doi:10.1111/j.1420-9101.2008.01504.x

Unequal allelic frequencies at the self-incompatibility locus within local populations of *Prunus avium* L.: an effect of population structure?

S. STOECKEL,*,† V. CASTRIC,‡ S. MARIETTE*,†,§ & X. VEKEMANS‡

*Cemagref, Unité de Recherche 'Ecosystèmes Forestiers', Domaine des Barres, Nogent-sur-Vernisson, France
†Université Paris XI & UPMC, UMR 8079, Laboratoire Ecologie, Systématique et Evolution, Orsay, France
‡Laboratoire de Génétique et Evolution des Populations Végétales, Université des sciences et technologies de Lille 1, CNRS UMR 8016, Villeneuve d'Ascq, France
§Unité de Recherche sur les Espèces Fruitières, INRA, Domaine de la Grande Ferrade, Villenave d'Ornon Cedex, France

Keywords:

balancing selection;
isoplethy;
population genetic structure;
S-alleles;
self-incompatibility.

Abstract

In this paper, we investigated the genetic structure and distribution of allelic frequencies at the gametophytic self-incompatibility locus in three populations of *Prunus avium* L. In line with theoretical predictions under balancing selection, genetic structure at the self-incompatibility locus was almost three times lower than at seven unlinked microsatellites. Furthermore, we found that S-allele frequencies in wild cherry populations departed significantly from the expected isoplethic distribution towards which balancing selection is expected to drive allelic frequencies (i.e. identical frequency equal to the inverse of the number of alleles in the population). To assess whether this departure could be caused either by drift alone or by population structure, we used numerical simulations to compare our observations with allelic frequency distributions expected : (1) within a single deme from a subdivided population with various levels of differentiation; and (2) within a finite panmictic population with identical allelic diversity. We also investigated the effects of sample size and degree of population structure on tests of departure from isoplethic equilibrium. Overall, our results showed that the observed allele frequency distributions were consistent with a model of subdivided population with demes linked by moderate migration rate.

Introduction

One obvious effect of natural selection is to alter the distribution of allelic frequencies at target loci in finite populations, when compared with loci subject to drift and mutation only (Nei, 1987). Such an effect will only be transient for positive and negative directional selection, whereas stable non-neutral distributions of allelic frequencies will occur under balancing selection. Balancing selection is a class of models of selection including negative frequency dependence, over-dominant selection and environmental heterogeneity, which typically increase the number of intermediate-frequency alleles when compared with the neutral model (Maruyama & Nei, 1981). A classical example of balancing selection involves self-incompatibility (SI) systems in plants (Wright, 1939). SI prevents self-fertilization in hermaphrodite plants through recognition and rejection of pollen expressing the same allelic specificity as that expressed in the pistils (de Nettancourt, 2001). Negative frequency-dependent selection arises in SI mainly through male function because pollen carrying a rare allelic specificity will be able to fertilize a larger proportion of mates than pollen with a common allele (Wright, 1939). In SI systems, pollen specificity may be determined either gametophytically (in gametophytic SI systems, hereafter GSI systems) or sporophytically (in sporophytic SI systems).

Theoretical investigations of GSI systems have considered the distribution of allelic frequencies at the locus controlling SI (S-locus) in infinite and finite

Correspondence: Solenn Stoeckel, INRA Agrocampus Rennes UMR 1099 BIO3P (Biology of Organisms and Populations applied to Plant Protection), F-35653 LE RHEU, France.
Tel.: +33 2 23 48 70 83; fax: +33 2 23 48 51 50;
e-mail: solenn.stoeckel@rennes.inra.fr

populations. In an infinite panmictic population, Wright (1939) pointed out that selective equivalence among the different allelic specificities should lead to equal allelic frequencies at equilibrium, the so-called 'isoplethy' hypothesis (see also Fisher, 1941; Finney, 1952; and Nagylaki, 1975; Boucher, 1993 for analytical derivations). Wright (1939) also derived the stationary frequency distribution of alleles at equilibrium under mutation, selection and drift in a finite panmictic population. This distribution presents an expected peak at an intermediate allelic frequency corresponding to the isoplethic equilibrium, hence a distribution that is highly distinct from the expected U-shaped distribution at a neutral locus (Kimura, 1983).

Several studies on plants with GSI attempted to test the 'isoplethy' hypothesis by sampling individuals within natural populations and typing alleles at the S-locus (referred to as S-alleles). The empirical frequency distributions obtained have generally been tested against isoplethy using a modification by Davies of Mantel's (1974) chi-squared test (Campbell & Lawrence, 1981). In a recent review, Castric & Vekemans (2004) reported that of 11 species with GSI for which this test had been performed, a highly significant rejection of the isoplethy hypothesis was obtained in two cases. This is certainly an underestimation of the occurrence of departure from Wright's hypothesis because applications of the isoplethy test, as it was performed, have low statistical power because of low sample sizes (Lawrence, 2000; this study). The explanations suggested by authors to explain such departure from isoplethy include either sampling effects (Campbell & Lawrence, 1981), departure from equilibrium because of demographic perturbations (Kato *et al.*, 2007), the occurrence of additional selective processes interacting with the sole negative frequency-dependent selection, such as S-lineage-specific loads of sheltered deleterious alleles (Uyenoyama, 1997) or positive directional selection on linked genes with independent function (Lane & Lawrence, 1995). We focus in this study on two important, yet largely overlooked, explanations for departure from isoplethy. The first one is that the test commonly used does not take into account the variance in allelic frequency expected in finite populations because of drift, whereas such variance can be large in small populations (Wright, 1939). Hence, the isoplethy test is not appropriate to check the predictions of the negative frequency-dependent selection model in finite populations. The second explanation has been suggested by Schierup (1998), and consists in the effect of population structure on the stationary frequency distribution of alleles within a local deme. Under an island model of migration, he showed that for intermediate levels of migration, the expected distribution of allelic frequencies within a deme is qualitatively different from that in a panmictic population. This is because of the fact that under intermediate levels of migration, the isoplethic equilibrium of allelic frequencies within a deme is never reached because of the constant influx of alleles from the overall metapopulation. This phenomenon is potentially important even for surprisingly low levels of migration because balancing selection leads to more effective migration (Schierup *et al.*, 2000a; Muirhead, 2001). Indeed, in a subdivided plant population with GSI, migrant S-alleles that are rare in the recipient populations will be favoured over resident alleles thanks to negative frequency-dependent selection. This process would lead to an increase in effective migration rate for S alleles when compared with neutral alleles. Theory thus predicts that differentiation at the S-locus should remain low even under very restricted migration (Schierup *et al.*, 2000a; Muirhead, 2001), and the influx of S-alleles into a population would often be high enough to alter the stationary frequency distribution of alleles (Schierup, 1998). Only the former prediction has been tested in studies from natural populations, with the conclusion that a lower genetic structure at the S-locus when compared with marker loci seems to be a general feature of plant SI systems, although most conclusive evidence comes from studies in plants with sporophytic SI (Glémin *et al.*, 2005; Brennan *et al.*, 2006; Kamau *et al.*, 2007). The effect of gene flow on distributions of allelic frequencies at the S-locus, by contrast, has not been studied empirically.

