npg

ORIGINAL ARTICLE

Effect of balancing selection on spatial genetic structure within populations: theoretical investigations on the self-incompatibility locus and empirical studies in *Arabidopsis halleri*

J-B Leducq[1,3], V Llaurens[1,3], V Castric[1], P Saumitou-Laprade[1], OJ Hardy[2] and X Vekemans[1]

[1]Laboratoire de Génétique et Évolution des Populations Végétales, UMR CNRS 8016, Université des Sciences et Technologies de Lille—Lille1, Villeneuve d'Ascq, France and [2]Service Evolution Biologique & Ecologie, CP 160/12, Faculté des Sciences, Université Libre de Bruxelles, Brussels, Belgium

The effect of selection on patterns of genetic structure within and between populations may be studied by contrasting observed patterns at the genes targeted by selection with those of unlinked neutral marker loci. Local directional selection on target genes will produce stronger population genetic structure than at neutral loci, whereas the reverse is expected for balancing selection. However, theoretical predictions on the intensity of this signal under precise models of balancing selection are still lacking. Using negative frequency-dependent selection acting on self-incompatibility systems in plants as a model of balancing selection, we investigated the effect of such selection on patterns of spatial genetic structure within a continuous population. Using numerical simulations, we tested the effect of the type of self-incompatibility system, the number of alleles at the self-incompatibility locus and the dominance interactions among them, the extent of gene dispersal, and the immigration rate on spatial genetic structure at the selected locus and at unlinked neutral loci. We confirm that frequency-dependent selection is expected to reduce the extent of spatial genetic structure as compared to neutral loci, particularly in situations with low number of alleles at the self-incompatibility locus, high frequency of codominant interactions among alleles, restricted gene dispersal and restricted immigration from outside populations. Hence the signature of selection on spatial genetic structure is expected to vary across species and populations, and we show that empirical data from the literature as well as data reported here on three natural populations of the herb *Arabidopsis halleri* confirm these theoretical results.

Heredity (2011) **106,** 319–329; doi:10.1038/hdy.2010.68; published online 9 June 2010

Introduction

Natural selection is believed to have a strong impact on patterns of SGS occurring within wild species. Many theoretical as well as empirical studies have assessed the effect of different forms of natural selection on such patterns. In plants, strong genetic structure is often encountered not only among natural populations (Hamrick and Godt, 1996), but also more locally within populations (Heywood, 1991; Vekemans and Hardy, 2004), as a consequence of restricted dispersal (for seeds, pollen or both). A common empirical approach used to detect the effect of natural selection on patterns of population genetic structure is to compare estimates of differentiation for a set of supposedly neutral 'control' loci and quantitative traits that may be under selection:

the F_{ST}/Q_{ST} comparison methods among populations (reviewed by Leinonen *et al.*, 2008) and spatial autocorrelation methods within populations (Sokal *et al.*, 1989). For traits whose genetic determinants have been identified, patterns of genetic structure can be directly compared at the gene level between control and selected loci. The expected patterns of SGS at target loci will strongly depend on the type of selection involved. Increased differentiation at selected loci has been documented associated with local adaptation, as for instance, the polymorphism of the lactase-persistence gene in humans (Burger *et al.*, 2007). In contrast, balancing selection is expected to reduce the extent of genetic differentiation among populations (Charlesworth *et al.*, 1997; Schierup *et al.*, 2000; Muirhead, 2001).

Multiallelic systems controlling self-incompatibility (SI) in plants are excellent model systems to characterize the effect of balancing selection on SGS. Indeed, plant SI systems are subject to strong negative frequency-dependent selection, a classical example of balancing selection, and have been the subject of comprehensive theoretical investigations as well as detailed molecular and genetic characterization in model plant families. SI is a widespread genetic system that prevents selfing in

Correspondence: *Professor X Vekemans, Laboratoire de Génétique et Évolution des Populations Végétales, UMR CNRS 8016, Université des Sciences et Technologies de Lille–Lille 1, Bat SN2–Cité Scientifique, Villeneuve d'Ascq 59655, France.*
E-mail: xavier.vekemans@univ-lille1.fr
[3]These authors contributed equally to this work.

hermaphroditic plant species through recognition between co-adapted proteins carried by pollen and pistils. SI phenotypes are generally controlled by a single multi-allelic genetic factor, the *S*-locus, which encodes pollen and pistil proteins. Two major SI systems have been identified: gametophytic SI (GSI), in which pollen phenotypes are determined by the haploid pollen genotype, and sporophytic SI (SSI), in which they are determined by the diploid paternal genotype through proteins expressed in the anther tissues, with dominance interactions often occurring among alleles at the *S*-locus (Takayama and Isogai, 2005). Negative frequency-dependent selection occurs at the *S*-locus because pollen carrying a rare allele has a higher chance of landing on a compatible pistil, and thus has higher reproductive success than pollen carrying a common allele (Wright, 1939). Theoretical and empirical studies on the genetic structure at the *S*-locus have mostly focused on the interaction between balancing selection and migration in subdivided populations. It has been shown theoretically that negative frequency-dependent selection will cause migrants introducing an allele at the *S*-locus (*S*-allele) that is absent or rare in the receiving population to be favored over resident individuals. This phenomenon causes a higher 'effective migration rate' (Barton and Bengtsson, 1986) at the *S*-locus as compared to neutral genes. A low population genetic structure at the *S*-locus is thus expected, even under very restricted migration (Schierup et al., 2000; Muirhead, 2001). These predictions have been tested in natural populations, and the results showed that the *S*-locus has indeed a lower genetic differentiation among populations than neutral loci (Glémin et al., 2005; Schierup et al., 2008; Stoeckel et al., 2008).

Whether this higher rate of effective dispersal for *S*-alleles could also occur within continuous populations is still unclear. In continuous plant populations, restricted pollen and seed dispersal causes spatial genetic structure (SGS), that is, a decrease of genetic similarity between individuals with spatial distance (Hardy and Vekemans, 1999; Rousset, 2000). Many empirical studies in plant populations have investigated SGS at marker loci, notably through the application of spatial autocorrelation methods (Heywood, 1991). It has been shown that the extent of SGS is dependent on species' characteristics such as life cycle and mating system (Vekemans and Hardy, 2004). In particular, mating systems influence SGS through their effect on pollen dispersal (for instance, strong SGS in highly selfing species) and genetic drift (for example, lower effective density in selfing species). Hence, strictly outcrossing plant species are expected to present lower levels of SGS on average than partially selfing species, and this has been confirmed empirically (Vekemans and Hardy, 2004). If balancing selection is increasing the rate of effective dispersal at the *S*-locus, this would lead to an even lower extent of SGS at the *S*-locus as compared to unlinked neutral loci.

