence envelope was estimated by random permutations (499 permutations) of the individual locations (Manly 1997) to test whether the observed values were significantly different from 0. Furthermore, assuming gene dispersal in two spatial dimensions, SGS was finally summarized with *Sp* statistics estimated as $-b/(1 - F_{(1)})$, with b the slope of the regression line between F_{ij} and the logarithm of geographical distance separating the pairs of isolates, and $F_{(1)}$, the mean kinship coefficient of the first distance class (see Vekemans & Hardy 2004 for details). The pairwise matrix of F_{ij} was calculated with SPAGeDI software (Hardy & Vekemans 2002). The *Sp* statistics were tested by comparing the observed value with the values obtained for 499 random permutations of the spatial locations of isolates.

For a single locus, l, the variogram of gene diversity is defined as:

$$\hat{\gamma}_l(r) = \sum_k \sum_{a<b} \frac{\chi_{ab}^{(r)}}{2n_r} (z_{lka} - z_{lkb})^2,$$

where n_r is the number of pairs of the sample falling into a series of distance classes r, $z_{ka} = 1$ if the gene copy a is of allele k and 0 if otherwise, and $\chi_{ab}^{(r)}$ is the Kronecker weight for the pair of observations a and b. $\chi_{ab}^{(r)}$ if a pair of the sample belongs to the distance class r and $\chi_{ab}^{(r)} = 0$ if otherwise. The variogram of multilocus gene diversity $\hat{H}(r)$ is the weighted sum of $\hat{\gamma}_l(r)$ for L loci (see Wagner *et al*. 2005 for details). Similarly, the variogram of genotypic diversity $\hat{D}(r)$ is obtained by coding each multilocus genotype by a dummy variable $Z_{g'}$ which takes the value of 1 if individual a is of genotype g and 0 if otherwise. $\hat{D}(r)$ may be considered to be the probability of sampling two individuals of different genotypes as a function of the distance separating those individuals (Wagner *et al*. 2005). Correlograms and variograms were estimated using data pairs (for individuals) grouped into 50 distance classes. The spatial lag was 10 m and the sampling locations in the first distance class were separated by a mean distance of 7 m. The number of pairs used to estimate kinship coefficient and the semivariances ranged from 260 to 915 per distance class. The significance of correlograms and variograms was assessed by performing 499 randomizations of the spatial location of individuals, resulting in estimates of the 97.5% and 2.5% statistical confidence envelopes. Calculations were carried out with R software (R Development Core Team, 2007).

We assessed the effect of repeated MLGs on SGS, by carrying out autocorrelation analyses using the weighting procedure proposed by Wagner *et al*. (2005), in which each MLG is assigned a weight inversely proportional to the number of repeats in the sample. Thus, pairwise coefficients were weighted $1/(n_i n_j)$, where n_i and n_j are the number of copies for the *ith* and *jth* MLG in the population, respectively. Pairs of isolates with the same MLG were not considered. This procedure was performed to generate the variogram for gene diversity and the correlogram for kinship coefficient.

Finally, we used the classical variogram formulation to analyse the spatial autocorrelation of the proportion of alleles shared between each isolate and one of the four most common genotypes. Again, 499 randomizations were used to assess statistical significance. When possible, we fitted a variogram model to the empirical values. The resulting variogram model was used for a punctual kriging interpolation procedure (Isaaks & Srivastava 1989) to produce isarithmic maps for the assessment of genetic similarity. The package geoR was used for all calculations (Ribeiro & Diggle 2001).

Results

Occurrence of the Cryphonectria *hypovirus (CHV-1) within the population*

Of the 276 strains analysed, 51.3% were white and were assumed to be infected with CHV-1. Infection rates did not differ between the four dominant MLGs (see below, $\chi^2 = 7.087$, d.f. = 3, $P = 0.069$) and the two most frequent MLGs: H15 and H75 ($\chi^2 = 0.278$, d.f. = 1, $P = 0.597$). Spatial analysis revealed a lack of spatial structure for the presence of the virus, the variogram having a flat shape (not shown).

Genetic diversity and recombination between MLGs

We genotyped 276 isolates based on the 10 microsatellite loci. The number of alleles per locus was estimated at 5.6, and the mean H_E was 0.61 (+/–0.03). For genotypic diversity, 92 MLGs were observed (g_{obs}), four of which were repeated more than 15 times (H75, H15, H44 and H47; Table 1), and 71 of which were observed only once in the sampled population. The maximum number of different genotypes based on the number of alleles observed (g_{max}) was greater than 20×10^6.