In this paper, we address the issue of detecting the effect of random genetic drift within a panmictic population, and that of population structure, in causing departure from isoplethic equilibrium at the S-locus in samples from natural populations of a tree species with GSI, *Prunus avium* L. (Rosaceae). We used previously published data on microsatellite loci in three populations of *P. avium* (Stoeckel *et al.*, 2006) to estimate neutral population genetic structure to compare with the differentiation at the S-locus, and to infer levels of gene flow among populations. We used numerical simulations to investigate quantitatively differences expected in stationary frequency distributions of alleles between finite panmictic populations and single demes within subdivided populations with variable gene flow. Indeed, in his study of the effect of population subdivision, Schierup (1998) has not explicitly compared the two types of distributions. The simulation results allowed to determine the effect of sample size and the level of population structure on tests of departure from stationary frequency distributions of alleles expected in panmictic populations. Finally, we tested observed distributions of allelic frequencies at the S-locus in samples from the three natural populations of *P. avium* against expectations generated by simulations under models of panmictic or subdivided populations.

Material and methods

Plant material and genotyping of the S-locus and microsatellite loci

The wild cherry (*P. avium* L., Rosaceae) is a scattered temperate tree species that naturally grows in European forests. It propagates both by asexual root-sprouting and sexual reproduction. Pollen is carried by insects, whereas seeds coated in fleshy fruits are mainly dispersed by birds and mammals.

We sampled three wild cherry populations from France [St Gobain (49°37′N, 3°26′E), Pagny (47°01′N, 5°12′E) and Comté (45°30′N, 3°20′E)]. Random samples of about 50 individuals were taken from the Pagny and Comté populations, whereas the St Gobain population was sampled exhaustively (245 trees) as described in Stoeckel et al. (2006). All sampled individuals were genotyped at the S-locus and at seven microsatellite loci using methods described in Stoeckel et al. (2006). S-alleles were identified using three pairs of consensus primers that result in an S-allele-specific combination of length polymorphisms (Sonneveld et al., 2003, 2006). Patterns of within-population genetic variation at the S-locus and at the microsatellite loci were analysed and discussed in Stoeckel et al. (2006). Here, we compared genetic structure among populations at both types of loci, and analyse the distribution of allelic frequencies at the S-locus. Details and references for microsatellite markers are given in Table 1.

Genetic data analysis

The F_{ST} statistic characterizing the level of population genetic structure was estimated for microsatellites and for the S-locus with Fstat v2.9.3.2 (Goudet, 1995). As wild cherry is a root-sprouting species, we performed all genetic analyses considering only a single individual per multilocus genotype (genet sample). To detect whether population structure was lower at the S-locus, a 95% confidence interval for F_{ST} was obtained by bootstrapping over all loci (GENETIX v4.05, Belkhir *et al.*, 1996–2004). Assuming an island model of migration at drift–migration equilibrium, the number of migrants per generation (Nm) was estimated from the value of the F_{ST} statistic computed on the microsatellite loci according to (Wright, 1931):

$$Nm = \frac{1 - F_{ST}}{4F_{ST}}.$$

Table 1 References, genomic location, number of alleles and population structure F_{ST} statistic for each microsatellite control locus and the S-locus in a sample from three populations of *Prunus avium*.

Marker names	References	Linkage group	Total number of alleles	F_{ST} (Genet)
UDP96-005	Cipriani *et al.* (1999)	G1	9	0.066
BPPCT034	Dirlewanger *et al.* (2002)	G2	14	0.090
PCEGA34	Downey & Iezzoni (2000)	G2	19	0.057
UDP98-411	Testolin *et al.* (2000)	G2	6	0.052
BPPCT040	Dirlewanger *et al.* (2002)	G4	8	0.052
PS12A02	Sosinski *et al.* (2000) and Joobeur *et al.* (2000)	G4	10	0.102
UDP98-021	Testolin *et al.* (2000)	G6	4	0.110
S-locus	Sonneveld *et al.* (2006)	G6	18	0.023
		Mean (±SE)		0.066 (±0.010)

Procedures of numerical simulations

We performed numerical simulations with three goals. First, to compare distributions of allelic frequencies at the S-locus expected at the deme level within a subdivided population with varying levels of interdeme migration with those expected in panmictic populations adjusted for identical allelic diversity (referred to as 'exploratory' simulations). Second, to investigate how the degree of population structure and the sample size affect results from tests of departure from expected distributions of allelic frequencies in panmictic populations, either infinite (isoplethic expectation) or finite, when samples are taken from a single deme within a subdivided population (referred to as 'test of panmixia' simulations). The third goal was to compare the observed distributions of allelic frequencies at the S-locus in the three population samples of *P. avium* with stationary frequency distributions of alleles expected either in a finite panmictic population or in a subdivided population at the deme level (referred to as *ad hoc* simulations). In all cases, we performed simulations of an S-locus with GSI under an n-island model of migration. The simulation procedure followed Schierup *et al.* (2000b). We performed forward two-locus simulations assuming a population of S demes, each with N diploid individuals, giving a total population size of $N_T = S \times N$, according to an n-island population model with pollen but not seed migration. Because gametes rather than individuals migrated, the migration rate was $2m$, with m = proportion of immigrating individuals per deme per generation in an equivalent model of migration of diploid individuals. For simulations of a finite panmictic population, we simply set $S = 1$ and $m = 0$. The model simulated a locus responsible for GSI, at which new functional alleles were arising at a rate u per generation, under the infinite allele model (Kimura & Crow, 1964). To calibrate simulations to the observed population genetic structure at microsatellite loci, a selectively neutral locus unlinked to the S-locus was added to the simulations, subject to a mutation rate v per generation, also according to the infinite allele model. Each

run was started with $2N_T$ different alleles in the population at both the selected and neutral loci, and allowed to evolve for 80 000 generations, at which time approximate mutation–selection–drift equilibrium had been reached (data not shown). One thousand replicate runs were performed for each type of simulation. All simulations were performed with $u = 5 \times 10^{-7}$ and $v = 2 \times 10^{-4}$. These values are biologically realistic and generated levels of allelic diversity at the S-locus and the microsatellite loci, respectively, of the same order of magnitude as in the observed population samples. Moreover, the number of alleles and the shape of stationary frequency distributions are not very sensitive to changes in the mutation rate at the S-locus in panmictic populations (Vekemans & Slatkin, 1994), and in subdivided populations as long as the mutation rate is much lower than the migration rate. All simulations of subdivided populations were performed with $S = 10$ demes. Although this number was chosen arbitrarily, we used more than three demes because the three populations sampled are members of an overall fragmented landscape where several populations are potentially interacting. Additional simulations showed that the qualitative results obtained were not dependent on the actual number of demes.