Three theoretical studies have explored models of GSI in a continuous population using simulations of a lattice model with restricted pollen and seed dispersal. Brooks et al. (1996) investigated the effect of genetic drift on the *S*-locus, measured as the among-allele variance of allele frequencies, under restricted pollen and seed dispersal as compared to a panmictic population. Under restricted dispersal, only a marginal increase in the variance of

allele frequencies was found, suggesting a low extent of SGS at the *S*-locus. Neuhauser (1999) investigated the rate of loss of alleles at the *S*-locus in a finite continuous population under restricted pollen dispersal as compared to random dispersal. Under restricted dispersal, an increase in rate of loss of alleles was detected, suggesting an increase in the rate of local genetic drift and hence a pattern of SGS at the *S*-locus. Recently, Cartwright (2009) estimated Wright's neighborhood size (Nb) at the *S*-locus and found an increase in Nb with increasing mean pollen dispersal distances, showing that the extent of SGS (inversely related to Nb) at the *S*-locus is sensitive to variation in dispersal within a continuous population. He also compared different mating systems (mixed-mating system vs strict outcrossing systems, such as physical separation of pollen and pistil, GSI or SSI systems) and found that (1) Nb at unlinked neutral loci under strict outcrossing systems was significantly higher than in a self-compatible population, but no difference was found between the different outcrossing systems; (2) Nb at the *S*-locus increased with increasing levels of dispersal and (3) Nb under SSI was lower when dominance occurs among alleles. However, conclusions from these theoretical studies do not allow detailed interpretations of studies comparing empirical patterns of SGS at the *S*-locus with those at a set of control marker loci because they did not compare explicitly patterns of SGS expected at the *S*-locus with those at unlinked neutral loci, and because key parameters influencing patterns of SGS at all loci (the rate of immigration) or specifically at the *S*-locus (the number of alleles, which determines the strength of selection) were not investigated in these studies. Two empirical studies investigated SGS within populations for both the *S*-locus and neutral loci. In a population of *Senecio squalidus*, Brennan et al. (2003) found no significant decrease of genetic similarity between individuals with spatial distance, at neither the *S*-locus nor allozyme markers. In contrast, in a population of *Prunus avium*, Schueler et al. (2006) found evidence for SGS at both the *S*-locus and microsatellite markers, with a similar extent of SGS for both types of loci. These studies concern two highly different plant species in terms of life history, type of SI system and level of allelic diversity at the *S*-locus (Senecio is an herbaceous plant with SSI and low allelic diversity; *P. avium* is a tree with GSI and high allelic diversity). Hence, theoretical studies aimed at assessing the effect of balancing selection on SGS in a continuous population should model a wide range of dispersal rates, different SI systems with different levels of allelic diversity to allow meaningful comparisons with empirical studies.

In this study, we investigated patterns of SGS for the *S*-locus as well as unlinked neutral loci in theoretical models of SI. We first performed numerical simulations to characterize the extent of SGS at the *S*-locus and at unlinked neutral loci with two main objectives: (1) to estimate the effect of frequency-dependent selection, induced by either GSI or SSI systems, on the SGS at the *S*-locus as compared to unlinked neutral loci; and (2) to determine the effects of the extent of pollen and seed dispersal, the number of alleles at the *S*-locus and the rate of gene flow into the population (immigration), on the difference in patterns of SGS between the *S*-locus and neutral loci. We also studied empirically patterns of SGS within three natural populations of *Arabidopsis halleri*

(Brassicaceae), a self-incompatible species under SSI control (Llaurens *et al.*, 2008b). We estimated the extent of SGS at 11 unlinked microsatellite loci as well as at the *S*-locus, and investigated the signature of balancing selection by comparing these empirical results with our theoretical predictions.

Materials and methods

Numerical simulations
We simulated a finite continuous population of N hermaphroditic diploid individuals under restricted pollen and seed dispersal according to three SI systems: (1) selfing and crosses between genetically incompatible individuals avoided through a GSI system (GSI model); (2) same as (1) but with an SSI system with codominance among all *S*-alleles (SSI-COD model, according to Schierup *et al.*, 1997) or (3) same as (1) but with an SSI system with strictly hierarchical dominance among *S*-alleles (SSI-DOM model). Individuals were simulated with an *S*-locus as well as 10 unlinked neutral loci. All simulations were performed with a population size of 2500. Simulated plants were uniformly distributed on a 50×50 spatial lattice, with a single individual per node. As starting conditions, the genotype of each individual was randomly chosen for all loci. For each neutral locus, the initial number of alleles was arbitrarily set to six, which is close to the average number of alleles in the overall data set from *A. halleri*. Because the simulations are performed until equilibrium between drift and immigration (see below) and with a large number of replicates, the number of alleles at neutral loci is not expected to influence patterns of SGS. For the *S*-locus we used a range of values for the initial number of *S*-alleles n_S, from the minimum possible value under strict SI (2 for SSI-DOM, 3 for GSI, 4 for SSI-COD; Bateman, 1952) to 100. We used $n_S = 25$ (a value expected under GSI at drift–mutation–selection equilibrium for a population of 2500 individuals under low mutation rates, as computed with the formulas derived by Yokoyama and Hetherington (1982)) when investigating the effect of variation in other parameters. We assumed that the dispersal of seeds and pollen followed an isotropic bivariate normal distribution (that is, normal distribution in a two-dimensional space with equal dispersal in all directions), with standard deviation σ_s and σ_p (σ represents the axial standard deviation of dispersal distances in units of grid steps, that is, the standard deviation of dispersal distances measured around a zero mean and relative to a single reference axis passing through the population). The axial variance of overall gene dispersal can be computed as $\sigma_t^2 = \sigma_p^2/2 + \sigma_s^2$ (Crawford, 1984). We simulated a range of values of σ_s (1–5) and σ_p (1–8) to get a range of σ_t values between its minimal value ($\sigma_t = 1.225$ obtained when $\sigma_s = \sigma_p = 1$) to 6. We used the minimum value of σ_t when investigating the effect of variation in other parameters. Mutation was not considered in simulations. The long-term effect of genetic drift on allelic frequencies was counteracted by introducing random immigration of pollen from a constant source of genotypes identical to the initial population, at rate m_p in the range 1.10^{-5} to 0.6 (corresponding to Nm in the range 0.0125 to 750 with $m = m_p/2 =$ diploid rate of migration). We used $m_p = 0.0006$ (Nm = 0.75, corresponding

to the observed level of population genetic structure based on a survey of microsatellite data in 65 populations covering the whole geographical distribution of *A. halleri*, $F_{ST} \approx 0.25$; M Pauwels, unpublished results) when investigating the effect of variation in other parameters. To determine genotypes of each individual i across the grid at generation t, we used a forward simulation algorithm that searched potential parents of i in the grid at generation $t-1$. The mother j was chosen in three steps: (1) a direction was randomly determined by drawing an angle α_{ij} from a uniform distribution; (2) the distance between the mother and the offspring d_{ij} was randomly drawn from the normal distribution of seed dispersal distances (with parameter σ_s); (3) the mother j was chosen as the nearest individual from the point given by the coordinate (α_{ij}; d_{ij}). The father was an immigrant randomly drawn from a constant source of genotypes identical to the initial population with probability m_p. With probability $(1-m_p)$, the father was chosen with the previously described method by using the mother j as the focal individual. Hence, the father, k, was chosen as the individual closest to the point (α_{jk}; d_{jk}), with d_{jk} randomly drawn from the normal distribution of pollen dispersal distances (with parameter σ_p). If j or k were outside the grid, new coordinates were obtained by repeating the process. The case $j = k$ (selfing) was rejected in all models. In all other cases, the phenotypes of j and k at the *S*-locus were checked for compatibility. For GSI, pollen with only one out of the two paternal *S*-alleles was randomly chosen, and then j and k were considered compatible when the *S*-allele of the pollen was different from the two *S*-alleles of j. For SSI-COD, j and k were compatible only when the two *S*-alleles of k were different from the two *S*-alleles of j. For SSI-DOM, we determined the SI phenotype of j and k taking into account the dominance interactions of their *S*-alleles. In SSI-DOM, the dominance scheme was strictly hierarchical, so the phenotype in heterozygotes corresponded to that of the more dominant allele. In case of incompatibility between j and k, a new potential father k was chosen with a probability m_p in the immigrant gene pool or $1-m_p$ in the grid, and the process was repeated until a compatible combination was found. This compatibility-checking algorithm thus simulated the negative frequency-dependent selection described by Wright (1939). When parents of i were defined, one allele at each locus from each parent was randomly taken to constitute the genotype of i. The process was repeated for each position in the grid, and then all individuals of generation $t-1$ were replaced at once by genotypes of generation t. Each simulation run was stopped after a predefined number of generations (10 000), which was chosen to allow stabilization of the statistics of interest (see below) at drift–migration equilibrium (data not shown).