The probability of the same MLG occurring more than once because of recombination in a randomly mating population (P_{sex}) was less than 0.05 for 14 of the 21 MLGs sampled at least twice (Table 1). However, the probability of sampling the most common MLGs at least three times,

MLG identity	No. of repeats	P_{gen}	P_{sex}	Mating type 1:2	Infection %
H75	79	0.0006	0.151	17:33	53.2
H15	45	0.0001	**0.029**	20:6	58.1
H44	16	1×10^{-08}	$\mathbf{4 \times 10^{-06}}$	7:1	25
H47	16	3×10^{-08}	$\mathbf{9 \times 10^{-06}}$	1:7	58.8
H36	6	3×10^{-11}	$\mathbf{8 \times 10^{-09}}$	ND	16.7
H86	5	2×10^{-05}	**0.006**	3:0	80
H91	5	0.0004	0.121	1:0	20
H74	4	0.0003	0.082	1:1	50
H25	3	9×10^{-05}	**0.024**	2:0	100
H72	3	0.0007	0.177	1:2	100
H80	3	0.0003	0.087	0:3	33.3
H01	2	0.0004	0.127	0:1	100
H06	2	8×10^{-07}	$\mathbf{2 \times 10^{-04}}$	2:0	50
H14	2	0.0002	**0.047**	0:2	50
H29	2	3×10^{-15}	$\mathbf{7 \times 10^{-13}}$	0:1	100
H31	2	2×10^{-10}	$\mathbf{6 \times 10^{-08}}$	1:1	50
H37	2	5×10^{-11}	$\mathbf{1 \times 10^{-08}}$	0:2	50
H48	2	2×10^{-10}	$\mathbf{6 \times 10^{-08}}$	2:0	50
H56	2	6×10^{-08}	$\mathbf{2 \times 10^{-05}}$	0:2	50
H59	2	8×10^{-07}	$\mathbf{2 \times 10^{-04}}$	ND	0
H65	2	0.0007	0.184	0:1	50

ND, not determined.

Table 1 Number of repeats, probability of first occurrence (P_{gen}) and second occurrence (P_{sex}) assuming random sexual reproductive events for the 21 multilocus genotypes repeated at least twice in the *C. parasitica* population studied (in bold, $P_{sex} < 0.05$, see text for details). The observed mating types for a subsample of individuals and the percentage infection are also given

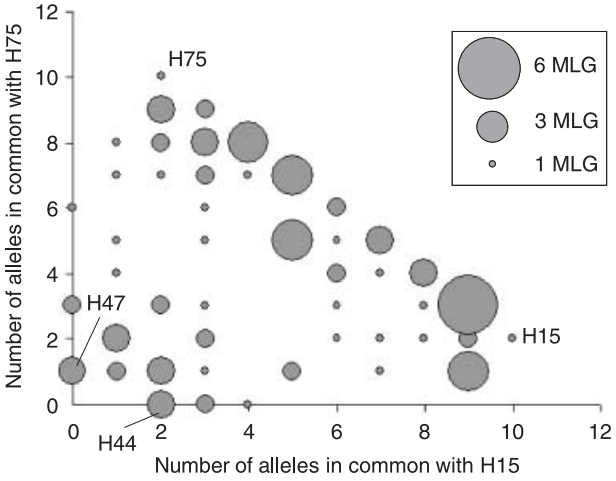

Fig. 2 Distribution of the various multilocus genotypes observed in the *Cryphonectria parasitica* population as a function of the number of alleles identical to H15 (*x*-axis) or to H75 (*y*-axis). The analysis was performed with a single copy of each multilocus genotype.

assuming random mating, was systematically < 0.05 (data not shown). We amplified MAT-1 from 65 and MAT-2 from 82 of the 156 isolates for which mating type was determined. Both mating types were amplified for eight isolates; these isolates were counted for both categories. The mating-type ratio did not deviate significantly from a 1:1 ratio if the total sample of isolates ($\chi^2 = 0.788$, $P = 0.375$) or H75 isolates only (mating type ratio was 17:33, $\chi^2 = 2.627$, $P = 0.105$) were considered. However, a significant deviation from the 1:1 ratio was detected for H15 isolates (ratio = 20:6, $\chi^2 = 4.064$, $P = 0.044$). Perithecia were observed in 19 cankers from the 90 cankers checked for the presence of sexual structures.