Exploratory simulations

For simulations of subdivided populations, we chose $N_T = 1500$, which gives a deme size $N = 150$, to approach the size and S-locus allelic diversity of the largest sampled population (St Gobain). Values of m were chosen to produce a wide range of N_m values, including 0.01, 0.1, 0.5, 1, 3, 5, 10, 20 and 50. At the end of each run, we computed the average number of different alleles within a single deme, n_a, and the F_{ST} statistic, computed according to Weir & Cockerham (1984). For simulations of panmictic populations, we adjusted values of $N_T = N$ to match the n_a values obtained in the subdivided population simulations for four selected Nm values (0.01, 0.5, 3 and 50). Distribution of allelic frequencies in panmictic populations and within demes of subdivided populations were constructed by recording across simulation runs the frequency of each allele present at the end of each run, and computing *a posteriori* the proportion of alleles recorded at each allele frequency.

Test of panmixia simulations

For subdivided populations, we simulated the same total population size ($N_T = 1500$) and range of Nm values as in the exploratory simulations. Random samples of 30 or 120 individuals were taken within a single deme at the end of each run and were used to compute n_a and distributions of allelic frequencies at the S-locus. For panmictic populations, we adjusted values of $N_T = N$ to match the n_a values obtained for each sample size and each Nm value. Tests of the isoplethy hypothesis (as expected in an infinite panmictic population) were performed on the average distributions computed for each parameter set in the subdivided populations using a modification by Davies of Mantel's (1974) chi-squared test, as described in Campbell & Lawrence (1981). The statistic was computed according to the following formula:

$$\chi^2_{(n_a-1)(Davies)} = \frac{(n_a - 1)\left(\sum_{i=1}^{n} (C_i^2) - \frac{4r^2}{n_a}\right)}{2r - \frac{4r}{n_a}}$$

where n_a is the number of alleles, C_i the observed sample frequency of the ith allele and r the sample size. As proposed by Mantel (1974), this statistic follows approximately a chi-squared distribution with $n_a - 1$ degrees of freedom. Goodness-of-fit tests against expected distributions in finite panmictic populations were performed by computing the average cumulated distribution of allelic frequencies in samples from subdivided populations and testing against 1000 replicate samples obtained by the simulations of panmictic populations adjusted for identical allelic diversity. We used the Kolmogorov–Smirnov goodness-of-fit approach to test the fit of a target distribution to that obtained according to the null hypothesis (Sokal & Rohlf, 1995). Briefly, the two cumulative distributions of allelic frequencies were binned into 0.02-wide frequency classes. The proportions of alleles in each bin in the two distributions were compared using the Kolmogorov–Smirnov test statistic D_{max}, which measures the maximal difference between the two distributions compared. The fit of the target distribution (from a single deme within subdivided populations) was then quantified as a P-value equivalent computed by comparing the D_{max} value obtained as described above with a distribution of D_{max} values obtained by comparing each replicate against the average values under the null hypothesis (panmictic populations).

Ad hoc simulations

The procedure was similar to that used for 'test of panmixia' simulations both in subdivided and panmictic populations, but total population size N_T was adjusted to match the observed n_a values at the S-locus in samples corresponding to each of the three populations investigated, and sample sizes were 32, 27 and 113 individuals, which correspond to samples from the Comté, Pagny and St Gobain populations respectively. To test the observed allele frequency distributions against expectations under a finite panmictic population model, we applied the Kolmogorov–Smirnov goodness-of-fit test described above with the target distribution being the observed one, and the null hypothesis distributions obtained with simulations of panmictic populations. The same procedure was used to test the observed distributions against expectations within a single deme from subdivided population simulations.

Results

Comparison of population genetic structure at the S-locus and at neutral loci for wild cherry populations

The average F_{ST} for all seven microsatellite loci was 0.074, and ranged from 0.052 to 0.110, which was substantially higher than the value for the S-locus (F_{ST} = 0.023, Table 1). This difference was highly significant, as the 95% confidence interval of F_{ST} obtained by bootstrap across loci was 0.049–0.086, which did not include the value for the S-locus. Assuming drift/migration equilibrium in an island model of migration, the F_{ST} for microsatellite loci amounts to a mean number of migrants among demes per generation (*Nm*) equal to 3.4.

Stationary frequency distributions of alleles in panmictic vs. subdivided population models

Results from the 'exploratory' set of simulations in subdivided populations showed that the number of alleles at the S-locus, n_a, ranged from 8.4 to 20.1 at the deme level, for a range in migration parameter, *Nm*, from 0.01 to 50 (Fig. 1). Large differences in population genetic structure were observed between the S-locus and the neutral locus, with values of F_{ST} up to 0.94 at the neutral locus, whereas the maximum F_{ST} value at the S-locus was 0.07. As expected by construction of the simulation procedure, the value of F_{ST} at the neutral locus that was closest to the observed value (F_{ST} = 0.074) was obtained for *Nm* = 3. At this level of migration, the F_{ST} value obtained by simulation at the S-locus was 0.022, which was remarkably similar to the observed value, 0.023.