At the end of each run, we analyzed the SGS on 1600 individuals located in a 40×40 grid centered on the initial grid, a procedure often used to minimize edge effects (Heuertz *et al.*, 2003). For each locus we computed the matrix of pairwise kinship coefficients among all pairs of individuals (F_r), using J Nason's estimator (Loiselle *et al.*, 1995). We computed the slope (b) of the linear regression of F_r values as a function of the logarithm of spatial distance between individuals. The absolute value of statistic b is a measure of the extent of SGS. It is inversely proportional to the overall axial

standard deviation of dispersal distances and is only weakly influenced by the shape of pollen and seed dispersal distributions (Hardy and Vekemans, 1999; Rousset, 2000; Heuertz *et al.*, 2003). To visualize SGS, we produced spatial autocorrelograms by plotting mean kinship coefficients, $F_r(d)$ (computed among pairs of individuals classified according to 40 nonoverlapping intervals of interindividual distance d) on the y axis and distance d expressed in number of steps along the grid on the x axis. The simulation program adapted from Hardy and Vekemans (1999) was written in C language.

Species studied
A. halleri (Brassicaceae) is a diploid herbaceous plant characterized by strong tolerance to heavy metals like zinc and cadmium. It is a perennial species that survives in winter as a rosette and is able to develop stolons for asexual reproduction. *A. halleri* has a functional SI system with sporophytic control of pollen phenotype (Llaurens *et al.*, 2008b).

Study sites and sampling
Our empirical study focused on three populations of *A. halleri* located in Hautes-Fagnes (Belgium, 50°29′63′N, 6°05′00′E), Auby (France, 50°24′14′N, 3°05′04′E) and Nivelle (France, 50°28′13′N, 3°28′06′E). The sampling strategies were designed so as to allow pairwise comparisons of individuals at different spatial scales (Supplementary Figure S1). The Hautes-Fagnes population grows in a site that has been recently colonized (Pauwels *et al.*, 2005). We sampled 66 individuals at regular intervals along a 400 m transect and 68 and 28 individuals, respectively, within two plots of a 0.5×3 m area. The Auby population is located in a site with high zinc content, due to soil pollution from the local mining industry (Van Rossum *et al.*, 2004). In this very large population, we sampled 134 individuals along a 400 m transect and 76 individuals within a plot of a 0.5×3 m area (see Van Rossum *et al.*, 2004 for a description of the sampling procedure). In the population of Nivelle, we sampled exhaustively all individuals in about half of the population and in the other half we sampled individuals along a 30 m transect (364 individuals in total; see Llaurens *et al.*, 2008a). In each population, spatial coordinates of all sampled individuals were collected.

Genotyping of microsatellite markers and the S-locus
For the Nivelle population, genotypic data for 11 microsatellite loci (*ATH*, *ELF3*, *GC16*, *LYR133*, *LYR417*, *GC22*, *H117*, *ICE13*, *MDC16*, *NGA112* and *NGA361*) were taken from Llaurens *et al.* (2008a). For the Auby population, DNA samples and genotypic data from five microsatellite loci (*LYR132*, *LYR133*, *LYR417*, *GC16* and *ATH*) were already available (Van Rossum *et al.*, 2004). We genotyped six additional microsatellite loci (*GC22*, *H117*, *ICE13*, *MDC16*, *NGA112* and *NGA361*) using a multiplex PCR procedure, as described in Llaurens *et al.* (2008a). For the Hautes-Fagnes population, leaves taken from each sampled individual were dried at 55 °C for 24 h. DNA was extracted from 10–15 mg of dried leaf material using the extraction kits Dneasy from Qiagen (Courtaboeuf, France). We genotyped 11 microsatellite loci (*GC22*, *H117*, *ICE113*, *MDC16*, *NGA112*, *NGA361*, *LYR132*, *LYR133*, *GC16*, *LYR104* and *ICE9*). For the first 6

markers, *GC22*, *H117*, *ICE113*, *MDC16*, *NGA112*, *NGA361*, we used the same multiplex PCR procedure as in Auby. For *LYR132*, *LYR133* and *GC16*, we used primers described in Van Rossum *et al.* (2004), for *ICE9*, those described in Clauss *et al.* (2002), and for *LYR104*, we used the following primers (courtesy of Thomas Mitchell-Olds: forward primer, GAGGCGAATGTAGTGGAAGG; reverse primer, CGACCTCCATCATCGATCTCAGCA). The reaction mixture (15 μl) contained 20 ng DNA, $1 \times$ buffer (Applied Biosystems), 2 mM of MgCl₂, 200 μM of Fermentas dNTP mix, 200 μg ml⁻¹ of bovine serum albumin, 0.2 μM of each microsatellite primer, 0.15 μM of M13 primer (fluorescence-labeled with either IRD-700 or IRD-800) and 0.025 U μl⁻¹ of *Taq* polymerase (Amplitaq DNA polymerase; Applied Biosystems, Courtaboeuf, France). The amplification was carried out 5 min at 95 °C, 8 cycles of 30 s at 95 °C, 45 s at 50 °C, 40 s at 72 °C then 30 cycles of 30 s at 95 °C, 20 s at 50 °C, 40 s at 72 °C, and one cycle of 7 min at 72 °C and performed in MJ Research PTC-200 thermocycler (Marnes-la-Coquette, France).

PCR products were separated on 6% polyacrylamide gels and visualized through fluorescence of M13 primers on a Li-Cor sequencer (Les Ulis, France). Size standards were run to allow accurate band sizing.

To determine genotypes at the *S*-locus, we used a collection of specific PCR primers that amplify each of the 26 alleles specifically (*AhSRK01* to *AhSRK26*) at the *SRK* gene previously identified in *A. halleri* see (Llaurens *et al.*, 2008b for primer sequences and PCR conditions). PCR products were mixed with loading dye and run at 110 V on 2% agarose gels in Tris-Borate-EDTA buffer for 45 min. Fragments, including a positive PCR control, were fluorescently labeled by ethidium bromide and visualized under UV light. We found a new allele (*AhSRK28*) that co-amplified with allele *AhSRK03*. To distinguish both alleles, we digested the PCR products with the restriction enzyme *Taq*I (which cuts *AhSRK28* but not *AhSRK03*) before electrophoresis. This *S*-locus genotyping method generated genotypes with 0, 1 or 2 *S*-alleles (Supplementary Table S1). This is because some alleles have not yet been identified due to their high nucleotide sequence divergence. Individuals with no *S*-allele detected were discarded. Individuals with a single *S*-allele were analyzed either by assuming that they are homozygotes, or that they are heterozygotes with a single unknown allele. Both assumptions gave very similar results (data not shown) and we thus report only results using the former.