The multilocus index of association (r_d) was 0.410 if all the isolates analysed were considered, and 0.141 if we considered only one copy per MLG. This index, with or without the repeated MLGs, was significantly different from 0 ($P < 10^{-3}$). For each MLG, we calculated the number of alleles identical in the MLG considered and in H75 and H15, the two most repeated MLGs within the population. This analysis showed that H75 and H15 shared two alleles for the 10 microsatellite loci analysed. About half the observed MLGs (47 of the 90 MLGs) had at least seven alleles in common with either H75 or H15 (Fig. 2). Twenty-four MLGs fell on the line linking H15 and H75 corresponding to genotypes potentially recombining between these two most dominant MLGs. Many of these MLGs were close to either H15 or H75. Another group consisted of 26 haplotypes sharing fewer than five alleles with either H15 or H75. H44 and H47 were included in this group. Finally, this comparison demonstrated the low frequency of MLGs with an allelic similarity intermediate between that of H75, that of H15 and the third group of MLGs (Fig. 2), suggesting a low rate of recombination between the four dominant MLGs.

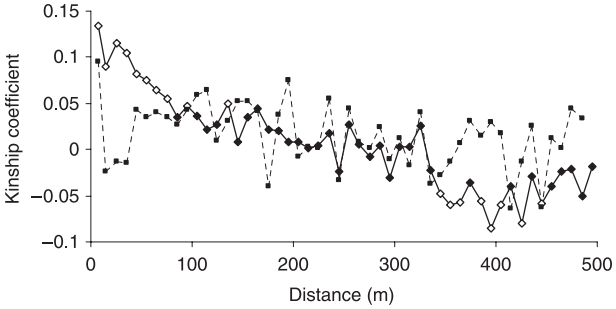

Fig. 3 Correlogram of kinship coefficients among pairs of *Cryphonectria parasitica* isolates as a function of pairwise distance classes from the French population studied. Diamond and square symbols indicate the kinship coefficient estimated with all isolates (nonweighting procedure) and after clone correction (weighting for repeated genotypes), respectively. Closed symbols do not differ significantly from the value obtained for a random structure ($P > 0.05$; obtained after 499 permutations of spatial location).

Spatial genetic structure

As expected under conditions of decreasing gene dispersal over spatial distance, kinship coefficients decreased steadily with increasing geographical distance between two isolates (Fig. 3). This pattern was observed only when all isolates were considered (i.e. with a nonweighting procedure). When repeated MLGs were weighted according to their frequency within the sampled population, the SGS did not differ significantly from a random pattern. The *Sp* statistic, quantifying SGS, was 0.062 for nonweighted and 0.001 for weighted autocorrelation. The slope and the *Sp* statistic differed significantly from those expected under the hypothesis of a random spatial distribution of genotypes when all isolates were considered ($P < 0.002$ for the two tests). After weighting for repeated genotypes, no mean value for Moran's index per distance class differed significantly from 0, but a slight decrease of genetic similarity over distance was detected, with a slope and *Sp* significantly different from those expected under the hypothesis of a random spatial distribution of genotypes ($P = 0.012$ and $P = 0.010$, respectively).

The variogram showed a regular increase in semivariance for gene diversity $\hat{H}(r)$ when all isolates were considered (Fig. 4a). The statistical envelope based on 499 randomizations indicated that a nonrandom spatial genetic structure occurred for distances of less than 100 m and more than 300 m (Fig. 4a). A bias may be introduced for larger distances of separation, because only observations from the edge of the sampled domain contribute to the calculations. Variogram analysis is therefore generally limited to one-half or two-thirds of the maximum distance between sampling points (Isaaks & Srivastava 1989). We report here the whole variogram, because we were interested in its general shape and whether or not it reached a plateau. The semivariance