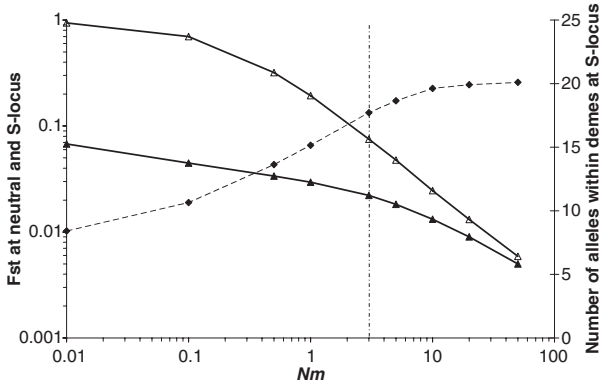

Fig. 1 Simulations of a subdivided population under GSI with 10 demes of size N = 150 across a range of migration rate corresponding to *Nm* = 0.01–50. Values of F_{ST} (solid lines) at the S-locus (filled triangles) and at a neutral unlinked locus (open triangles); values of the number of alleles at the S-locus maintained within a single deme (broken line with filled diamonds). The vertical line represents the level of migration inferred from microsatellite data in the populations of *Prunus avium*.

For comparison purposes, simulations of panmictic populations were adjusted to match each of the n_a values observed in subdivided populations at the deme level (Fig. 1). We found that panmictic population sizes between 300 and 1500 individuals were needed to maintain numbers of alleles similar to those obtained in a deme of size 150 individuals under a subdivided population model (N_T = 1500) for a range in *Nm* from 0.01 to 50. We compared stationary allelic frequency distributions at the S-locus between exploratory simulations under panmictic and subdivided population models (Fig. 2). A striking result was that for intermediate levels of migration (*Nm* = 0.5 and 3), that represent the most commonly observed values for outcrossing plant species, the distributions obtained within demes of a subdivided population were substantially more platikurtic (more flat) than those from corresponding panmictic populations: the former presenting a clear excess of rare and high-frequency alleles when compared with the latter. Although the trend was similar for cases with very low (*Nm* = 0.01) or very high (*Nm* = 50) migration rate, the difference in shape of the distributions between panmictic and subdivided populations was much smaller.

Effect of sample size and migration rates in subdivided populations on tests of departure from panmictic population models

Distributions of allelic frequencies in samples taken within single demes of subdivided populations were compared with expectations under panmictic population models, as obtained from the 'test of panmixia' set of simulations. Results were compared for two sample sizes (30 and 120 individuals) and a range of *Nm* values between 0.01 and 50. Results of Davies' test of the isoplethy hypothesis, corresponding to expectation in an infinite panmictic population, are shown in Table 2. For a sample size of 120 individuals, rejection of isoplethy was highly significant through the whole range of migration rates investigated (*Nm* = 0.01–50). However, for a sample size of 30 individuals, rejection was significant only for intermediate levels of migration (*Nm* = 0.5–5). Table 2 also shows the results of the Kolmogorov–Smirnov goodness-of-fit test against samples from simulations in finite panmictic populations. For a sample size of 120 individuals, average cumulated distributions of allelic frequencies within demes in subdivided populations were significantly different from those obtained in panmictic populations only at intermediate migration levels (*Nm* = 1–5). Moreover, for a sample size of 30 individuals, none of the tests were significant. Hence, both sample size and level of migration have a large effect on these tests of departure from panmictic population models, which suggests that subdivided population effects can only be detected for intermediate levels of migration and sufficiently large sample sizes.

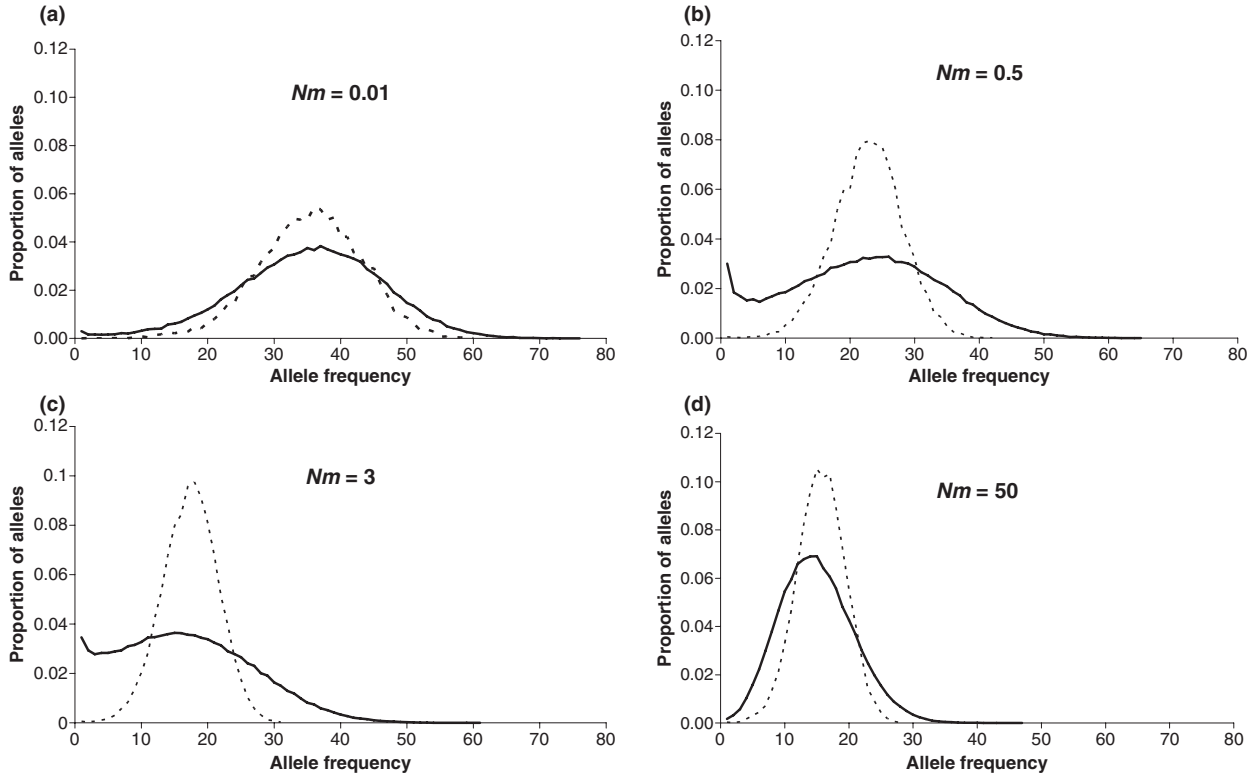

Fig. 2 Comparison of stationary distributions of allelic frequencies at the S-locus between a panmictic population (dotted line) and a single deme from a subdivided population (solid line) for a range of *Nm* values. (a) *Nm* = 0.01; (b) *Nm* = 0.5; (c) *Nm* = 3; (d) *Nm* = 50. For simulation conditions, see 'exploratory simulations' in the text.