Data analysis
Individuals of *A. halleri* may reproduce vegetatively through stolons (Van Rossum *et al.*, 2004; Llaurens *et al.*, 2008a). To avoid bias in estimates of SGS in relation to clonal reproduction, we considered ramets with identical multilocus genotypes at microsatellite loci as belonging to the same genet (noting that the distance between two identical multilocus genotypes never exceeded 1 m) and we discarded pairwise comparisons between those ramets. SGS within each population was analyzed with the program SPAGeDi (Hardy and Vekemans, 2002). Multilocus kinship coefficients for microsatellite loci, and a single locus coefficient for the *S*-locus, were computed between all pairs of individuals using J Nason's estimator of kinship (Loiselle *et al.*, 1995). Different genets were coded as different 'categories' and computations

of kinship coefficients were performed among categories only. The occurrence of SGS was tested by a Mantel test between the matrix of pairwise kinship coefficients and that of the logarithm of pairwise spatial distances between individuals, using 1000 random permutations of spatial locations among individuals. The extent of SGS was characterized by the slope b of the regression of pairwise kinship coefficients on the logarithm of pairwise spatial distances. For microsatellite loci, the mean and standard error of b were estimated using a jackknife procedure over loci. We compared this mean with the value of b obtained for the S-locus using a t-test. To compare the extent of SGS at microsatellite loci between populations and between *A. halleri* and other species, we also computed the S_p statistic for microsatellite loci according to $S_p = -b/[1-F_r(1)]$ where $F_r(1)$ is the mean kinship coefficient between individuals belonging to a first distance interval that should include all pairs of neighbors (Vekemans and Hardy, 2004). This statistic was proposed as a mean to compare quantitatively the SGS between populations and species, and corresponds under certain conditions to the reciprocal of Wright's neighborhood size. For visualization purpose, we plotted autocorrelograms with mean pairwise kinship coefficients computed for each of 20 nonoverlapping intervals of interindividual distances.

Results

Expected spatial genetic structure at neutral loci and the *S*-locus

As expected under the isolation by distance model under highly restricted dispersal, an approximately linear decrease in the average multilocus kinship coefficient $F_r(d)$ with the logarithm of distance d was observed for the unlinked neutral loci (Figure 1). The comparison between patterns of SGS at unlinked neutral loci and the S-locus revealed a strong effect of selection, with a strikingly weaker slope for the autocorrelograms of the S-locus as compared to neutral loci (Figure 1). The type of SI system was found to slightly influence the slope of the autocorrelograms at unlinked neutral loci, with a weaker slope for SSI-COD as compared to GSI models, and an intermediate slope for the SSI-DOM model. Differences among models of SI were more pronounced at the S-locus than for neutral loci, with a substantially weaker slope for GSI and SSI-COD, as compared to the SSI-DOM model. Frequency-dependent selection also affected the shape of the autocorrelograms: although kinship coefficients decreased nearly linearly with the logarithm of distance for neutral loci (at least for distances higher than two step units), it approached an asymptote close to the x axis for the S-locus, leading to a concave autocorrelogram (Figure 1).

The effect of gene dispersal distances on the extent of SGS (summarized by the statistic b, the slope of the linear regression of pairwise kinship coefficients as a function of the logarithm of spatial distance) at unlinked neutral loci and at the S-locus is shown in Figure 2. Under highly restricted dispersal ($\sigma_t < 2$), the slope b was about four times more negative for neutral loci as compared with the S-locus. However, b values for neutral loci and the S-locus became very similar for $\sigma_t > 4$. The model of SI had a very small effect on patterns of SGS at neutral loci, whereas the slope b for the S-locus was about 1.5 times higher under SSI-DOM than under both GSI and SSI-COD models. We also tested the effect of varying the relative pollen and seed dispersal distances (σ_p and σ_s, respectively), under constant overall gene dispersal (σ_t), and found that all combinations showed very similar results both for unlinked neutral loci and for the S-locus (Supplementary Figure S2).

We investigated the effect of the number of S-alleles (n_S) on the patterns of SGS, for a case with highly restricted dispersal ($\sigma_t = 1.225$; Figure 3). The number of
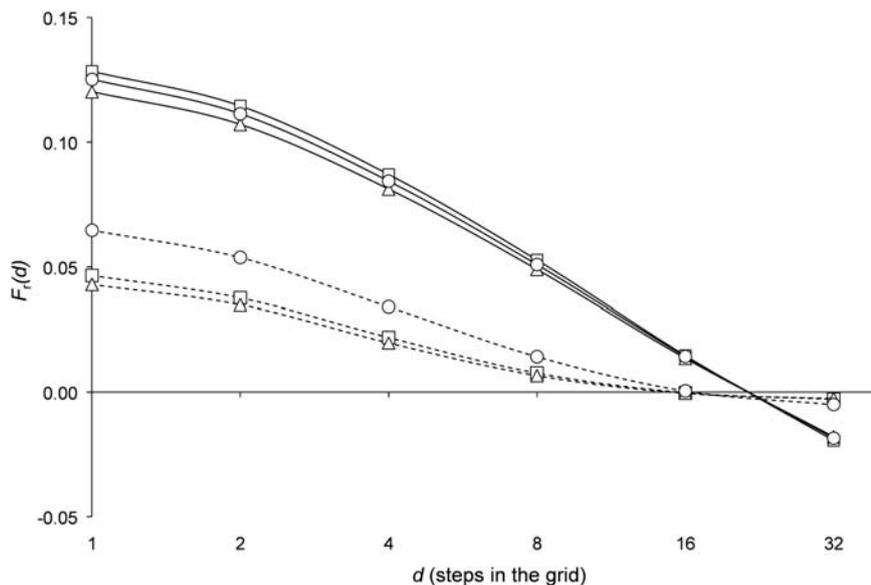