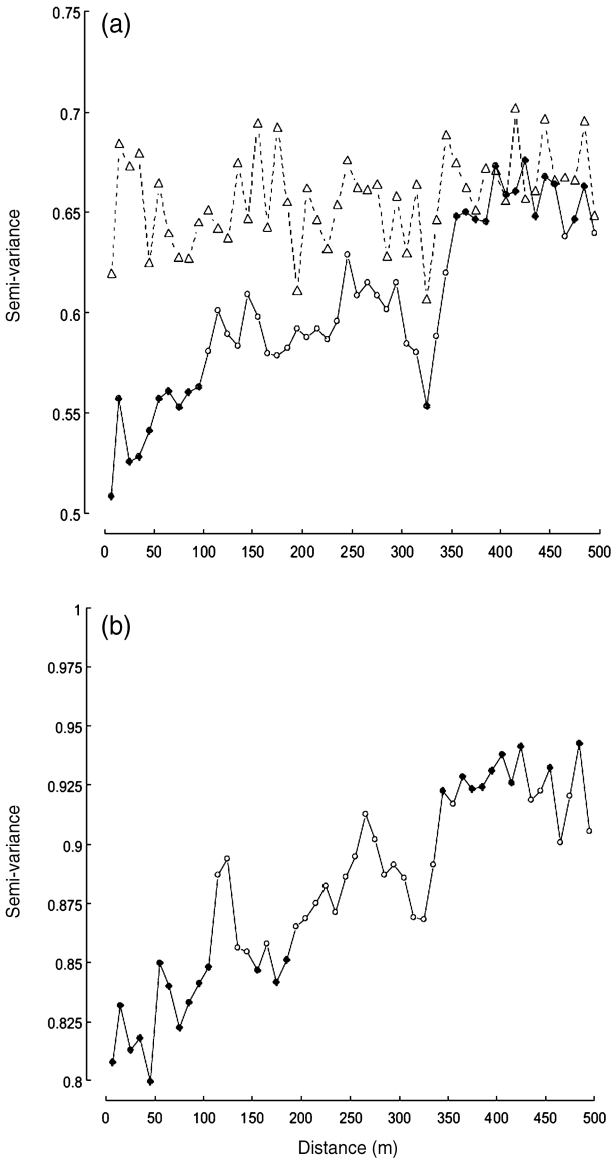

Fig. 4 Variogram of gene diversity among *Cryphonectria parasitica* isolates as a function of pairwise distance classes (4a) considering all the isolates (circles) or weighted for repeated genotypes (triangles), or of genotypic diversity (4b) from the French population studied. Open symbols indicate the absence of a significant difference from the value obtained for a random structure ($P > 0.05$; obtained after 499 permutations of spatial location).

did not reach a plateau, instead increasing regularly with spatial distance (Fig. 4a). Using the weighting scheme proposed by Wagner *et al*. (2005) to correct for repeated genotypes, we detected no significant spatial genetic structure and the variogram was flat (Fig. 4a). The variogram of genotypic diversity had a shape very similar to that described above. Genotypic diversity ($\hat{D}(r)$) increased

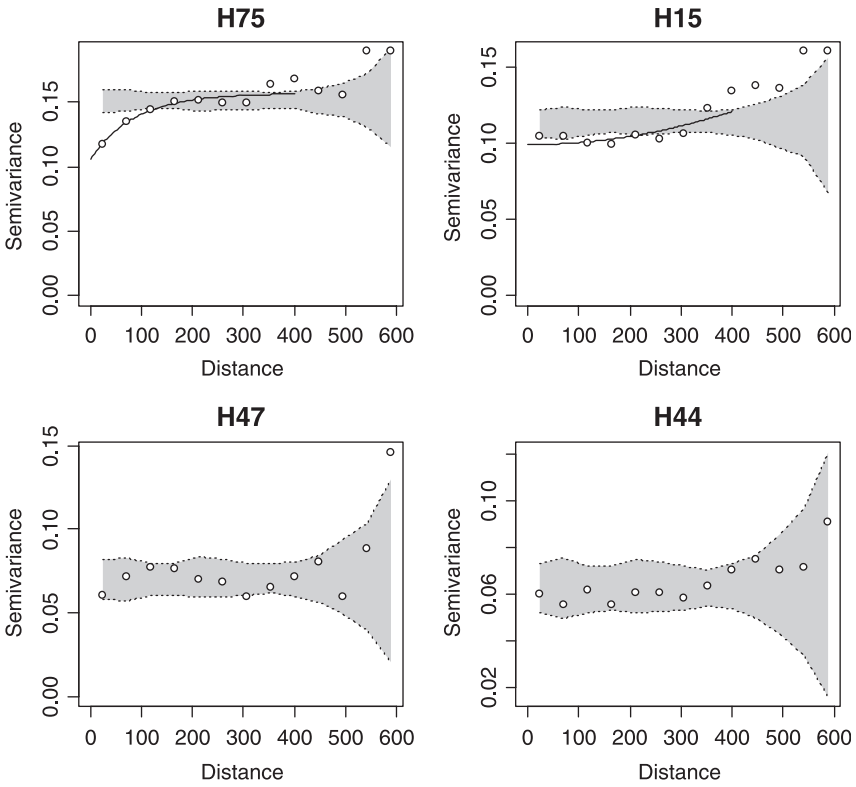

Fig. 5 Variogram of genetic similarity (in number of shared alleles) to the multilocus genotypes H75, H15, H44 and H47 among *Cryphonectria parasitica* isolates as a function of pairwise distance classes from the French population studied. The grey area indicates values not significantly different from those obtained for a random structure ($P > 0.05$; obtained after 499 permutations of spatial location).

linearly with the distance separating the isolates (Fig. 4b). Again, no plateau was observed.

