Heredity (2008) 101, 351–358

www.nature.com/hdy

ORIGINAL ARTICLE

Inferences on the number and frequency of *S*-pollen gene (*SFB*) specificities in the polyploid *Prunus spinosa*

J Vieira, RAM Santos, SM Ferreira and CP Vieira
Molecular Evolution Group, Instituto de Biologia Molecular e Celular, University of Porto, Porto, Portugal

In flowering plants, self-incompatibility is a genetic mechanism that prevents self-fertilization. In gametophytic self-incompatibility (GSI), pollen specificity is encoded by the haploid genotype of the pollen tube. In GSI, specificities are maintained by frequency-dependent selection, and for diploid species, at equilibrium, equal specificity frequencies (isoplethy) are expected. This prediction has been tested in diploid, but never in polyploid self-incompatible species. For the latter, there is no theoretical expectation regarding isoplethy. Here, we report the first empirical study on specificity frequencies in a natural population of a polyploid self-incompatible species, *Prunus spinosa*. A total of 32 *SFB* (the pollen *S* gene) putative specificities are observed in a large sample from a natural population. Although *P. spinosa* is polyploid, the number of specificities found is similar to that reported for other diploid Rosaceae species. Unequal specificity frequencies are observed.
Heredity (2008) **101,** 351–358; doi:10.1038/hdy.2008.60; published online 2 July 2008

Keywords: gametophytic self-incompatibility; *SFB*; allele number; *Prunus spinosa*; isoplethy

Introduction

One effective way to assure genetic variability in hermaphrodite plants is self-incompatibility (SI), that consists of recognition of self- or self-related pollen by cells of the pistil, followed by rejection of the incompatible pollen through aborted development, either immediately after pollination on the stigma surface, or at a subsequent stage during pollen tube growth into the stigma or style (de Nettancourt, 1977). Recognition is a 'lock-and-key' system, and each 'lock-and-key' is defined as a specificity (Charlesworth *et al*., 2005). In gametophytic SI (GSI), pollen specificity is determined by the *S*-locus genotype of the haploid gametophyte. Therefore, for a given specificity, the rate of successful fertilization is inversely related to the specificity frequency in the population (frequency-dependent selection). The number and frequency of specificities is expected to follow an equilibrium between selection, mutation and drift. In diploid species, specificities are expected to be maintained at equal frequencies—isoplethy, if they are selectively equivalent, and if the population size is infinite (Wright, 1939; see review by Castric and Vekemans, 2004).

Castric and Vekemans (2004) reviewed the empirical studies in natural populations of plant species with GSI (twelve diploid species analyzed, five Solanaceae, three Rosaceae, and one Onagraceae, Papaveraceae and Polemoniaceae species), and found that the expectation of equal frequencies in populations at equilibrium holds in general. Nevertheless, the use of small samples in most studies may be responsible for the failure to reject isoplethy (Stoeckel *et al*., 2008). Indeed, that the specificity frequencies appear to be stochastically equal in populations is surprising (Lawrence, 2000). The time to steady state after specificity frequencies have been perturbed can take many generations, and this time is greater for populations containing a large number of specificities in comparison with those with a small number of specificities (Brooks *et al*., 1996, 1997a, b). Unequal frequencies for different specificities have been reported for all Rosaceae species studied so far (*Crataegus monogyna* (Maloideae), *Prunus lannesiana* and *P. avium* (Prunoideae)) but *Sorbus aucuparia* (Maloideae) (Kato and Mukai, 2004; de Cuyper *et al*., 2005; Schueler *et al*., 2006; Raspé and Kohn, 2002, 2007). Two general alternative hypotheses to account for the observed unequal specificity frequencies have been proposed by these authors: some specificities can be subject to natural selection indirectly by close linkage with other genes, or founder effects. Nevertheless, as recently suggested by Stoeckel *et al*. (2008), the apparent rejection of isoplethy may be due to the noninclusion of genetic drift in the commonly used statistical tests.

There is no theoretical expectation regarding specificity frequencies in polyploid self-incompatible species. Although, polyploid Solanaceae and Scrophulariaceae species usually display a self-compatible phenotype (de Nettancourt, 1977), in *Prunus*, allopolyploid species can be self-incompatible (Hauck *et al*., 2006; Nunes *et al*., 2006). In this genus, self-compatibility is observed only when a minimum of two nonfunctional *S*-haplotypes are present in the same individual (Hauck *et al*., 2006). The

Correspondence: Dr CP Vieira, Molecular Evolution Group, IBMC, University of Porto, Rua do Campo Alegre 823, Porto 4150-180, Portugal. E-mail: cgvieira@ibmc.up.pt
Received 4 March 2008; revised 23 May 2008; accepted 29 May 2008; published online 2 July 2008

issue of isoplethy in natural populations of polyploid self-incompatible species was never addressed from the empirical point of view.

P. spinosa is a nondomesticated allopolyploid species, that is self-incompatible (Salesses, 1970; Nunes *et al*., 2006). This species is thought to be the result of crossbreeding of diploid species *P. cerasifera* and *P. microcarpa* (Eryomine, 1990; Salesses and Bonnet, 1993; Reynders-Aloisi and Grellet, 1994), but varying ploidy levels are described (2*n* varing from 16 to 64; Baiashvili, 1980; Castroviejo- Bolivar *et al*., 1986).