Figure 1 Spatial autocorrelograms for unlinked marker loci (mean over 10 loci; solid lines) and for the S-locus (broken lines) under three mating system models (gametophytic SI: (squares); sporophytic SI with codominance, SSI-COD: (triangles); sporophytic SI with dominance, SSI-DOM (circles)). Average results from 100 replicate simulations after 10 000 generations under a lattice model with $N = 2500$ individuals, $\sigma_t = 1.225$ ($\sigma_s = \sigma_p = 1$), $Nm = 1.5$ and 25 alleles at the S-locus. The y axis represents the mean kinship coefficient between individuals $F_r(d)$. The x axis represents the spatial distance intervals between individuals, d in units of grid steps (log scale).
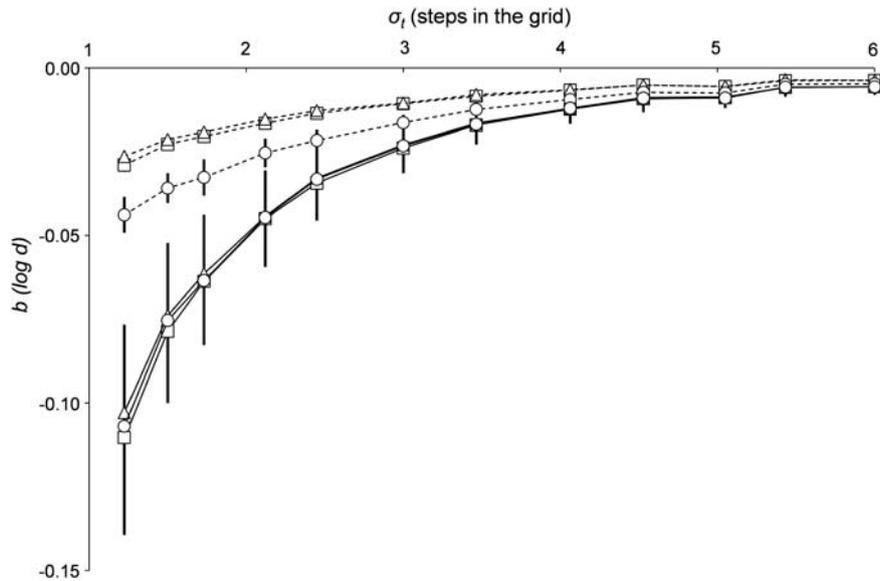
Figure 2 Effect of the variance of overall gene dispersal σ_t (in units of grid steps) on the extent of SGS measured by the slope b of kinship–distance curves for unlinked marker loci (mean over 10 loci; solid lines) and for the S-locus (broken lines) under three models of SI. Results are from 100 replicate simulations (error bars are standard deviations) after 10 000 generations under a lattice model with $N = 2500$ individuals, Nm = 1.5 and 25 alleles at the S-locus. The y axis represents the mean overall slope b of kinship–distance curves according to the logarithm of spatial distance. SI systems: gametophytic SI (squares); sporophytic SI with codominance, SSI-COD (triangles); sporophytic SI with dominance, SSI-DOM (circles).
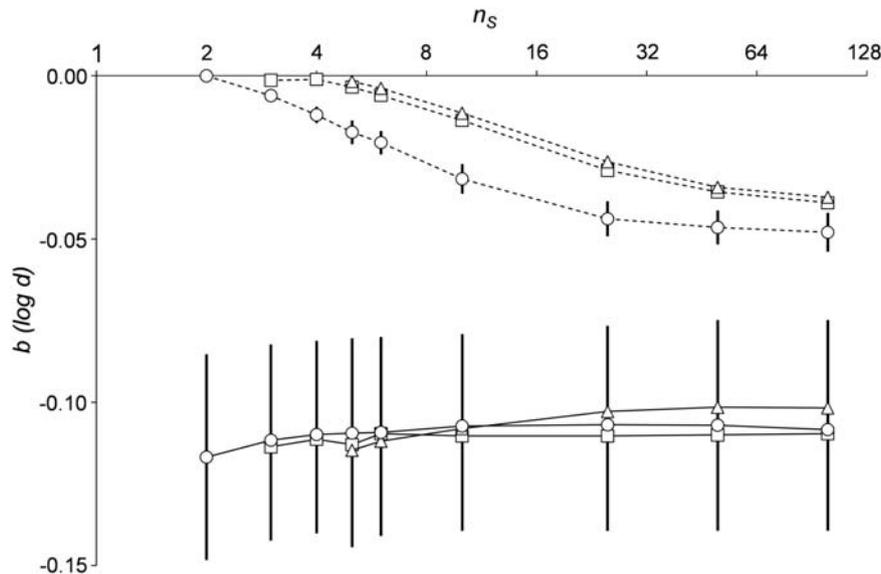

Figure 3 Effect of the number of alleles at the S-locus n_S, on the extent of SGS measured by the slope b of kinship–distance curves for unlinked marker loci (mean over 10 loci; solid lines) and for the S-locus (broken lines) under three models of SI. Results from 100 replicate simulations (error bars are standard deviations) after 10 000 generations under a lattice model with $N = 2500$ individuals, Nm = 1.5 and $\sigma_t = 1.225$. The y axis represents the mean overall slope b of kinship–distance curves according to the logarithm of spatial distance. SI systems: gametophytic SI (squares); sporophytic SI with codominance, SSI-COD (triangles); sporophytic SI with dominance, SSI-DOM (circles).

S-alleles did not influence noticeably the extent of SGS at the unlinked neutral loci, whereas the effect on the S-locus itself was substantial for all SI models. Large standard deviations of the statistic b for neutral loci are due to the high stochasticity generated by genetic drift under highly restricted dispersal distances. For the S-locus, the strength of selection increases when the number of S-alleles decreases, so that stochastic effects are strongly reduced. Despite highly restricted dispersal, an isolation by distance pattern was not or barely

detected at the S-locus (that is, the slope b was close to zero) when the number of alleles was close to its theoretical minimum (two for SSI-DOM, three for GSI and four for SSI-COD; Bateman, 1952). When the number of S-alleles increased, the extent of SGS at the S-locus increased (that is, values of b became more negative) and asymptotically reached a level lower than that expected for the unlinked neutral loci. We also observed that the extent of SGS under the SSI-DOM model was consistently higher than under the GSI and SSI-COD models.
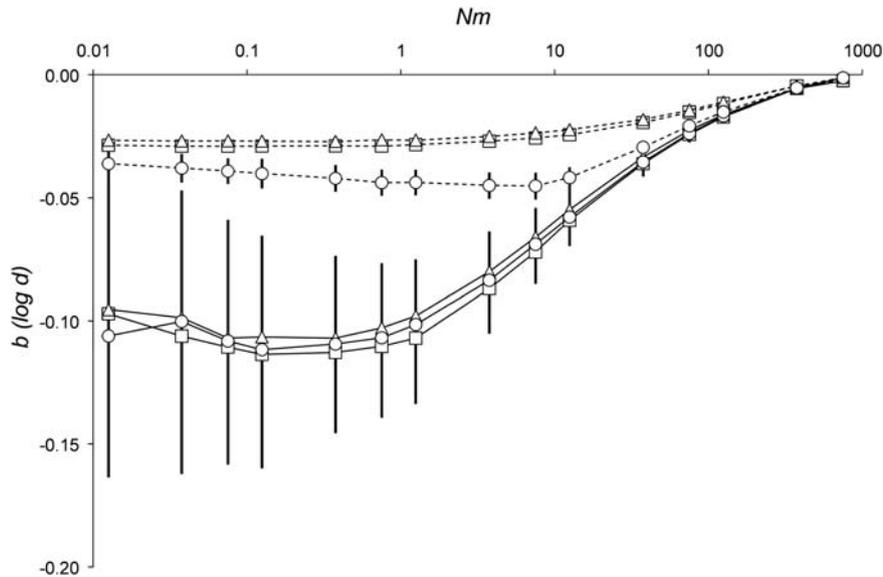
Figure 4 Effect of immigration rate Nm on the extent of SGS measured by the slope b of kinship–distance curves for unlinked marker loci (mean over 10 loci; solid lines) and for the S-locus (broken lines) under three models of SI. Results from 100 replicate simulations (error bars are standard deviations) after 10 000 generations under a lattice model with $N = 2500$ individuals, $\sigma_t = 1.225$; and n_S (number of alleles at the S-locus) = 25. The y axis represents the mean overall slope b of kinship–distance curves according to the logarithm of spatial distance. SI systems: gametophytic SI (squares); sporophytic SI with codominance, SSI-COD (triangles); sporophytic SI with dominance, SSI-DOM (circles).

The effect of the immigration rate (m) on the extent of SGS at unlinked neutral loci and at the S-locus is shown in Figure 4. For neutral loci, the extent of SGS was not much affected by changes in the rate of immigration below a value of Nm = 1. However, the standard deviation of the statistic b increased substantially under low immigration, probably in relation to the loss of diversity at neutral loci due to genetic drift (Supplementary Figure S3). Under SSI-DOM, low immigration (Nm <10) also led to a slight increase in the standard deviation of the statistic b for the S-locus (Figure 4), which resulted from a loss of the most dominant S-alleles under strong genetic drift (data not shown). Above Nm = 1.5, the extent of SGS for neutral loci decreased rapidly. For the S-locus, we observed a similar effect, with lower amplitude, as the slope b was substantially weaker than at neutral loci. Hence, under high rates of immigration, which is a homogenizing force, the difference in the extent of SGS between unlinked neutral loci and the S-locus was vanishing.