Analysis of the proportion of alleles common to each isolate and one of the four most common genotypes (i.e. genetic similarity) yielded the variograms reported in Fig. 5. Spatial autocorrelation occurred in two cases, for the proportion of alleles shared with H75 and H15 (Fig. 5). Models were fitted to the empirical variograms, using the observed semivariance values for lag distances below 400 m, corresponding to two-thirds of the maximum distance between sampling points (Isaaks & Srivastava 1989). No plateau occurred for H15, whereas the variogram obtained for genetic similarity to H75 reached a constant value for distances greater than 200 m. The plateau value was determined by fitting a parametric Gaussian model (Isaaks & Srivastava 1989) to an empirical variogram, using weighted least-square estimates. The weighting of each semivariance estimate was proportional to the number of data pairs involved in its calculation (Cressie 1985). No significant spatial structure was detected in terms of similarity to the other two dominant MLGs (H47 and H44; Fig. 5), probably because frequencies were too low to detect spatial structures. We report only the map of genetic similarity between the isolates and H75 (Fig. 1). The lack of significant spatial structure or the existence of a structure beyond the spatial scale studied for the other variograms resulted in maps that were uninterpretable.

Discussion

Both methods of spatial analysis (variogram and autocorrelogram) indicated a strong SGS when all the isolates were considered (i.e. with the nonweighting procedure). These results were associated with a significant clustering of repeated genotypes over distances of 0 to 100 m, particularly for the two dominant MLGs (H75 and H15). We cannot completely rule out an effect of selection due to microenvironmental heterogeneity (e.g. microclimatic conditions or SGS at a fine spatial scale for *Castanea sativa*), but this clustering of markers assumed to be neutral is consistent with the dispersal over short distances of asexual spores. The high prevalence of the hypovirus in this population may favour asexual dispersal and the large clonal patches observed (Fig. 1), because fungal infection results in a loss of female fertility (Nuss 2005). However, repeated genotypes were observed at greater distances than those reported for American populations, in which identical genotypes were observed mostly in the same or in neighbouring chestnut trees (i.e. a few metres apart; Milgroom & Lipari 1995). Our results therefore suggest that asexual spores disperse over larger distances than previously thought. Asexual spore dispersal over distances of a few metres probably involves rain or small animals, where insects may occasionally disperse spores over distances of up to several hundred metres (Russin *et al*. 1984). Alternatively, spores produced

by intrahaploid mating may contribute to the spread of identical genotypes (see Introduction). Marra *et al*. (2004) estimated the intrahaploid mating rate in European populations to be up to 20%. Both mating types were detected within many repeated MLGs, and thus, intrahaploid mating may also occur in this French population. The sexual spores generated through intrahaploid mating are thought to disperse over greater distances than asexual spores (see for example Milgroom & Lipari 1995). This may thus favour the establishment of clonal genotypes several hundred metres away from the initial disease focus.

We were unable to determine average clonal subrange in this study because the observed SGS was close to that for a random structure after the removal of the effect of repeated genotypes by the weighting procedure. Thus, the spatial distribution of repeated genotypes accounted for most of the SGS at distances of 0 to 600 m. Both the weighting procedure and the steady increase in the semivariance of genotypic diversity suggested that the spatial distance over which clonal repetitions affect SGS was larger than the spatial scale investigated in this population. The random SGS observed after removing the effect of repeated genotypes could be due to the small number of different genotypes detected in this analysis (only 92 different MLGs), limiting the power of this study to detect IBD. However, the absence of a significant IBD pattern may also be due to sexual spores being dispersed over larger distances than estimated in previous studies (i.e. a few hundred metres, Heald *et al*. 1915). Spores may disperse over larger distances in European populations than in American populations, on which all previous estimates were based. This may be due to differences in climatic conditions, the density and spatial distribution of host species, or the activities of animal vectors. Furthermore, direct estimates by capture, as used by Heald *et al*. (1915), do not generally estimate long-distance dispersal efficiently (Nathan *et al*. 2003). This random SGS might also result from an absence of drift–dispersal equilibrium, due to recent colonization, leading to a random distribution of related genotypes. *C. parasitica* was first detected in this area in the 1970s (Grente 1981), but the population studied here may have been founded more recently. Colonization may have occurred a few sexual generations ago, because an IBD structure generally becomes established soon after the founding event (between 10 and 20 sexual generations later; Hardy & Vekemans 1999; Leblois *et al*. 2004).