Table 2 Effect of sample size (*K* = 30 or 120 individuals) and migration rates (*Nm* = 0.01–50) in subdivided populations sampled within a single deme on tests of departure from panmictic population models.

	Davies' test of isoplethy (*P*-value)		Goodness-of-fit test against a finite panmictic population (*P*-value)	
Nm	*K* = 30	*K* = 120	*K* = 30	*K* = 120
0.01	0.088	< 0.0001	> 0.999	0.876
0.1	0.053	< 0.0001	> 0.999	0.580
0.5	0.028	< 0.0001	0.908	0.131
1	0.035	< 0.0001	0.846	0.027
3	0.044	< 0.0001	0.357	0.033
5	0.089	< 0.0001	0.830	0.034
10	0.134	< 0.0001	0.880	0.241
20	0.417	< 0.0001	0.918	0.614
50	0.578	< 0.0001	> 0.999	> 0.999

Goodness-of-fit tests comparing observed distributions against panmictic and subdivided population models

Eighteen distinct S-alleles were found across the three populations with 18, 10 and 13 S-alleles identified in the St Gobain, Pagny and Comté populations respectively (Table 3). Using these numbers of alleles in each wild cherry population, the expected deterministic equilibrium frequencies for each allele were 0.077, 0.100 and 0.056 in the Comté, Pagny and St Gobain populations respectively. By contrast, the observed allelic frequencies within the samples ranged widely from 0.016 to 0.172 in Comté, from 0.019 to 0.185 in Pagny and from 0.004 to 0.137 in St Gobain (Table 3 and Fig. 1). The isoplethy hypothesis was thus highly significantly rejected in each population sample (Table 4). Cumulative allelic frequency distributions in each population sample were compared with their expected distributions obtained by *ad hoc* simulations of GSI in finite panmictic populations with sample size and number of *S*-alleles matching the observed values (Fig. 3). In each case, the observed distribution showed an excess of low- and high-frequency alleles, and a deficit of alleles at intermediate frequency when compared with expected distributions under the hypothesis that samples were derived from panmictic populations. This discrepancy was highly significant in St Gobain, marginally significant in Pagny ($P = 0.085$) but not significant in Comté (Table 4). Comparisons were also made with expected distributions in subdivided populations in samples taken within a

Table 3 The observed and expected S-allele frequencies under isoplethy [$\overline{fq}_{(iso)}$] in three populations of *Prunus avium* considering only one representative per genotype (genet sample).

S-allele name	Comté	Pagny	St Gobain
S_1	0.172	ND	0.084
S_2	0.063	0.111	0.080
S_3	0.094	0.185	0.062
S_4	ND	ND	0.004
S_6	0.047	0.056	0.049
S_7	0.109	0.130	0.137
S_9	ND	ND	0.004
S_{10}	0.078	ND	0.093
S_{12}	0.141	ND	0.102
S_{13}	0.016	ND	0.066
S_{14}	0.109	0.185	0.066
$S_{x/16}$	ND	ND	0.018
S_{17}	0.063	0.019	0.009
S_{18}	0.063	0.130	0.049
S_{19}	0.031	0.130	0.031
S_{20}	ND	0.037	0.124
S_{21}	ND	0.019	0.004
S_{22}	0.016	ND	0.018
n/n_t	0.684	0.526	0.947
$\overline{fq}_{(iso)}$	0.077	0.100	0.056

n/n_t was the proportion of S-alleles identified in the population considering the number of known S-alleles already described in the species (n_t = 19) (see Sonneveld *et al.*, 2003; De Cuyper *et al.*, 2005). ND represents S-allele not detected in the sample.

Table 4 Test of departure of allelic frequency distributions at the S-locus in samples from three natural populations of *Prunus avium* from expected distributions for an infinite (Davies' test of isoplethy) or finite (Kolmogorov–Smirnov's test against simulation results) panmictic or subdivided population with matching number of S-alleles.

	St Gobain	Comté	Pagny
Observed number of S-alleles	18	13	10
Sample size	113	32	27
Test against infinite panmictic population			
Davies' χ^2	495.26	34.43	47.42
P-value	< 0.0001	0.0006	< 0.0001
Test against finite panmictic population			
Simulated population size	1250	650	400
D_{max} expected 95% CI	0.072–0.238	0.089–0.302	0.113–0.331
D_{max} observed	0.469	0.168	0.310
P-value	< 0.001	0.444	0.085
Test against finite subdivided population			
Simulated total population size*	1500	800	500
D_{max} expected 95% CI	0.087–0.281	0.102–0.336	0.107–0.374
D_{max} observed	0.197	0.069	0.254
P-value	0.215	0.988	0.324

*Simulation of a sample from a single deme within a subdivided population with 10 demes and a migration rate corresponding to Nm = 3.

single deme, whose sample size, number of S-alleles and migration rates matched the observed values. In accordance with the hypothesis that population structure caused the observed unequal allelic frequencies, expected distributions under the subdivided population model fitted more closely the observed distributions than those obtained under a panmictic model (Fig. 3). Further, goodness-of-fit tests comparing observed and expected distributions under the subdivided population model were nonsignificant for the three population samples (Table 4).

Discussion

Weak population genetic structure at the S-locus

The level of population genetic structure observed among the three wild cherry populations at microsatellite loci (F_{ST} = 0.074) was typical of an outcrossing temperate tree species (Hamrick & Godt, 1990). Similar values were found with allozyme markers among populations of the same species: Frascaria *et al.* (1993) and Mariette *et al.* (1997) reported F_{ST} values, respectively, of 0.049 and 0.052 for studies with four and five population samples respectively. At the global European scale, Tavaud (2002) estimated the population differentiation at nearly 0.071 with microsatellites, considering a western and two eastern (Georgian and Romanian) populations. Those results thus confirm the occurrence in France, and so far at the European scale, of some intermediate genetic structure of *P. avium* species.