Here, we study the *SFB* gene, that is the pollen component of GSI (Entani *et al*., 2003; Ushijima *et al*., 2003; Yamane *et al*., 2003; Romero *et al*., 2004; Nunes *et al*., 2006). Using a sample of 26 *P. spinosa* individuals from a natural population we find 36 *SFB* alleles. Those 36 alleles are predicted to correspond to 32 different specificities. Each putative specificity differs from all others at amino-acid sites identified as being positively selected. Those sites are likely responsible for determining specificity differences (Nunes *et al*., 2006). The frequency distribution of putative specificities is determined. This work differs from that of Nunes *et al*. (2006), that used the same population, as these authors report the frequencies of 19 likely different specificities using 11 individuals only. Moreover, no effort was made by these authors to systematically search for all specificities present in the population, as done here. Isoplethy is not observed in the *P. spinosa* population studied.

Materials and methods

Plant material and DNA extraction

P. spinosa is a common shrub in all Portugal (Franco, 1984). Leaves from individuals of one wild Portuguese *P. spinosa* population located in Rabal-Bragança (41°52′0N, 6°45′0W, assigned as B) were collected individually and frozen. A total of 26 individuals were sampled along the Sabor river margins (the river is depicted by a continuous black line in Figure 1).

In *P. spinosa* populations, individuals with varying ploidy levels have been described (Baiashvili, 1980). Directly determining ploidy levels in a large number of individuals, in large shrubs such as *P. spinosa*, is difficult, as chromosome counts are usually performed from the root tip meristems (Salesses, 1970). Moreover, in natural populations, chromosome numbers are not always a multiple of the chromosome base number (Baiashvili, 1980). As it is not known in which chromosome the *S*-locus region is located, determining the chromosome number does not necessarily indicate how many *S*-alleles are present in a given individual. Different approaches aimed at fully characterizing the allele set present in the 26 individuals from the Bragança population were thus used (see below). One of the approaches involved, looking for the presence of *SFB* alleles known to be present in a small sample ($N=7$) from the unrelated Santo António- Pêso da Régua population, here characterized (41°10′0N, 7°46′60W, assigned as Re; see Figure 2). As described below, three specificities were found in the Bragança population, only when specific primers based on sequences obtained from Re population were used.

Genomic DNA was extracted from leaves of individual plants using the method of Ingram *et al*. (1997).

SFB PCR amplifications

Primers 62F and 1010R (Supplementary Table 1; Nunes *et al*., 2006) were used to obtain an amplification product of 930 bp. As in the coding region of the *SFB* gene there are no introns, the different alleles of an individual, amplified using this general set of primers, cannot be distinguished on the basis of length differences on an agarose gel. This is in contrast with what is usually done when studying the *S-RNase* gene, where significant intron size variation is often observed between alleles (Vieira *et al*., 2003). Thus, the amplification product was cloned using the TA cloning kit (Invitrogen, Carlsbad, CA, USA). The *Dde*I restriction pattern was analyzed on average for 60–80 colonies. For each restriction pattern, three colonies of the same type were sequenced to correct for possible nucleotide miss-incorporations that may have occurred during the PCR reaction. Sequencing was performed by STABVIDA (Lisbon, Portugal), using BigDye cycle-sequencing kit (PerkinElmer), and specific primers or the primers for the M13 forward and reverse priming sites of the pCR2.1 vector.

Due to chance, *SFB* alleles could have been missed during the screening of clones. Therefore, specific primers were designed for each putative *SFB* specificity

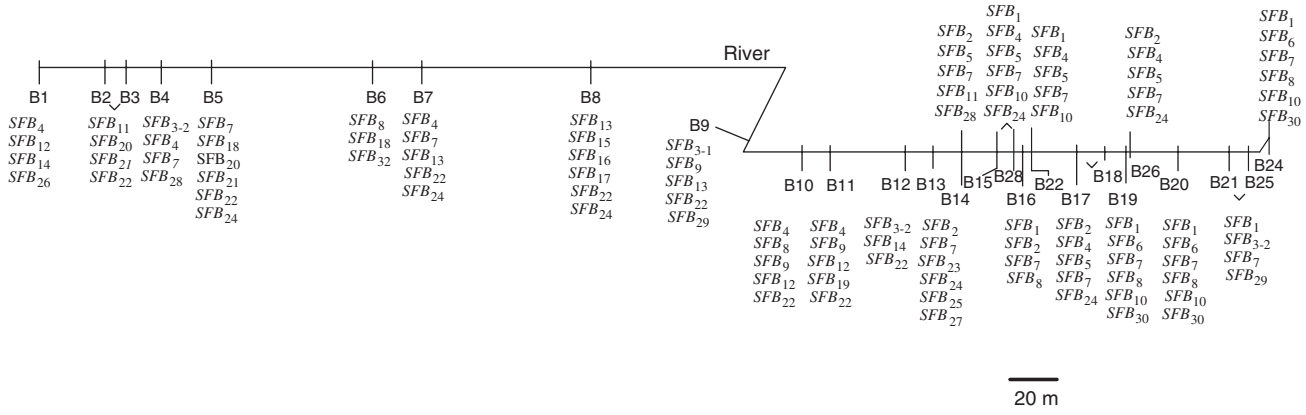

Figure 1 Distribution of the individuals studied of the wild *P. spinosa* Rabal-Bragança population. The zigzag line represents the margin of the Sabor river where all individuals were sampled. For each individual, *SFB* alleles are presented. ˇ—individuals having the same *SFB* alleles.