In the absence of historical data, we were unable to draw any firm conclusions concerning the absence of an IBD pattern and its relationship to recent colonization. However, we assume that the colonization process might affect the overall SGS (i.e. when all isolates were included in the analysis) of this *C. parasitica* population, even several years after the founding event. Figure 1 indicates that there is a large difference in the density of the two dominant MLGs (H15 and H75) and their related MLGs between the southwestern and northeastern parts of the study plot. This pattern may be attributed to colonization by two MLGs at opposite sides of the stand, followed by an expansion of these genotypes over short distances, mostly by asexual means and through intrahaploid mating. A colonization process of this type was suggested as a possible cause of the recent spread of *C. parasitica* populations in Switzerland (Bissegger & Heiniger 1994; Hoegger *et al*. 2000). The H15 and H75 clonal lineages were highly differentiated genetically and displayed only low levels of mixing in the plot, creating a cline of allelic frequencies. As argued by Wagner *et al*. (2005), a genetic cline may account for the continuous, linear increase in the semivariogram for genotypic and genetic diversity observed in this study. Furthermore, the semivariogram of genetic similarity to H15 did not reach a plateau as it did for H75. This result is also suggestive of the formation of a clinal structure. H15 was more clustered in the western part of the population, and thus, our study did not totally capture the SGS of this dominant MLG and its related genotypes. If this hypothesis is correct for the French population, it should be confirmed by spatial analyses at a larger spatial scale. We would expect to find a higher frequency of H15 than of H75 in the western part of the new studied area, and the opposite result in the eastern part.

Surprisingly, this study showed that rare MLGs were highly similar to the four dominant MLGs (H75, H15, H44 or H47). Most of these rare MLGs differed from the four dominant MLGs by only one or two alleles over the 10 loci analysed. These differences between MLGs could have been due to genotyping errors or somatic mutations. However, these two hypotheses were rejected, because most of the isolates were genotyped at least twice, and because the one or two alleles differentiating rare MLGs from dominant MLGs were generally identical to those found in other dominant MLGs (data not shown). Size differences between alleles generally exceeded one microsatellite repeat motif, strongly suggesting that alleles were inherited by a limited process of recombination between dominant MLGs rather than being generated through mutations. There are at least two possible reasons for this. First, recombining genotypes, other than those genetically similar to the dominant genotypes, may be poorly adapted to local abiotic (climate) and biotic (*Castanea sativa* genotypes, local CHV-1 strains, etc.) conditions. However, there is no experimental evidence to suggest that such genotypes have a higher fitness, and the components of fitness (e.g. mycelium growth or sporulation rates) should be compared between recombined genotypes to test this hypothesis more directly. Furthermore, assuming that most of the microsatellite loci used in this study were neutral and physically independent, such a mechanism would occur only if selected genes were linked to these loci. Second, a recent study in an American *C. parasitica* population suggested that parasexual recombination may occur after heterokaryon formation between incompatible vegetative

lineages (McGuire *et al*. 2005). This process, which has been described for *C. parasitica* only in the laboratory (see for example Rizwana & Powell 1995), would occur after the fusion of the two mycelia and the transient fusion of two different nuclei. At this stage, the nuclei exchange chromosomes (or parts of chromosomes) by mitotic recombination. Parasexuality may occur frequently between the four dominant MLGs, generating many of the nondominant MLGs observed in this study. Sexual structures were observed in this population, but the observed genotypic structure suggests that sexual reproduction between different clonal lineages would contribute less to local epidemic spread than asexuality, intrahaploid mating, inbreeding and putative parasexuality.

The semivariogram of genetic and genotypic diversity clearly indicated a continuous increase in the variance of genetic similarity with increasing spatial distance separating two isolates. This finding illustrates the efficiency of the variogram method for assessing the relevance of the spatial scale studied for the inference of spatial dispersal processes. As argued by Wagner *et al*. (2005), in an IBD model with dispersal processes occurring mostly at the scale studied, the semivariogram is characterized by a plateau (i.e. a constant variance of genetic similarity between individuals) at the largest distances investigated. Thus, the sampling window (600 m) used in this study was smaller than the scale at which the processes studied shape the SGS. The relevance of spatial scales for the correct inference of dispersal distances was much more difficult to assess using Moran's correlogram alone (Vekemans & Hardy 2004). Our results confirm that semivariograms are potentially useful for determining the spatial scale of dispersal processes, constituting a feasible alternative to the iterative procedure developed by Vekemans & Hardy (2004). Furthermore, Wagner *et al*. (2005) showed that variogram modelling can facilitate estimation of the *Sp* parameter and historical gene flow. The *Sp* parameter is an interesting index for quantifying spatial structure (Vekemans & Hardy 2004). However, it can be used to measure historical gene flow if, and only if, the scale of the study matches the spatial scale of genetic variability under a hypothesis of drift–dispersal equilibrium. Our variogram analyses strongly suggested that either the spatial scale investigated was too small or that the *C. parasitica* population was not at equilibrium. Thus, in this case, *Sp* can be used only to estimate the level of structuration, and inferences about dispersion should be avoided. This study provides insight into how variograms for genetic diversity can facilitate estimation of the spatial scale over which dispersal processes occur in various species (whether partially clonal or not).