Our data allowed comparing F_{ST} estimates between a gametophytic S-locus and unlinked microsatellite loci in samples from natural populations. The selected locus showed a much lower genetic structure (almost reduced by a factor of three) when compared with control loci. This empirical observation is thus qualitatively consistent with predictions from theoretical models of strong balancing selection in subdivided populations (Schierup *et al.*, 2000a; Muirhead, 2001), and results from the effect of selection that increases the effective migration rate of S-alleles. A similar effect was observed in sporophytic SI species. In *Brassica insularis*, Glemin *et al.* (2005) found a significantly lower population structure at the S-locus when compared with microsatellite markers. By contrast, in *Senecio squalidus*, Brennan *et al.* (2006) found no significant difference in F_{ST} values between the sporophytic S-locus and allozyme markers. However, allozyme markers showed a significant pattern of isolation by distance not found at the S-locus. In *Arabidopsis lyrata*, using nucleotide sequences at the pistil component of the S-locus, Charlesworth *et al.* (2003) did not detect significant genetic differentiation among five populations, whereas it was found at four of six control genes from the same gene family. In addition, Kamau *et al.* (2007) reported a lower population structure in

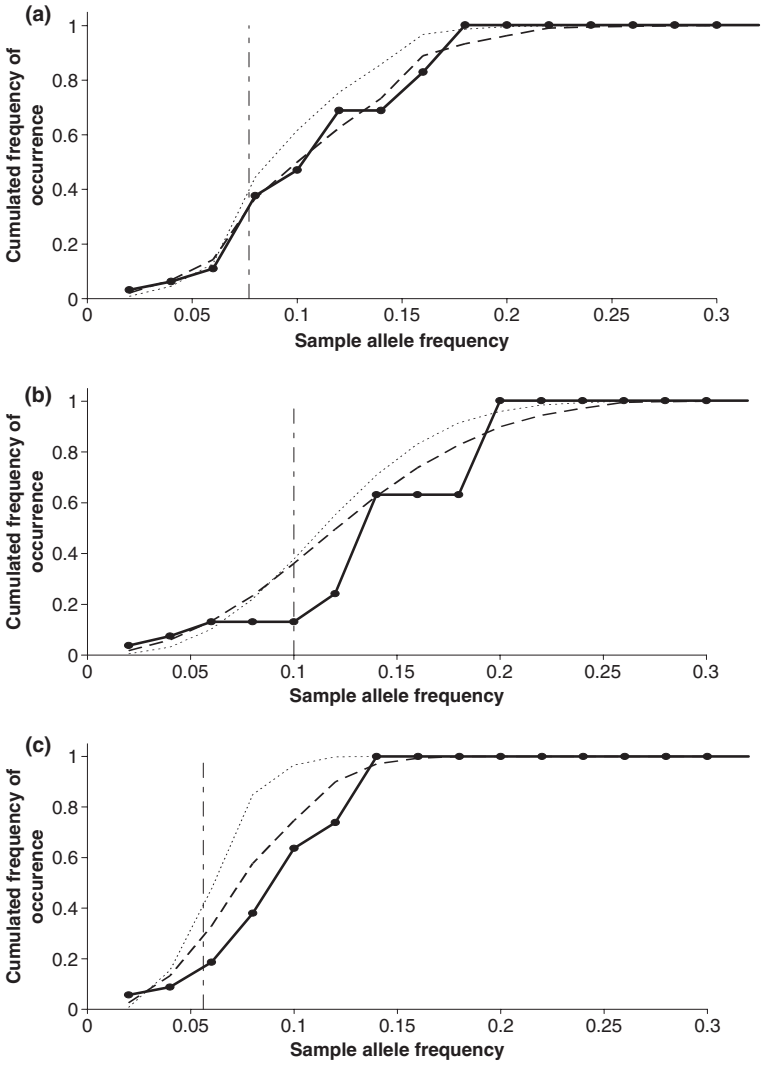

Fig. 3 Observed and expected cumulative distributions of allelic frequencies at the S-locus in samples from three populations of wild cherry trees and from *ad hoc* simulations of panmictic and subdivided population models. (a) Comté population; (b) Pagny population; (c) St Gobain population. The observed distribution is represented by a thick solid line with solid dots representing the cumulated frequency of occurrence in the sample of alleles with a frequency in the corresponding interval. Expected distributions from panmictic and subdivided population simulations are represented by dotted and broken lines respectively. The vertical line represents the expected frequency under the isoplethy hypothesis.

the genomic region surrounding the S-locus in *A. lyrata*, when compared with more distant or unlinked regions. On species with GSI, other studies report a low population genetic structure at the S-locus (F_{ST} = 0.014 in *Prunus lannesiana*; Kato *et al.*, 2007) or a high degree of allelic overlap at the S-locus in comparisons of geographically distant pairs of populations (*Oenothera organensis*, Emerson, 1939; *Solanum carolinense*, Richman *et al.*, 1995; *Papaver rhoeas*, O'donnell *et al.*, 1993; *Sorbus aucuparia*, Raspé & Kohn, 2007), but these studies did not report differentiation levels at neutral genetic markers. Nevertheless, combining concordant results from all these studies and ours suggests that low population genetic structure is indeed a general property of SI genes. The magnitude of the decrease in genetic differentiation for the S-locus when compared with unlinked neutral loci depends on the level of migration (Fig. 1), and in our study we obtained a remarkable quantitative agreement between observed and expected reductions in F_{ST} at the S-locus.

Departure from isoplethy

We showed that the S-alleles in these natural populations of *P. avium* are not maintained at equal frequencies as reflected by the highly significant tests against the isoplethy hypothesis (Table 4). The observed distributions of S-allele frequencies within each population sample also showed discrepancies with expected distributions in finite panmictic populations, i.e. allowing for variances in allele frequencies because of genetic drift was not sufficient to explain the observed patterns. Indeed, the population samples displayed an excess of S-alleles at low and high frequencies, but a deficit of alleles at intermediate frequencies, when compared with expectations from panmictic population simulations (Fig. 3).