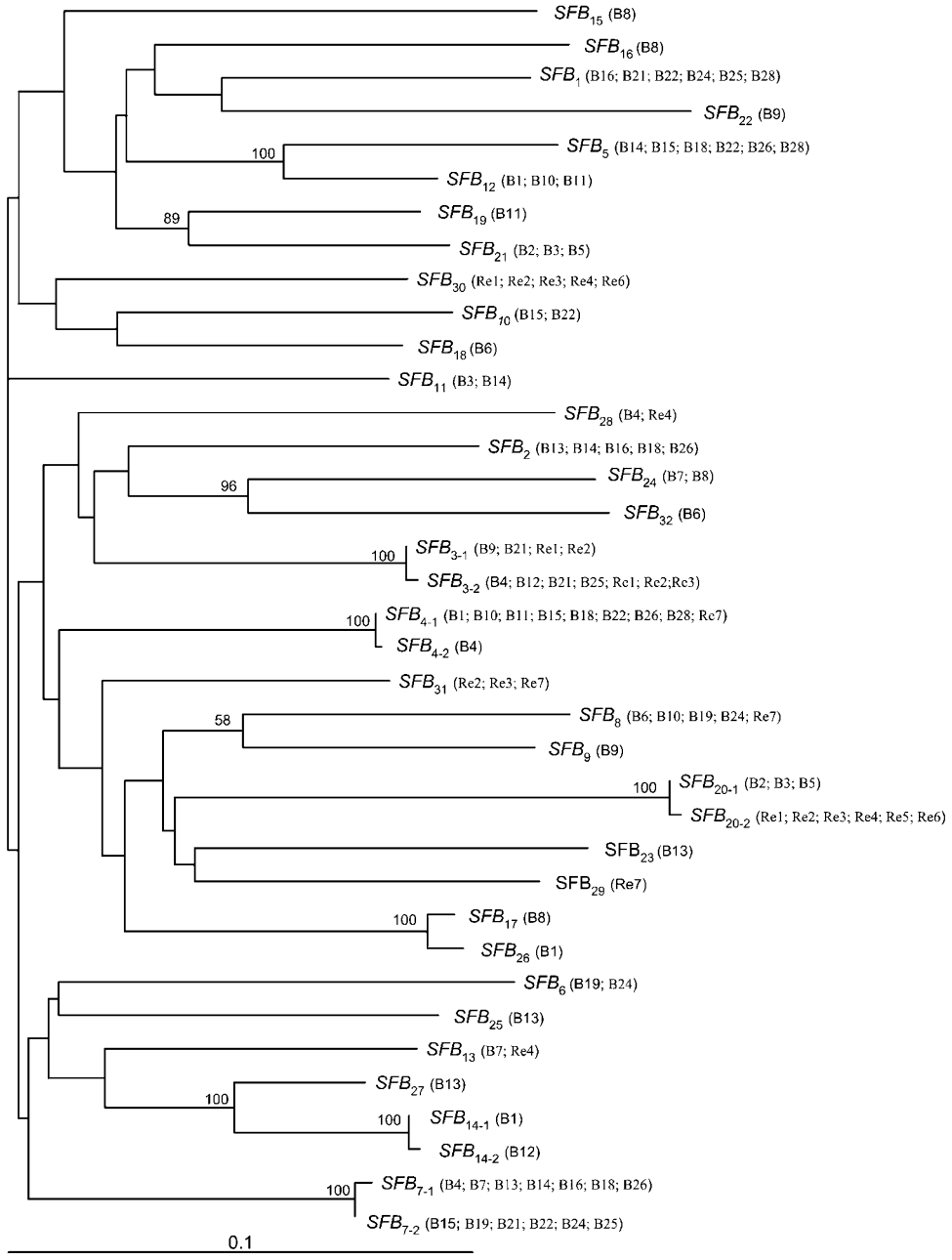

Figure 2 Maximum likelihood phylogenetic tree of *P. spinosa SFB* sequences. Individuals that present identical sequences are indicated in the brackets. B codes for Bragança and Re for Pêso da Régua populations, respectively. Using Modeltest (Posada and Crandall, 1998) the model that best describes the *P. spinosa SFB* data set is determined to be the TVM + G model, with nucleotide frequencies A = 0.27940; C = 0.19720; G = 0.20950; T = 0.31390, the substitution rate matrix A ↔ C = 1.3819; A ↔ G = 3.1999; A ↔ T = 0.9383; C ↔ G = 1.8718; C ↔ T = 3.1999 and G ↔ T = 1.0000, a proportion of invariable sites of 0.0803 and a γ-distribution for the rates at variable sites with shape parameter 0.9095. A molecular clock was not enforced.

(including those found in the Re population; Supplementary Table 1) except for SFB_{3-1}/SFB_{3-2}. The latter two putative specificities were distinguished using primers 1624F (TCTAACCTTCTTCTTTGA) and 2108R (GGGGGTTTTGTTTTTGTG) that amplify the *S-RNase* and *SFB* intergenic region, and digestion with *Sau*3AI. For the 26 individuals of the Bragança population, PCR amplifications were performed using genomic DNA, and each primer combination separately. To confirm the specificity of the primer pair, amplification products not expected in a given individual according to the above-described cloning experiment, were cloned and sequenced. We also used all possible combinations of each specific primer and the appropriate general primer (either the 62F or the 1010R primer; Supplementary Table 1), to complete the *SFB* characterization of the individuals of the Bragança population. Unexpected amplification products, according to the characterization thus far performed were cloned and sequenced.