Patterns of SGS at microsatellite loci and at the S-locus in natural populations

The number of alleles per microsatellite locus per population varied from 2 to 10 with an average over loci ranging from 3.45 in Hautes-Fagnes to 4.91 in Auby (Table 1). Observed heterozygosities were also higher in Auby ($H_{obs} = 0.57$) than in Hautes-Fagnes (0.39) or in Nivelle (0.46), in agreement with the much larger census size of the Auby population. Clonality was substantial in the Auby population (with a ratio of the number of genets to the number of ramets equal to 0.82) but weak in Hautes-Fagnes and Nivelle, whereas local plant density was highest in the Nivelle population. The three populations showed a highly significant pattern of isolation by distance within population, as revealed by

the Mantel test between matrices of the pairwise multi-locus kinship coefficients and spatial distances (Table 1). The spatial autocorrelograms confirmed that the pairwise kinship coefficients decreased with spatial distance between individuals, at least in the range 0–3 m (Supplementary Figure S4). We thus restricted computation of the slope b of the autocorrelograms to the range 0–3 m. Estimates of the slope b ranged from −0.013 (Nivelle) to −0.099 (Hautes-Fagnes), giving values of the S_p statistic ranging from 0.014 (Nivelle) to 0.116 (Hautes-Fagnes). This statistic measures the extent of SGS at neutral loci, and the results thus indicate that patterns of genetic structure were strongest in Haute-Fagnes, intermediate in Auby and lowest in the Nivelle population.

For the S-locus, the number of S-alleles detected ranged from four (in Hautes-Fagnes) to nine (in Auby, Table 1). These numbers are minimum estimates as the molecular typing method does not ensure exhaustive recovery of S-alleles. However, the numbers of individuals showing a null genotype (not a single allele detected) were low (ranging from 0/162 to 27/322; Supplementary Table S1), suggesting that the bias was low. A significant relationship between pairwise kinship coefficients at the S-locus and the log of spatial distances was detected in Nivelle and Hautes-Fagnes, but not in Auby (Table 1). In the Nivelle population, the slope b computed in the range 0–3 m for the S-locus was similar to that observed for microsatellite loci ($b = -0.01$), whereas in Hautes-Fagnes the slope for the S-locus ($b = -0.03$) was strongly and significantly weaker (an almost fourfold decrease) as compared to neutral loci ($b = -0.1$). For Auby, in the absence of a detectable pattern of isolation by distance for neutral loci, the value of zero for slope b for the S-locus was found to be significantly different than that for neutral loci ($b = -0.06$). Hence the pattern of SGS differed between

Table 1 Estimates of genetic variation and patterns of SGS at microsatellite loci and the S-locus in three natural populations of *A. halleri*

	Nivelle	Auby	Hautes-Fagnes
Sample size			
Number of ramets	364	210	162
Number of genets[a]	356	173	148
D (individuals per m^2)	39.9	25.4	30.3
Microsatellite loci			
$A \pm$ s.d.	3.67 ± 1.67	4.91 ± 2.30	3.45 ± 1.44
$H_{obs} \pm$ s.d.	0.46 ± 0.18	0.57 ± 0.20	0.39 ± 0.16
Mantel test (P-value)[b]	<0.001	<0.001	<0.001
b	-0.0132 ± 0.0017	-0.0583 ± 0.0192	-0.0994 ± 0.0377
S_P	0.014	0.061	0.116
S-locus			
n_s [9]	8	9	4
H_{obs}	0.82	0.77	0.69
Mantel test (P-value)[b]	<0.001	NS	<0.01
b	-0.0113	0	-0.0265
Microsatellites vs S-locus t-test (P-value)[c]	0.150	0.006	0.042

Abbreviations: D, mean observed density of individuals; A, mean number of alleles per locus; H_{obs}, mean observed heterozygosity; b, slope of the regression of pairwise kinship coefficients on the log of spatial distances between individuals; S_P, statistic measuring the extent of SGS; n_s, number of alleles detected at the S-locus; NS, not significant.
[a]Number of distinct multilocus genotypes at microsatellite loci.
[b]Result of a Mantel test to detect a pattern of isolation by distance.
[c]Result of a t-test comparing the observed value of slope b for the S-locus with the average b for microsatellite loci.

the S-locus and microsatellite loci in two (Auby and Hautes-Fagnes) out of three populations with, as predicted by theory, a lower extent of SGS observed for the S-locus.

Discussion

Our theoretical results showed that frequency-dependent selection acting on the genes controlling gametophytic and SSI is expected to lead to a lower extent of SGS at the S-locus than at unlinked neutral loci within continuous populations. Previous studies had shown that under restricted dispersal, patterns of SGS would develop at the S-locus (Brooks et al., 1996; Neuhauser, 1999; Cartwright, 2009), but they did not explicitly compare those with SGS at unlinked neutral loci. Our study results thus constitute an extension of the theoretical results obtained in island models of subdivided populations, for which a substantially lower level of genetic differentiation was found among populations at the S-locus as compared with unlinked neutral loci (Schierup et al., 2000; Muirhead, 2001). These predictions were interpreted as resulting from an increase in the effective migration rate at the S-locus due to frequency-dependent selection. By analogy, one can interpret our results as due to an increase in effective dispersal distances at the S-locus within a continuous population, caused by selection. The results of our study also show that the difference in patterns of SGS between the S-locus and unlinked neutral loci depends strongly on population and species characteristics, such as the extent of pollen and seed dispersal, the degree of population isolation, the type of SI system and the allelic diversity at the S-locus. Hence, the signature of selection in terms of patterns of SGS is not expected to be detected in all populations or species, and thus the effect of population characteristics should be considered when interpreting empirical results from the literature (Brennan et al., 2003; Schueler et al., 2006), or from this study.

Influence of the self-incompatibility model on SGS at the S-locus

Beside a lower extent of SGS at the S-locus than at unlinked neutral loci under restricted dispersal, our simulations also showed that frequency-dependent selection affected the shape of the autocorrelogram at the S-locus, with a faster approach to its asymptote. This pattern is also expected under the action of other homogenizing forces such as mutation, migration (see below) or random dispersal.

For a given number of S-alleles, we observed that the impact of frequency-dependent selection was weaker for the SSI-DOM model than for GSI or SSI-COD. We suggest that this arises because the dominance relationships among S-alleles in SSI-DOM increase the number of compatible crosses in the population, thus decreasing the strength of frequency-dependent selection (Schierup et al., 1997). This suggestion is supported by the observation that patterns of SGS for individual alleles at the S-locus in SSI-DOM indicate stronger SGS for recessive alleles (subject to weaker selection) than for dominant alleles (Supplementary Figure S6). These quantitative differences among the three SI systems are in close agreement with those obtained for the analysis of population subdivision under SI (Schierup et al., 2000). In models of SI, the strength of frequency-dependent selection is strongly affected by the number of S-alleles segregating in the population, with stronger selection when the number of alleles is lower (Wright, 1939). Accordingly, we found that the extent of SGS at the S-locus was more strongly reduced as compared to neutral loci when the number of S-alleles was close to its minimum value. In those situations, virtually no signature of SGS is expected for the S-locus even in the case of very restricted gene dispersal (see also Neuhauser, 1999). Actually, for SSI systems with the theoretical minimum number of alleles, SGS cannot possibly occur at the S-locus because each individual's offspring are segregating for the two or four extent alleles in SSI-DOM and SSI-COD

systems, respectively. SSI with distyly is an example of an SSI-DOM model with only two S-alleles. Accordingly, Van Rossum and Triest (2006) found a high extent of SGS at neutral loci, but no such pattern at the *S*-locus in a population of *Primula elatior*. Similarly, in most populations of the tristylous *Eichhornia paniculata*, Husband and Barrett (1992) found no evidence for nonrandom spatial distribution of morphs.