Although the patterns were similar in all three population samples, results from goodness-of-fit tests were highly significant in the St Gobain population, marginally significant in the Pagny population and nonsignificant in the Comté population. These differences could be due either to true differences in distributions of allelic frequencies in these populations, or to differences in statistical power of the test, as the sample size in St Gobain is more than three times higher than in the other populations. This latter explanation is supported by the results of our simulations showing that samples of 30 individuals from single demes within subdivided populations could not lead to rejection of the fit to panmictic populations irrespective of the level of population structure, in contrast to samples of size 120 for a range of Nm = 1–5 (Table 2). We hypothesized that the discrepancy between the distribution of allelic frequencies in St Gobain and expectations for a finite panmictic population is because of population structure, as suggested by the theoretical results of Schierup (1998). Several lines of evidence are in agreement with this suggestion. First, formal comparisons of stationary allelic frequency distributions between simulations within demes of a subdivided population and within panmictic populations with matching allelic diversity have shown that an excess of rare and high-frequency alleles is expected in subdivided populations for intermediate levels of migration, as observed in St Gobain. Second, comparisons between observed distributions of allelic frequencies in each population sample with *ad hoc* simulations from subdivided populations showed a better fit than with those on panmictic populations, and the goodness-of-fit tests were nonsignificant even for the large sample of the St Gobain population. Similar deviation from the isoplethy hypothesis at the S-locus has been reported in another wild cherry population (Schueler *et al.*, 2006), and in several populations of another congeneric species (*P. lannesiana*, Kato & Mukai, 2004 and Kato *et al.*, 2007). These studies also used large sample sizes; so, it is possible that the deviations observed by rejection of the isoplethy test are due solely to the fact that they did not take into account the variance in allelic frequencies because of genetic drift. The authors invoked recent demographic perturbations and the long generation time of these tree species as the most likely explanation for these observations. However, in the study on *P. lannesiana*, an F_{ST} of 0.014 was reported for the S-locus, which corresponds to Nm = 5 (Fig. 1), thus precisely in the range where population structure is expected to influence the shape of allelic frequency distributions. We thus offer the alternative explanation that the subdivided population effect contributed to some extent to the observed pattern in *P. lannesiana.* More generally, we believe that this effect is pervasive in many plant species with GSI, because most outcrossing plant species show levels of population differentiation within the intermediate range (Nm = 1–5, corresponding to F_{ST} = 0.045–0.2 for neutral markers Hamrick *et al.*, 1992; Nybom, 2004), shown to influence stationary distributions of allelic frequencies within single demes. This conjecture seems to be in contradiction to the observation that most other attempts to test isoplethy in samples from natural populations of species with GSI systems failed to reject the isoplethy hypothesis (for references, see Castric & Vekemans, 2004). However, because these studies were based on samples of 12–25 individuals, with the exception of the *O. organensis* pioneer study which sampled 65 individuals (Emerson, 1939), our results suggest that they are irrelevant for proper tests of the deviation from the isoplethy hypothesis.

Selection effects on individual S-alleles, or on linked genes, have also been invoked as a potential cause of deviation from the isoplethy hypothesis, as for instance in studies on *P. rhoeas* (O'donnell *et al.*, 1993). Although such selection effects may be occurring, we think that future studies on natural populations of species with GSI should first consider the most parcimonious explanations such as the effect of drift and the effect of population structure.

Our simulation study also allowed us to investigate how Davies' test of isoplethy performs for two different sampling efforts. We found this test to be exceedingly sensitive to sample sizes, with large samples (here N = 120 individuals) frequently rejecting isoplethy even in cases very close to panmixia. We showed that this undesirable statistical property of the isoplethy test is because of the fact that it is not appropriate for samples from finite populations. According to Wright's (1939) model, the property of isoplethy is predicted in infinite panmictic populations. By contrast, genetic drift under finite populations would cause S-allele frequencies to fluctuate around the deterministic values. Hence, even when mutation–selection–drift has been reached, S-allele frequencies in a finite population are expected to display some variance, whose magnitude is determined by the effective population size (Wright, 1939). Genetic drift is ignored by Davies' test of isoplethy, which is based on the expected symmetric allelic frequency in infinite populations. We introduced an alternative test using numerical simulations under a finite panmictic population to compare the observed distribution with an expected distribution explicitly taking genetic drift into account. A possible perspective of this work could be to design a test based on Wright's predicted distribution in finite populations. However, this numerical simulation-based test is now capable of detecting the effect of subdivided populations in situations of large sample sizes and intermediate levels of population structure.

In conclusion, our results showed an excess of low and high-frequency S-alleles in samples from populations of *P. avium*. Our simulations revealed that the observed frequency distributions were consistent with a model of subdivided population with demes linked by moderate migration rate. As predicted by Schierup (1998), such an

intermediate genetic structure may be part of the explanation for the observed unequal allelic frequencies at the SI locus in *P. avium* and other species.

Acknowledgments

We warmly thank Jérôme Grange for the time he spent to help us in acquiring genetic and field data. We also thank Nathalie Frascaria-Lacoste for helpful discussions, and the two anonymous reviewers for comments that greatly improved the manuscript. The self-incompatibility team at the University of Lille 1 is financially supported through an ATIP-plus grant from the environment and sustainable development Department of the CNRS. Solenn Stoeckel was supported by Region Centre.

References

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Bonhomme, F. 1996–2004. GENETIX 4.05, Logiciel sous Windows TM pour la Génétique des Populations, Laboratoire Génome, Populations, Interactions. CNRS UMR 5171, Université de Montpellier II, Montpellier, France.

Boucher, W. 1993. A Deterministic Analysis of Self-Incompatibility Alleles. *J. Math. Biol.* **31**: 149–155.

Brennan, A.C., Harris, S.A. & Hiscock, S.J. 2006. The population genetics of sporophytic self-incompatibility in *Senecio squalidus* L. (Asteraceae): The number, frequency, and dominance interactions of S alleles across its British range. *Evolution* **60**: 213–224.

Campbell, J. & Lawrence, M. 1981. The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas.* II. The number and frequency of S-alleles in a natural population. *Heredity* **46**: 81–90.

Castric, V. & Vekemans, X. 2004. Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. *Mol. Ecol.* **13**: 2873–2889.

Charlesworth, D., Mable, B.K., Schierup, M.H., Bartolome, C. & Awadalla, P. 2003. Diversity and linkage of genes in the self-incompatibility gene family in *Arabidopsis lyrata*. *Genetics* **164**: 1519–1535.

Cipriani, G., Lot, G., Huang, W.G., Marrazzo, M.T., Peterlunger, E. & Testolin, R. 1999. AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L) Batsch]: isolation, characterisation and cross-species amplification in *Prunus*. *Theor. Appl. Genet.* **99**: 65–72.

De Cuyper, B., Sonneveld, T. & Tobutt, K.R. 2005. Determining self-incompatibility genotypes in Belgian wild cherries. *Mol. Ecol.* **14**: 945–955.

Dirlewanger, E., Cosson, P., Tavaud, M., Aranzana, M.J., Poizat, C., Zanetto, A., Arus, P. & Laigret, F. 2002. Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *Theor. Appl. Genet.* **105**: 127–138.

Downey, S.L. & Iezzoni, A.F. 2000. Polymorphic DNA markers in black cherry (*Prunus serotina*) are identified using sequences from sweet cherry, peach, and sour cherry. *J. Am. Soc. Hort. Sci.* **125**: 76–80.

Emerson, S. 1939. A preliminary survey of the *Oenothera organensis* population. *Genetics*, **24**: 524–537.

Finney, D. 1952. The equilibrium of a self-incompatible polymorphic species. *Genet. Res.* **26**: 33–64.