Analyses of sequences

The DNA sequences were deposited in GenBank (accession numbers EU760896–EU760574 for sequences

obtained with 62F and 1010R; for *SFB* sequences using specific primers, EU760575–EU760619). Here, we use two data sets, namely, *P. spinosa SFB* alleles only, and all *Prunus SFB* alleles available. The nucleotide sequences were aligned using ClustalX v. 1.64b (Thompson *et al*. 1997), and minor manual adjustments were performed using Proseq version 2.43 (http://helios.bto.ed.ac.uk/evolgen/filatov/proseq.html). Analyses of DNA polymorphism were performed using DnaSP 4.1 (Rozas *et al*. 2003) and ProSeq version 2.43 software. Maximum likelihood trees were obtained using PAUP (Swofford, 2002), after running Modeltest (Posada and Crandall, 1998) to find the simplest model of nucleotide sequence evolution that best fits the data (using the Akaike information criterion (AIC)).

Nunes *et al*. (2006) used a set of 37 *Prunus SFB* sequences to identify amino-acid sites under positive selection. Nevertheless, a set of 70 *Prunus SFB* sequences (including those here reported) is now available in GenBank. Therefore, in principle, there is more power to identify amino-acid sites under positive selection. The analyses performed by Nunes *et al*. (2006) were thus repeated using the larger data set now available (Supplementary Table 2). Such information is needed to identify all putative *P. spinosa* specificities (sets of amino-acid sequences that differ from all other amino-acid sequences at least in one site identified as positively selected). The codeml software, as implemented in PAML 3.13 (Yang, 1997), and the method of Wilson and McVean (2006), as implemented in the omegaMap v 0.5 software (www.danielwilson.me.uk), that uses a population genetics approximation to the coalescent with recombination, were used. The model that best describes this data set is TVM + I + G model, with nucleotide frequencies A = 0.2842; C = 0.2215; G = 0.2047 and T = 0.2896, the substitution rate matrix A ↔ C = 1.1161; A ↔ G = 2.8711; A ↔ T = 0.8283; C ↔ G = 1.4636; C ↔ T = 2.8711 and G ↔ T = 1.0000, a proportion of invariable sites of 0.0803 and a γ-distribution for the rates at variable sites with shape parameter 0.9095. A molecular clock was not enforced. When using the Yang (1997) method, of all models tested (M0, M1, M2, M3, M7 and M8) the simplest model that best fits the data is the discrete model with 3 categories (M3). In the method of Wilson and McVean (2006) two 250 000 iterations were performed. A burn-in of 25 000 was also used and the results checked for convergence. All codons were assumed to be at equal frequencies. Ten random sequence orders were used. The parameters to be estimated are the selection parameter ($\omega = K_a/K_s$), the population recombination rate (ρ), the rate of synonymous transversion (μ), the transition–transversion ratio (κ) and the insertion/deletion rate (φ) (Wilson and McVean, 2006). The first two parameters may vary along the sequence. A block of 30 codons (approximately 10% of the sequence size) is used when estimating both ω and ρ. For the Wilson and McVean (2006) method both an objective and a subjective approach to prior specification were used. Inverse distributions were used as priors for ω and ρ, and improper inverse distributions were used for the other parameters (μ, κ and φ). The bounds for ω are 0.01–1000 and for ρ are 0.00000001–1000. Thus the posterior density outside this range should be about 0. In the second approach, exponential distributions were used for all parameters ($\mu = 1$, $\kappa = 3$, $\omega = 1$, $\rho = 0.0001$, $\varphi = 0.2$). The values chosen reflect the prior belief that there is little recombination and that only a few sites along the sequence are positively selected.

To test whether all putative specificities are present at equal frequency in the Bragança population, we used a resampling method, excluding individuals suspected to be the result of vegetative propagation. A total of 50 000 samples were generated with the same number of individuals and ploidy levels as observed in the Bragança sample, under the assumption that the population had a given total number of specificities (varying from 32 to 42) and that all specificities were present at the same frequency in the population. When 33 specificities are assumed, about 33% of the times, the simulated samples show the same number of putative specificities as in the Bragança population sample. If a different number of specificities is assumed, 32 specificities are not observed so often in a sample with the same configuration as that from Bragança population. We then took the 16 387 simulations that, under the assumption of 33 specificities, showed 32 specificities in the simulated sample. On the basis of these simulated samples we calculated the average number of times the commonest, the second commonest and so forth, specificity occurs. The fit of the expected and observed distribution was tested using a χ^2-test after merging the appropriate classes in such a way that for each cell the expected number of events is higher than five.

Results

P. spinosa SFB sequences

Three to six different amplification products per individual (16 Bragança and seven Pêso da Régua) were revealed after the restriction digestion of 60–80 colonies resulting from the cloning of the 930 bp amplification product obtained from genomic DNA (with primers 62F and 1010R). Using tblastx, it can be seen that the 79 sequences obtained share more than 80% of amino-acid identity with described *SFB* alleles from *Prunus* species.

These steps had been already performed for other 11 individuals from the Bragança population analyzed by Nunes *et al*. (2006), and 10 were here used. Therefore 106 nucleotide sequences (the 79 here reported plus 27 reported by Nunes *et al*., 2006), are available for *P. spinosa SFB* gene. The 106 nucleotide sequences define 37 different nucleotide sequences. The phylogenetic relationship of these 37 nucleotide sequences is presented in Figure 2.