The influence of the number of *S*-alleles on patterns of SGS at the *S*-locus has some consequences for conservation genetics issues. Indeed, a concern in managing very small plant populations with functional SI is that the number of *S*-alleles present in the population may be very small, such that any given individual may mostly receive pollen from individuals sharing the same alleles, hence reducing seed-set (Kirchner *et al.*, 2006; Busch and Schoen, 2008). This effect (called the 'S-Allee effect' by Wagenius *et al.*, 2007) was expected to be stronger under low pollen dispersal because of reproduction within local neighborhoods. However, our study results suggest that in populations with a low number of S-alleles, the *S*-locus would show only a very weak pattern of SGS, so that spatial structure is unlikely to further restrict the availability of compatible pollen. Hence allelic diversity at the *S*-locus, rather than SGS, appears to be a key factor for the persistence of endangered populations of self-incompatible species.

Influence of within-population dispersal and rate of immigration

Simulations showed an important decrease in the extent of SGS for both the *S*-locus and unlinked neutral loci with increasing local dispersal distances and increasing rate of immigration. For a given value of the standard deviation of overall gene dispersal, patterns of SGS at either the *S*-locus or the unlinked neutral loci were similar when varying the relative dispersal distances of pollen and seeds. This simulation result has been obtained previously for a neutral locus by Heuertz *et al.* (2003), and is a consequence of the general theoretical result by Rousset (2000), that the slope of the kinship–distance curves is only weakly influenced by the shape of gene dispersal distributions. Our study results indicate that this prediction holds for loci subject to frequency-dependent selection. Under high dispersal or immigration, both types of loci showed convergent patterns of SGS. One thus expects that a significant difference in the extent of SGS between neutral loci and the *S*-locus would only be observed in populations or species with restricted pollen and seed dispersal and restricted immigration, for example, in highly fragmented populations. This arises because frequency-dependent selection can be seen as a homogenizing force so that its effect will only be detectable under strong local genetic drift and in the absence of other homogenizing forces such as high immigration flow.

Influence of SI systems on SGS at unlinked neutral loci

Self-incompatibility systems are mostly known as outcrossing devices in hermaphroditic plant populations, but it is also generally assumed that they contribute to a reduction in overall levels of biparental inbreeding within populations (Bos and van der Haring, 1988). This is because they cause a restriction in mating between close relatives sharing identical genotypes at the *S*-locus. Cartwright (2009) addressed this issue by comparing patterns of SGS at neutral loci under restricted dispersal between models of a continuous population with an SI system vs a strict outcrossing system not involving mate limitation among individuals (for example, physical separation of pollen and pistil). He found no difference between the two systems, and concluded that SI may be considered essentially as a selfing avoidance system, and that restriction of biparental inbreeding is merely a side effect. However, he simulated only SI systems with a large number of alleles at the *S*-locus, which correspond to situations with weak frequency-dependent selection. Our study results showed that a decrease in the number of *S*-alleles does not promote a decrease in the extent of SGS at unlinked neutral loci (Figure 3). In contrast, the overall pattern seems opposite, that is, a higher extent of SGS (more negative slope of the autocorrelograms) when the number of *S*-alleles decreases, although the standard deviations of the slope estimates were high. We suggest that this is due to a stronger limitation in mate availability when the number of *S*-alleles is low (Vekemans *et al.*, 1998). Indeed, stronger limitation in mate availability also causes an increase in the variance of reproductive success among individuals (Supplementary Figure S5), which would cause a decrease in effective density influencing patterns of SGS at neutral loci. This is consistent with the observation that this effect is stronger for SSI-COD (Figure 3; Supplementary Figure S5), which constitutes the SSI model causing the strongest limitation in mate availability (Vekemans *et al.*, 1998). Our study results thus show that even under strong selection, the SI system does not contribute to a reduction in overall levels of biparental inbreeding within continuous populations. Thus, we do not expect that predominantly outcrossing species would show contrasted patterns of SGS at unlinked markers depending on the system enforcing outcrossing. This conclusion is consistent with the results from a literature survey on patterns of SGS in plant species as a function of the mating system, where very similar extents of SGS were observed in predominantly outcrossing species with ($S_P = 0.0134 \pm 0.0077$) or without ($S_P = 0.0126 \pm 0.0101$) SI systems (Vekemans and Hardy, 2004).

Interpretation of empirical results

Previous empirical studies, together with our own results, showed three contrasted types of situations: (1) no pattern of SGS whatsoever—neither at the *S*-locus nor at neutral marker loci (Brennan *et al.*, 2003); (2) significant and consistent patterns of SGS at the *S*-locus and neutral markers (Schueler *et al.*, 2006; population Nivelle in this study); (3) significantly higher extent of SGS at neutral marker loci as compared to the *S*-locus (populations Auby and Hautes-Fagnes in this study). Situation (1) could be due to the absence of SGS in the case of random dispersal, to recent demographic disturbance or to a lack of resolution of the analysis when sample sizes are small (for example, in the study of Brennan *et al.*, 2003, in *S. squalidus* only 24 individuals were sampled). According to our theoretical results, situation (2) would occur when the tail of the pollen and/or seed dispersal distributions are very fat, implying a significant fraction of nearly random dispersal, when the number of alleles at the *S*-locus is high and/or when the immigration rate is

high. The observation of this pattern in *P. avium* is consistent with the high number of *S*-alleles (15), the high level of pollen and seed dispersal and the high rate of immigration in this temperate tree species producing a low extent of SGS ($S_p = 0.0122$, Schueler *et al.*, 2006; $F_{ST} = 0.074$, Stoeckel *et al.*, 2008). In our study, we observed this pattern only in the population of Nivelle, which has the lowest extent of SGS at microsatellite loci ($S_p = 0.014$) as a result of higher gene dispersal and higher plant density than the other two populations. Situation (3) indicates that frequency-dependent selection is affecting the extent of SGS at the *S*-locus. We observed this pattern in two populations (Auby and Hautes-Fagnes), and a similar observation was reported for the species *A. lyrata* although the geographical scale of that study was larger (Schierup *et al.*, 2008). For the Hautes-Fagnes population, this effect is consistent with our theoretical results as this population shows highly restricted dispersal, a very low allelic diversity at the *S*-locus and a complete isolation. For the Auby population, the absence of a pattern of isolation by distance at the *S*-locus in contrast to a strong pattern of SGS at marker loci is more difficult to interpret. Because the number of *S*-alleles is much higher in Auby than in Hautes-Fagnes, one would have expected a weaker signature of frequency-dependent selection on patterns of SGS in the former. However, an important characteristic of the Auby population is the high level of clonal reproduction that shows a strong spatial structure (Supplementary Figure S1; Van Rossum *et al.*, 2004). One can suggest that this situation would generate high levels of incompatible geitonogamous pollination, which could increase the strength of frequency-dependent selection through an effect on female fitness (this effect, called fecundity selection, has been described by Vekemans *et al.*, 1998). Hence, putatively, the absence of SGS at the *S*-locus in Auby, as computed based on the distribution of genets, could result from a strong spatial structure at the level of ramets. Because clonal reproduction is often associated with SI in plants (Vallejo-Marin and O'Brien, 2007), this suggestion should be investigated in more detail.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

References

Barton NH, Bengtsson BO (1986). The barrier to genetic exchange between hybridizing populations. *Heredity* **56**: 357–376.