Acknowledgements

We would like to thank F. Dercq and F. De Blay for having allowed us to collect samples from the chestnut stand, X. Capdevielle for assistance in the field and with GPS data, S. Werth and R. Leblois for providing us with data for testing the R program for variogram analysis. We also thank S. Richard-Cervera and C. Aguilar for assistance in the laboratory, and S. Arnaud-Haond, F. Delmotte, T. Giraud, F. Halkett, R. Leblois, and MG Milgroom for helpful comments on the manuscript. This project was supported by a grant from INRA (Département Santé des Plantes et Environnement and Département Ecologie des Forêts, Prairies et milieux Aquatiques: Projet Epidémiologie et dynamique des populations).

References

Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, **1**, 101–102.

Alberto F, Gouveia L, Arnaud-Haond S *et al*. (2005) Within-population spatial genetic structure, neighbourhood size and clonal subrange in the seagrass *Cymodocea nodosa*. *Molecular Ecology*, **14**, 2669–2681.

Anagnostakis SL (1987) Chestnut blight: the classical problem of an introduced pathogen. *Mycologia*, **79**, 23–37.

Arnaud-Haond S, Belkhir K (2007) GenClone: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Molecular Ecology Notes*, **7**, 15–17.

Arnaud-Haond S, Duarte CM, Alberto F, Serrao EA (2007) Standardizing methods to address clonality in population studies. *Molecular Ecology*, **16**, 5115–5139.

Bissegger M, Heiniger U (1994) Aerial photographs reveal chestnut-blight foci. *European Journal of Forest Pathology*, **24**, 419–423.

Bissegger M, Rigling D, Heiniger U (1997) Population structure and disease development of *Cryphonectria parasitica* in European chestnut forests in the presence of natural hypovirulence. *Phytopathology*, **87**, 50–59.

Breuillin F, Dutech C, Robin C (2006) Genetic diversity of the chestnut blight fungus *Cryphonectria parasitica* in four French populations assessed by microsatellite markers. *Mycological Research*, **110**, 288–296.

Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics*, **96**, 523–536.

Cressie N (1985) Fitting variogram models by weighted least squares. *Mathematical Geology*, **17**, 563–586.

Davis JE, Kubisiak TL, Milgroom MG (2005) Polymorphic sequence-characterized codominant loci in the chestnut blight fungus, *Cryphonectria parasitica*. *Molecular Ecology Notes*, **5**, 195–197.

Gilbert GS (2002) Evolutionary ecology of plant diseases in natural ecosystems. *Annual Review of Phytopathology*, **40**, 13–43.

Giraud T, Yockteng R, Lopez-villavicencio M, Refrégier G, Hood ME (2008) Mating system of the anther smut fungus *Microbotryum violaceum*: selfing under heterothallism. *Eukaryotic Cell*, **7**, 765–775.

Goovaerts P (1997) *Geostatistics for Natural Resources Evaluation*. Oxford University Press, Oxford, UK.

Grente J (1981) *Les variants hypovirulents de l' Endothia parasitica et la lutte biologique contre le chancre du châtaignier*. PhD Dissertation, Université de Bretagne Occidentale, Brest, France.

Hämmerli A, Reusch TBH (2003) Genetic neighbourhood of clone structures in eelgrass meadows quantified by spatial autocorrelation of microsatellite markers. *Heredity*, **91**, 448–455.

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, **83**, 145–154.

Hardy OJ, Vekemans X (2002) SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Hardy OJ, Maggia L, Bandou E *et al*. (2006) Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology*, **15**, 559–571.

Heald FD, Gardner MW, Studhalter RA (1915) Air and wind dissemination of ascospores of the chestnut blight fungus. *Journal of Agricultural Research*, **3**, 493–526.

Heiniger U, Rigling D (1994) Biological control of chestnut blight in Europe. *Annual Review of Phytopathology*, **32**, 581–599.

Hoegger PJ, Rigling D, Holdenrieder O, Heiniger U (2000) Genetic structure of newly established populations of *Cryphonectria parasitica*. *Mycological Research*, **104**, 1108–1116.

Isaaks E, Srivastava R (1989) *Applied Geostatistics*. Oxford University Press, Oxford, UK.

Kubisiak TL, Dutech C, Milgroom MG (2007) Fifty-three polymorphic microsatellite loci in the chestnut blight fungus, *Cryphonectria parasitica*. *Molecular Ecology Notes*, **7**, 428–432.

Legendre P, Legendre L (1998) *Numerical Ecology*. Elsevier, Amsterdam, The Netherlands.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of tropical understorey shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Manly B (1997) *Randomization and Monte Carlo Methods in Biology*. Chapman & Hall, Boca Raton, Florida.

Marra RE, Cortesi P, Bissegger M, Milgroom MG (2004) Mixed mating in natural populations of the chestnut blight fungus, *Cryphonectria parasitica*. *Heredity*, **93**, 189–195.