There are five phylogenetic groups that show two closely related sequences each (Figure 2). To determine whether they represent the same specificity, controlled crosses must be performed. This is, however, not an easy task, as this species is polyploid, thus multiple controlled crosses and analyses of large numbers of progeny are required. Nevertheless, each of the five pairwise comparisons differs in one (SFB_{3-1} and SFB_{3-2}, SFB_{4-1}–SFB_{4-2} and SFB_{20-1}–SFB_{20-2}) or two (SFB_{7-1}–SFB_{7-2} and SFB_{14-1}–SFB_{14-2}) amino-acid sites. In the pairwise comparisons, none of the differing amino-acid sites is identified as positively selected (see below). In the absence of data on controlled crosses, we thus assume that the amino-acid differences found in the SFB_{4-1}–SFB_{4-2}, SFB_{20-1}–SFB_{20-2}, SFB_{7-1}–SFB_{7-2} and SFB_{14-1}–SFB_{14-2} comparisons represent amino-acid

polymorphism within specificity classes. In agreement with this assumption, for SFB_{7-1}–SFB_{7-2} the corresponding *S-RNase* gene is identical at the amino-acid level in the region analyzed (DQ677586 for SFB_{7-1} and EU833958 for SFB_{7-2}). For SFB_{3-1} and SFB_{3-2} there is, however, evidence to suggest that they do represent two different specificities, as both alleles are amplified from genomic DNA of the same individual (Re1, Re2 and B21; Figure 2). Furthermore, the corresponding *S-RNase* sequences show two amino-acid differences (Nunes *et al*., 2006), and one of these has been identified as a positively selected amino-acid site (Vieira *et al*., 2007). Although, it is unclear, how many amino-acid changes at sites under diversifying selection are required to create a new specificity, we assume that they correspond to two different specificities. The remaining 27 phylogenetic groups (Figure 2) likely represent different specificities, as the number of amino-acid differences between these phylogenetic groups varies from 6 (SFB_{17} and SFB_{26}, $K_a = 0.011$) to 85 (SFB_{13} and SFB_{22}, $K_a = 0.1557$). In all cases, at least one of the amino-acid differences is a site identified as under positive selection (see below). *P. spinosa* thus likely harbors at least 33 different specificities.

To fully characterize the 26 Bragança individuals shown in Figure 1, specific primers were designed for 32 out of 33 putative *SFB* specificities (as described in 'Materials and methods'). PCR amplifications using each of the 32 primer sets were then performed using genomic DNA. All unexpected amplification products were cloned and sequenced (see 'Materials and methods'). The full characterization of the *SFB* alleles in Bragança is presented in Figure 1. Only one primer set did not support amplification in any of the 26 individuals from the Bragança population. Therefore, the *P. spinosa* Bragança population likely harbors at least 32 different specificities.

None of the 32 primer sets tested supports amplification in all individuals, as expected for GSI alleles. In the Bragança population, individuals B2–B3, B15–B28, B17–B18–B26, B19–B20–B24 have the same set of *SFB* alleles.

The number of *SFB* alleles characterized per individual using the 32 sets of primers varies from three to six (Figure 1). *P. spinosa* is described as being an allotetraploid species thus four *S*-alleles were expected per individual (Halliday and Beadle, 1983). Variations in ploidy levels are however described in this species ($2n$ can vary from 16, 24, 32, 40, 43, 44, 48, 53, 56, 59, 64; $n = 16$; Flora Iberica, http://www.rjb.csic.es/floraiberica/PHP/cientificos.php) even in the same population (Baiashvili, 1980).

The combination of 46 specific *SFB* primers and the appropriate general primer (either 72F or 1010R) revealed no further alleles in the individuals studied (data not shown). We take this as suggestive evidence that the inferred number of specificities per individual is correct.

Positively selected sites in the *P. spinosa* SFB sequences

The 70 *Prunus SFB* sequences have been used to identify positively selected sites (Supplementary Table 2), those likely responsible for defining specificity differences. Two different approaches were used, namely, a phylogenetic one (Yang, 1997) and a population genetics method that uses an approximation to the coalescent with recombination (Wilson and McVean, 2006). The two methods were used previously, but with 37 *Prunus* sequences only (Nunes *et al*., 2006). Therefore, it was unclear whether more sites under positive selection would be identified when using many more sequences. Thirty-one amino-acid sites were identified by at least one of the methods used with a posterior probability of selection higher than 95%, and these are shown in Table 1 (see Supplementary Figure 2). Twenty-two of these amino-acid sites were previously identified as being under positive selection by Nunes *et al*. (2006). It should be noted that we cannot be sure that all positively selected amino-acid sites have been identified. Indeed, the power to detect sites under positive selection may be low when sequences are highly diverged (Castric and Vekemans, 2007).

Both the phylogenetic and population genetics approach present potential problems that can affect the identification of sites under positive selection (see Vieira *et al*., 2007). Nevertheless, 18 out of 25 amino-acid sites without alignment gaps were identified by both methods

Table 1 Amino-acid sites under positive selection

Regions	*Amino-acid positions*	*Class*				
		1	*2*	*4*	*5*	*6*
F-box	1–21					
	22–40					
V_1	41–54		43*			
			44^{+}			
	55–228		140		83$^{\#}$	
			156		130	
					131$^{\#}$	
					201	
Vn	229–233	229*		230^{+}		
		231*				
		232*				
		233$^{\#}$				
	234–244					
V_2	245–260		260		248$^{\#}$	
					258$^{\#}$	
	261–277					
HV_a	278–287	285$^{\&}$		281^{+}		
		287$^{\&}$		282^{+}		
				283^{+}		
				284$^{\&}$		
				286*		
	288–297					
HV_b	298–312	299*	306			300$^{\&}$
		301*	308			
		304*	309			
			312			