Bateman AJ (1952). Self-incompatibility systems in angiosperms I Theory. *Heredity* **6**: 285–310.

Bos M, Van der Haring E (1988). Gene flow in Plantago. II. Gene flow pattern and population structure: a simulation study. *Heredity* **61**: 1–11.

Brennan AC, Harris SA, Hiscock SJ (2003). Population genetics of sporophytic self-incompatibility in *Senecio squalidus* L Asteraceae) II: a spatial autocorrelation approach to determining mating behaviour in the presence of low S allele diversity. *Heredity* **91**: 502–509.

Brooks RJ, Tobias AM, Lawrence MJ (1996). The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. XI. The effects of limited pollen and seed dispersal, overlapping generations and variation in plant size on the variance of S-allele frequencies in populations at equilibrium. *Heredity* **76**: 367–376.

Burger JM, Kirchner M, Bramanti B, Haak W, Thomas MG (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc Natl Acad Sci USA* **104**: 3736–3741.

Busch JW, Schoen JD (2008). The evolution of self-incompatibility when mates are limiting. *Trends Plant Sci* **13**: 128–136.

Cartwright RA (2009). Antagonism between local dispersal and self-incompatibility systems in a continuous plant population. *Mol Ecol* **18**: 2327–2336.

Charlesworth B, Nordborg M, Charlesworth D (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res* **70**: 155–174.

Clauss MJ, Cobban H, Mitchell-Olds T (2002). Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaceae). *Mol Ecol* **11**: 591–601.

Crawford TJ (1984). The estimation of neighbourhood parameters for plant populations. *Heredity* **53**: 273–283.

Glémin S, Gaude T, Guillemin ML, Lourmas M, Olivieri I, Mignot A (2005). Balancing selection in the wild: testing population genetics theory of self-incompatibility in the rare species *Brassica insularis*. *Genetics* **171**: 279–289.

Hamrick JL, Godt MJW (1996). Effects of life history traits on genetic diversity in plant species. *Philos Trans R Soc London B Biol Sci* **351**: 1291–1298.

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). A versatile computer program to analyse spatial genetic structure at the individual or population model. *Mol Ecol Notes* **2**: 618–620.

Heuertz M, Vekemans X, Hausman J-F, Palada M, Hardy OJ (2003). Estimating seed versus pollen dispersal from spatial genetic structure in the common ash. *Mol Ecol* **12**: 2483–2495.

Heywood JS (1991). Spatial analysis of genetic variation in plant populations. *Annu Rev Ecol Syst* **22**: 335–355.

Husband BC, Barrett SCH (1992). Genetic drift and the maintenance of the style length polymorphism in tristylous populations of *Eichhornia paniculata* (Pontederiaceae). *Heredity* **69**: 440–449.

Kirchner F, Robert A, Colas B (2006). Modelling the dynamics of introduced populations in the narrow-endemic *Centaurea corymbosa*: a demo-genetic integration. *J Appl Ecol* **42**: 1011–1021.

Leinonen T, O'Hara RB, Cano JM, Merila J (2008). Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *J Evol Biol* **21**: 1–17.

Llaurens V, Castric V, Austerlitz F, Vekemans X (2008a). High paternal diversity in the self-incompatible herb *Arabidopsis halleri* despite clonal reproduction and spatial restricted pollen dispersal. *Mol Ecol* **17**: 1577–1588.

Llaurens V, Billiard S, Leducq J-B, Castric V, Klein E, Vekemans X (2008b). Does frequency-dependant selection with complex dominance interactions accurately predict allelic frequencies at the self-incompatibility locus in *Arabidopsis halleri*? *Evolution* **62**: 2545–2557.

Loiselle BA, Sork VL, Nason J, Graham C (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am J Bot* **82**: 1420–1425.

Muirhead CA (2001). Consequences of population structure on genes under balancing selection. *Evolution* **55**: 1532–1541.

Neuhauser C (1999). The ancestral graph and gene genealogy under frequency-dependent selection. *Theor Popul Biol* **59**: 203–214.

Pauwels M, Saumitou-Laprade P, Holl A-C, Petit D, Bonnin I (2005). Multiple origin of metallicolous populations of the pseudometallophyte *Arabidopsis halleri* (Brassicaceae) in Central Europe: the cpDNA testimony. *Mol Ecol* **14**: 4403–4414.

Rousset F (2000). Genetic differentiation between individuals. *J Evol Biol* **13**: 58–62.

Schierup MH, Bechsgaard JS, Christiansen FB (2008). Selection at work in self-incompatible *Arabidopsis lyrata*. II. Spatial distribution of S-haplotypes in Iceland. *Genetics* **180**: 1051–1059.

Schierup MH, Vekemans X, Charlesworth D (2000). The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genet Res* **76**: 51–62.

Schierup MH, Vekemans X, Christiansen FB (1997). Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. *Genetics* **147**: 835–846.

Schueler S, Tusch A, Scholz F (2006). Comparative analysis of the within-population genetic structure in wild cherry (*Prunus avium* L at the self-incompatibility locus and nuclear microsatellites. *Mol Ecol* **15**: 3231–3243.

Sokal RR, Jacquez GM, Wooten MC (1989). Spatial autocorrelation analysis of migration and selection. *Genetics* **121**: 845–855.

Stoeckel S, Castric V, Mariette S, Vekemans X (2008). Unequal allelic frequencies at the self-incompatibility locus within local populations of *Prunus avium* L: an effect of population structure? *J Evol Biol* **21**: 889–899.

Takayama S, Isogai A (2005). Self-incompatibility in plants. *Annu Rev Plant Biol* **59**: 467–489.

Vallejo-Marin M, O'Brien HE (2007). Correlated evolution of self-incompatibility and clonal reproduction in *Solanum* (Solanaceae). *New Phytol* **173**: 415–421.

Van Rossum F, Bonnin I, Fénart S, Pauwels M, Petit D, Saumitou-Laprade P (2004). Spatial genetic structure within a metallicolous population of *Arabidopsis halleri*, a clonal, self-incompatible and heavy-metal-tolerant species. *Mol Ecol* **13**: 2959–2967.

Van Rossum F, Triest L (2006). Fine-scale genetic structure of the common *Primula elatior* (Primulaceae) at an early stage of population fragmentation. *Am J Bot* **93**: 1281–1288.

Vekemans X, Schierup MH, Christiansen FB (1998). Mate availability and fecundity selection in multi-allelic self-incompatibility systems in plants. *Evolution* **52**: 19–29.

Vekemans X, Hardy O (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol Ecol* **13**: 921–935.

Wagenius S, Lonsdorf E, Neuhauser C (2007). Patch aging and the S-Allee effect: breeding system effects on the demographic response of plants to habitat fragmentation. *Am Nat* **169**: 383–397.

Wright S (1939). The distribution of self sterility alleles in populations. *Genetics* **24**: 538–552.

Yokoyama S, Hetherington LE (1982). The expected number of self-incompatibility alleles in finite plant populations. *Heredity* **48**: 299–303.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)