Marra RE, Milgroom MG (2001) The mating system of the fungus *Cryphonectria parasitica*: selfing and self-incompatibility. *Heredity*, **86**, 134–143.

Maynard Smith J, Smith NH, O'rourke M, Spratt BG (1993) How clonal are bacteria? *Proceedings of the National Academy of Sciences, USA*, **90**, 4384–4388.

McGuire IC, Marra RE, Milgroom MG (2004) Mating-type heterokaryosis and selfing in *Cryphonectria parasitica*. *Fungal Genetics and Biology*, **41**, 521–533.

McGuire IC, Davis JE, Double ML *et al*. (2005) Heterokaryon formation and parasexual recombination between vegetatively incompatible lineages in a population of the chestnut blight fungus, *Cryphonectria parasitica*. *Molecular Ecology*, **14**, 3657–3669.

Milgroom MG, Lipari SE (1995) Spatial analysis of nuclear and mitochondrial RFLP genotypes in populations of the chestnut blight fungus, *Cryphonectria parasitica*. *Molecular Ecology*, **4**, 633–642.

Milgroom MG, Lipari S, Wang KR (1992) Comparison of genetic diversity in the chestnut blight fungus, *Cryphonectria (Endothia) parasitica*, from China and the United States. *Mycological Research*, **96**, 1114–1120.

Nathan R, Perry G, Cronin JT, Strand AE, Cain ML (2003) Methods for estimating long-distance dispersal. *Oikos*, **103**, 261–273.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nuss DL (1992) Biological control of chestnut blight: an example of virus-mediated attenuation of fungal pathogenesis. *Microbiological Reviews*, **56**, 561–576.

Nuss DL (2005) Hypovirulence: mycoviruses at the fungal-plant interface. *Nature Reviews Microbiology*, **3**, 632–642.

Pimentel D, McNair S, Janecka J *et al*. (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture Ecosystems and Environment*, **84**, 1–20.

Reusch TBH, Hukriede W, Stam WT, Olsen JL (1999) Differentiating between clonal growth and limited gene flow using spatial autocorrelation of microsatellites. *Heredity*, **83**, 120–126.

Ribeiro PJ, Diggle PJ (2001) geoR: a package for geostatistical analysis. *R-NEWS*, **1**, 14–18.

Rizwana R, Powell WA (1995) Ultraviolet light-induced heterokaryon formation and parasexuality in *Cryphonectria parasitica*. *Experimental Mycology*, **19**, 48–60.

Robin C, Heiniger U (2001) Chestnut blight in Europe: diversity of *Cryphonectria parasitica*, hypovirulence and biocontrol. *Forest Snow and Landscape Research*, **76**, 361–367.

Robin C, Anziani C, Cortesi P (2000) Relationship between biological control, incidence of hypovirulence, and diversity of vegetative compatibility types of *Cryphonectria parasitica* in France. *Phytopathology*, **90**, 730–737.

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.

Russin JS, Shain L, Nordin GL (1984) Insects as carriers of virulent and cytoplasmic hypovirulent isolates of the chestnut blight fungus. *Journal of Economic Entomology*, **77**, 838–846.

Slatkin M (1985) Gene flow in natural populations. *Annuals Review of Ecology and Systematics*, **16**, 393–430.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, **82**, 561–573.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 931–935.

Wagner HH, Holderegger R, Werth S *et al*. (2005) Variogram analysis of the spatial genetic structure of continuous populations using multilocus microsatellite data. *Genetics*, **169**, 1739–1752.

Wang XR, Ennos RA, Szmidt AE, Hansson P (1997) Genetic variability in the canker pathogen fungus, *Gremmeniella abietina*. 2. Fine scale investigation of the population genetic structure. *Canadian Journal of Botany-Revue Canadienne de Botanique*, **75**, 1460–1469.

Werth S, Wagner HH, Holderegger R, Kalwij JM, Scheidegger C (2006) Effect of disturbances on the genetic diversity of an old forest-associated lichen. *Molecular Ecology*, **15**, 911–921.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Yonezawa K, Ishii T, Nagamine T (2004) The effective size of mixed sexually and asexually reproducing populations. *Genetics*, **166**, 1529–1539.

Cyril Dutech's research interests focus on the evolutionary processes that shape the genetic diversity of natural fungal pathogen populations, particularly dispersal processes and modes of reproduction. Jean-Pierre Rossi is an ecologist with a particular interest in landscape and spatial ecology. Olivier Fabreguettes is laboratory assistant, and in charge of the microsatellite analyses. Cécile Robin is a forest pathologist with interests in the biology and ecology of the chestnut blight fungus and its interaction with the Cryphonectria Hypovirus.