Class 1: *Phy* and *Pop* > 95%; Class 2: *Phy* > 95% and 50% < *Pop* < 95%; Class 3: 50% < *Phy* < 95% and *Pop* > 95%; Class 4: sites with alignment gaps and *Pop* > 95%; Class 5: *Phy* > 95% and *Pop* < 50%; Class 6: *Phy* < 50% and *Pop* > 95%; Class 7: 50% < *Phy* < 95% and 50% < *Pop* < 95%; where *Phy* and *Pop* stands for the posterior probability of the amino-acid site being positively selected using Yang (1997) phylogenetic approach and Wilson and McVean (2006) population genetics method, respectively.
Amino-acid sites identified using 37 *Prunus SFB* sequences (Nunes *et al*., 2006) are marked as:
*amino acids identified by *Phy* and *Pop*;
#amino-acid sites identified using only *Phy*;
&amino-acid sites identified using only *Pop* and
+sites with alignment gaps detected as being positively selected by *Pop*.

as being positively selected with a posterior probability of selection higher than 50%. It is possible that the six amino-acid sites that are in class 5 ($Phy > 95\%$ and $Pop < 50\%$; Table 1) and the one that is in class 6 ($Pop > 95\%$ and $Phy < 50\%$; Table 1) are false positives. Sites in class 5 could be explained if the phylogenetic relationship between the sequences used was inaccurately derived, or if there is recombination meaning that a single tree cannot accurately describe the relationship of the sequences used (Wilson and McVean, 2006). The Bruen *et al*. (2006) *Phi* test (as implemented in SplitsTree4; Huson and Bryant, 2006), looks for an excess of homoplasies in a data set relative to those expected given the observed divergence. This test performs well in diverse situations such as exponential growth (star-like topologies) and patterns of substitution rate correlation (Bruen *et al*., 2006). The *Phi* test can also be efficiently used to detect recombination between closely, as well as distantly related samples, regardless of the suspected rate of recombination (Bruen *et al*., 2006). When this test is applied to the 70 *SFB* sequences here analyzed, a statistically highly significant excess is observed ($P < 0.00005$). A similar result is obtained when a set of 88 *Prunus S-RNases* compiled by Vieira *et al*. (2007) is used ($P < 0.05$). For *Prunus S-RNases*, other features compatible with recombination have been presented, although the evidence is still not unequivocal (Vieira *et al*., 2003; Nunes *et al*., 2006; Ortega *et al*., 2006). It is tempting to take such an excess of homoplasies as evidence for recombination. Nevertheless, in GSI, many old specificities are maintained by strong frequency-dependent selection. It remains to be determined how many times such an excess is observed when comparable simulated data (without recombination) is used, but this is beyond the scope of this article.

Specificity numbers and frequencies

As *P. spinosa* individuals are polyploid, the total number of specificities present in the population cannot be estimated as done by other authors (Castric and Vekemans, 2004). The 127 alleles sampled in the Bragança population revealed 32 different putative specificities (Figure 1). The other *Prunus* species exhaustively sampled (*P. dulcis*, *P. armeniaca*, *P. salicina* and *P. avium*; see the GenBank data compiled by Vieira *et al*., 2008a) revealed in between 21 and 31 specificities. Furthermore, in two Maloideae species, namely *S. acuparia* and *C. monogyna*, 32 and 17 alleles were found, respectively (Raspé and Kohn, 2002, 2007). It is thus likely that most *P. spinosa* specificities were sampled. Figure 3 shows for the Bragança population, the observed and expected distribution under the assumption of equal frequencies. The two distributions are significantly different ($P < 0.05$). There is an excess of low and high frequency specificities in the Bragança sample (Figure 3). It should be noted, that there is no theoretical expectation regarding specificity frequencies in polyploid self-incompatible species.

Discussion

The thorough search performed in a large sample from the Bragança population revealed 32 putative specificities. Although *P. spinosa* is a polyploid species with varying chromosome numbers, the number of putative specificities found is similar to that found for diploid *Prunus* species (Kato and Mukai, 2004; de Cuyper *et al*., 2005; Schueler *et al*., 2006).

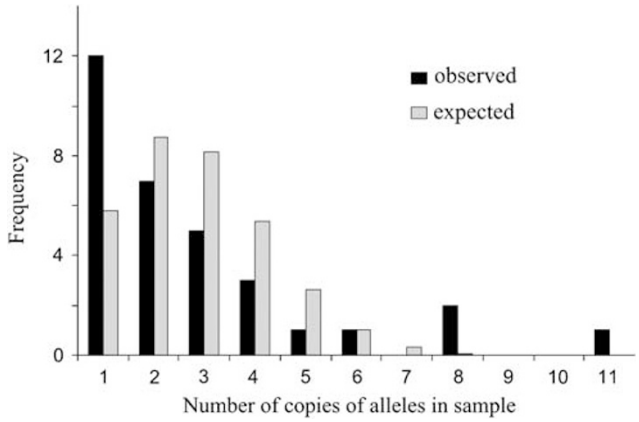

Figure 3 Observed and expected distribution under the assumption of equal specificity frequencies for the Bragança population.

Only 12% of the *P. spinosa* putative specificities show amino-acid polymorphism, even when sequences from different populations are compared (data not shown). This finding is in agreement with that of Nunes *et al*. (2006), that suggested little amino-acid polymorphism within specificities. Such a result is expected, because of the small effective population size of each specificity class (Charlesworth *et al*., 2005). As pointed out by Nunes *et al*. (2006), it is possible that the amino-acid variation could have been inherited from *P. spinosa*'s two diploid ancestors. In addition, the variation observed within specificities could have been obtained by gene flow from *P. insititia* that may hybridize with *P. spinosa* (Nunes *et al*., 2006).