Fisher, R. 1941. The theoretical consequences of polyploid inheritance for the mid style form in *Lythrum salicaria*. *Ann. Eugen.* **11**: 31–38.

Frascaria, N., Santi, F. & Gouyon, P.H. 1993. Genetic differentiation within and among populations of chestnut (*Castanea sativa* Mill) and wild cherry (*Prunus avium* L). *Heredity* **70**: 634–641.

Glémin, S., Gaude, T., Guillemin, M.L., 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.

Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. *J. Hered.* **86**: 485–486.

Hamrick, J. & Godt, M. 1990. Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding, and Genetic Resources* (A. Brown, M.T. Clegg, A.L. Kahler & B. Weir, eds), pp. 43–63. Sinauer, Sunderland.

Hamrick, J., Godt, M. & Sherman-Broyles, S. 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forests* **6**: 95–124.

Joobeur, T., Periam, N., Vicente, M.C., King, G.J. & Arus, P. 2000. Development of a second generation linkage map for almond using RAPD and SSR markers. *Genome* **43**: 649–655.

Kamau, E., Charlesworth, B. & Charlesworth, D. 2007. Linkage disequilibrium and recombination rate estimates in the self-incompatibility region of *Arabidopsis lyrata*. *Genetics* **176**: 2357–2369.

Kato, S. & Mukai, Y. 2004. Allelic diversity of S-RNase at the self-incompatibility locus in natural flowering cherry populations (*Prunus lannesiana var. speciosa*). *Heredity* **92**: 249–256.

Kato, S., Iwata, H., Tsumura, Y. & Mukai, Y. 2007. Distribution of S-alleles in island populations of flowering cherry, *Prunus lannesiana var. speciosa*. *Genes Genet. Syst.* **82**: 65–75.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, MA.

Kimura, M. & Crow, J. 1964. The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725–738.

Lane, M.D. & Lawrence, M.J. 1995. The population genetics of the self-incompatibility polymorphism in *Papaver Rhoeas* .10. An association between incompatibility genotype and seed dormancy. *Heredity* **75**: 92–97.

Lawrence, M.J. 2000. Population genetics of the homomorphic self-incompatibility polymorphisms in flowering plants. *Ann. Bot.* **85**: 221–226.

Mantel, N. 1974. Comment and a suggestion. *J. Am. Stat. Assoc.* **89**: 378–380.

Mariette, S., Lefranc, M., Legrand, P., Taneyhill, D., FrascariaLacoste, N. & Machon, N. 1997. Genetic variability in wild cherry populations in France. Effects of colonizing processes. *Theor. Appl. Genet.* **94**: 904–908.

Maruyama, T. & Nei, M. 1981. Genetic variability maintained by mutation and overdominant selection in finite populations. *Genetics* **98**: 441–459.

Muirhead, C.A. 2001. Consequences of population structure on genes under balancing selection. *Evolution* **55**: 1532–1541.

Nagylaki, T. 1975. Deterministic behavior of self-incompatibility alleles. *Genetics* **79**: 545–550.

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

de Nettancourt, D. 2001. *Incompatibility and Incongruity in Wild and Cultivated Plants*. Springer-Verlag, Berlin, Germany.

Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol. Ecol.* **13**: 1143–1155.

O'donnell, S., Lane, M.D. & Lawrence, M.J. 1993. The population genetics of the self-incompatibility polymorphism in *Papaver Rhoeas* .6. Estimation of the overlap between the allelic complements of a pair of populations. *Heredity* **71**: 591–595.

Raspé, O. & Kohn, J.R. 2007. Population structure at the S-locus of *Sorbus aucuparia* L. (Rosaceae : Maloideae). *Mol. Ecol.* **16**: 1315–1325.

Richman, A.D., Kao, T.H., Schaeffer, S.W. & Uyenoyama, M.K. 1995. S-allele sequence diversity in natural populations of *Solanum Carolinense* (HORSENETTLE). *Heredity* **75**: 405–415.

Schierup, M.H. 1998. The Number of Self-Incompatibility Alleles in a Finite, Subdivided Population. *Genetics* **149**: 1153–1162.

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

Schierup, M.H., Charlesworth, D. & Vekemans, X. 2000b. The effect of hitch-hiking on genes linked to a balanced polymorphism in a subdivided population. *Genet. Res.* **76**: 63–73.

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, R. & Rohlf, F. 1995. *Biometry*. W.H. Freeman and Company, New York.

Sonneveld, T., Tobutt, K.R. & Robbins, T.P. 2003. Allele-specific PCR detection of sweet cherry self-incompatibility (S) alleles S1 to S16 using consensus and allele-specific primers. *Theor. Appl. Genet.* **107**: 1059–1070.

Sonneveld, T., Robbins, T.P. & Tobutt, K.R. 2006. Improved discrimination of self-incompatibility S-RNase alleles in cherry and high throughput genotyping by automated sizing of first intron polymerase chain reaction products. *Plant Breed.* **125**: 305–307.

Sosinski, B., Gannavarapu, M., Hager, L.D., Beck, L.E., King, G.J., Ryder, C.D., Rajapakse, S., Baird, W.V., Ballard, R.E. & Abbott, A.G. 2000. Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. *Theor. Appl. Genet.* **101**: 421–428.

Stoeckel, S., Grange, J., Fernandez-Manjarres, J.F., Bilger, I., Frascaria-Lacoste, N. & Mariette, S. 2006. Heterozygote excess in a self-incompatible and partially clonal forest tree species - *Prunus avium* L. *Mol. Ecol.* **15**: 2109–2118.

Tavaud, M. 2002. *Diversité du Cerisier doux (Prunus avium* L.) sur son Aire de Répartition : Comparaison Avec ses Espèces Apparentées (P. cerasus et P. gondouinii) et son Compartiment Sauvage. INRA Bordeaux, ENSAM, Montpellier, France: 98.

Testolin, R., Marrazzo, T., Cipriani, G., Quarta, R., Verde, I., Dettori, M.T., Pancaldi, M. & Sansavini, S. 2000. Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. *Genome* **43**: 512–520.

Uyenoyama, M.K. 1997. Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics* **147**: 1389–1400.

Vekemans, X. & Slatkin, M. 1994. Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics* **137**: 1157–1165.

Weir, B.S. & Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population-structure. *Evolution* **38**: 1358–1370.

Wright, S. 1931. Evolution in mendelian populations. *Genetics* **16**: 97–159.

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

Received 13 September 2007; revised 17 December 2007; accepted 20 December 2007