In the *P. spinosa* Bragança population isoplethy is not observed. *P. spinosa* is a polyploid species, and no theoretical expectation regarding specificity frequencies in polyploid self-incompatible species is available. Nevertheless, as this is the first report on specificity frequencies distribution in polyploidy species, it seems useful to discuss possible deviations from population equilibrium.

Identical individuals were found in this sample. Considering the total number of alleles found in the population, and the ploidy level of the individuals, it seems unlikely that two identical individuals would be found in the sample by chance alone. As the maximum physical distance between the identical individuals is less than 11 m (Figure 1), vegetative propagation could be a likely reason for this observation. It should be noted that *P. spinosa* does not show apomixis (Guitian *et al*., 1992), and that although natural vegetative multiplication is rare in the genus *Prunus*, clones of *P. spinosa* can be maintained by tillering (OECD Environment, Health and Safety Publications, 2002). Therefore, individuals suspected to be the result of vegetative propagation were removed from the analyses related to specificity frequency distributions. It should, be noted, nevertheless, that old clonal individuals may have contributed to a putative initial deviation from expected equilibrium frequencies.

In the Bragança population there is an excess of low and high frequency specificities (Figure 3). This type of

deviation is similar to that observed in wild cherry populations, where a significant departure from the isoplethic distribution is also observed (Stoeckel *et al*., 2008). These authors have shown that the observed allele frequency distribution is compatible with genetic drift and a model of subdivided populations with demes linked by moderate migration rate. The way we tested for the expectation of equal frequencies does not include the effect of drift, but this is also a caveat of other approaches commonly used for diploid species (see for instance Castric and Vekemans, 2004).

For polyploid species such as *P. spinosa* it is possible that not all *S*-allele combinations are equally present in pollen grains, as this would imply that every possible chromosome pairing is equally likely during meiosis. In the polyploid *P. cerasus* this, however, seems to be the case (Hauck *et al*., 2006). Moreover, we did not find any obvious association between pairs of alleles.

Furthermore, in *Prunus,* heteroallelic pollen retains its SI phenotype, and two nonfunctional *S*-haplotypes are needed to be present in the same individual for self-compatibility to arise (Hauck *et al*., 2006). Thus, in principle, nonfunctional haplotypes may reach significant frequencies in the population. At the *SFB* gene, no evidence has been found suggestive of nonfunctionality. So far only one nonfunctional haplotype (the $S\text{-}RNase_6$-SFB_6 haplotype that shows a stop codon in the *S-RNase* gene; data not shown) has been found. It should be noted that this haplotype seems to be at low frequency in the Bragança sample. Nunes *et al*. (2006) studied a sample of 10 randomly chosen individuals from Bragança population and found them to be self-incompatible.

The pollen contribution of a large tree may be very different from that of a young small tree. If for any reason, in a local population and for a given age cohort, the frequency of a given specificity increases beyond its predicted equilibrium frequency, then the local population will be out of the equilibrium for many years as large trees and shrubs live for many years. Nevertheless, Schueler *et al*. (2006) studied a large population of German diploid *P. avium* trees with different age cohorts and found that the frequency distribution did not change in the different age cohorts. This can be taken either as evidence that the population is at equilibrium or that, once disturbed, it takes many years for the frequencies to go back to equilibrium. The two main explanations (founder effects and selection at linked loci) that have been put forward for the observed unequal frequencies in diploid Rosaceae species may also, in principle apply. *P. spinosa* is a long-lived perennial species. Because of recent dispersal and establishment of secondary populations, local populations of long-lived species are not expected to be at evolutionary equilibrium (Raspé and Kohn, 2007).

S-alleles with different sets of linked recessive deleterious alleles cannot coexist in the same diploid individual (Uyenoyama, 1997). Therefore, isoplethy may not be observed if *S*-alleles are commonly linked with recessive deleterious alleles. This effect is expected to be more pronounced in species where recombination is suppressed in a large region around the *S*-locus, as thought for Maloideae (Sassa *et al*., 2007). Nevertheless, in *Prunus* the *S*-locus seems to encompass at most three genes (Entani *et al*., 2003; Ushijima *et al*., 2003; Vieira *et al*., 2008b).

More data are needed from natural populations of other polyploid self-incompatible species to establish the generality of the observation made for *P. spinosa*, regarding specificities frequencies distribution.

Acknowledgements

We thank the anonymous reviewers for the constructive criticisms on earlier versions of the article. This work has been partially funded by Fundação para a Ciência e Tecnologia (research projects POCTI/AGG/44800/2002 and POCI/BIA-BDE/59887/2004 funded by POCI 2010, co-funded by FEDER funds).

References

Baiashvili EI (1980). Karyological study of *Prunus spinosa* L. *Bull Georgian Acad Sci* **100**: 645–647.

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.

Brooks RJ, Tobias AM, Lawrence MJ (1997a). A time series analysis of the population genetics of the self-incompatibility polymorphism. 1. Allele frequency distribution of a population with overlapping generations and variation in plant size. *Heredity* **79**: 350–360.

Brooks RJ, Tobias AM, Lawrence MJ (1997b). A time series analysis of the population genetics of the self-incompatibility polymorphism. 2. Frequency-equivalent population and the number of alleles that can be maintained in a population. *Heredity* **79**: 361–364.

Bruen TC, Philippe H, Bryant D (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**: 2665–2681.

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.

Castric V, Vekemans X (2007). Evolution under strong balancing selection: how many codons determine specificity at the female self-incompatibility gene SRK in Brassicaceae? *BMC Evol Biol* **7**: 132.

Castroviejo-Bolivar S, Aedo-Pérez C, Castroviejo S, Laínz M, López González G, Montserrat P *et al.* (1986). Flora Ibérica. In: Castroviejo S, Laínz M, LópesGonzáles G, Montserrat P *et al.* (eds). *Plantas vasculares de la Península Ibérica e Islas Baleares*, vol. 6. Real Jardín Botánico: Madrid, Spain. pp 448–449.

Charlesworth D, Vekemans X, Castric V, Glemin S (2005). Plant self-incompatibility systems: a molecular evolutionary perspective. *New Phytol* **168**: 61–69.

de Cuyper B, Sonneveld T, Tobutt KR (2005). Determining self-incompatibility genotypes in Belgian wild cherries. *Mol Ecol* **14**: 945–955.

de Nettancourt D (1977). *Incompatibility in Angiosperms*. Springer-Verlag: Berlin.

Entani T, Iwano M, Shiba H, Che FS, Isogai A, Takayama S (2003). Comparative analysis of the self-incompatibility (*S*-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic diversity. *Genes Cells* **8**: 203–213.

Eryomine GV (1990). New data on origin of *Prunus domestica* L. ISHS. *Acta Horticulturae* **283**: IV International Symposium on Plum and Prune Genetics, Breeding and Pomology, pp 27–30.

Franco JA (1984). *Nova Flora de Portugal (Continente e Açores)*. Sociedade Astória, Lda.: Lisboa.

Guitian J, Guitian P, Sanchez JM (1992). Reproductive biology of two *Prunus* species (Rosaceae) in the Northwest Iberian Peninsula. *Plant Syst Evol* **185**: 153–165.

Halliday G, Beadle M (1983). *Flora Europaea*. Cambridge University Press: Cambridge.

Hauck NR, Yamane H, Tao R, Iezzoni AF (2006). Accumulation of non-functional S-haplotypes results in the breakdown of gametophytic self-incompatibility in tetraploid *Prunus*. *Genetics* **172**: 1191–1198.

Huson DH, Bryant D (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254–267.

Ingram GC, Doyle S, Carpenter R, Schultz EA, Simon R, Coen ES (1997). Dual role for fimbriata in regulating floral homeotic genes and cell division in *Antirrhinum*. *EMBO J* **16**: 6521–6534.

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.

Lawrence MJ (2000). Population genetics of the homomorphic self-incompatibility polymorphisms in flowering plants. *Ann Bot* **85**: 221–226.

Nunes MD, Santos RA, Ferreira SM, Vieira J, Vieira CP (2006). Variability patterns and positively selected sites at the gametophytic self-incompatibility pollen *SFB* gene in a wild self-incompatible *Prunus spinosa* (Rosaceae) population. *New Phytol* **172**: 577–587.

OECD Environment, Health and Safety Publications (2002). Consensus document on the biology of *Prunus* sp. (stone fruits). JT00124192. ENV/JM/MONO(2002)13. Organisation de Coopération et de Développement Economiques. Electronic publication available at http://www.oecd.org/ehs/.

Ortega E, Boskovic RI, Sargent DJ, Tobutt KR (2006). Analysis of *S-RNase* alleles of almond (*Prunus dulcis*): characterization of new sequences, resolution of synonyms and evidence of intragenic recombination. *Mol Genet Genomics* **276**: 413–426.

Posada D, Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.

Raspé O, Kohn JR (2002). *S*-allele diversity in *Sorbus aucuparia* and *Crataegus monogyna* (Rosaceae: Maloideae). *Heredity* **88**: 458–465.

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

Reynders-Aloisi S, Grellet F (1994). Characterization of the ribosomal DNA units in two related *Prunus* species (*P. cerasifera* and *P. spinosa*). *Plant Cell Rep* **13**: 641–646.

Romero C, Vilanova S, Burgos L, Martinez-Calvo J, Vicente M, Llacer G *et al*. (2004). Analysis of the *S*-locus structure in *Prunus armeniaca* L. Identification of S-haplotype specific S-RNase and F-box genes. *Plant Mol Biol* **56**: 145–157.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.

Salesses G (1970). Études cytologiques chez les *Prunus*. I- Espèces de la section Euprunus. *Ann Amélor Plantes* **20**: 469–483.

Salesses G, Bonnet A (1993). Meiotic behaviour of hybrids between *P. cerasifera*, *P. spinosa* and *P. persica*: an approach to the peach-plum genome relationship. *Cytologia Tokyo* **58**: 257–262.

Sassa H, Kakui H, Miyamoto M, Suzuki Y, Hanada T, Ushijima K *et al*. (2007). *S* locus F-box brothers: multiple and pollen-specific F-box genes with *S* haplotype-specific polymorphisms in apple and Japanese pear. *Genetics* **175**: 1869–1881.

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.

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.

Swofford DL (2002). *PAUP*: Phylogenetic Analysis Using Parsimony (and other methods), ver. 4.0b10*. Sinauer: Sunderland.

Thompson J, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The ClustalX window interface: flexible stategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882.

Ushijima K, Sassa H, Dandekar AM, Gradziel TM, Tao R, Hirano H (2003). Structural and transcriptional analysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. *Plant Cell* **15**: 771–781.

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

Vieira CP, Charlesworth D, Vieira J (2003). Evidence for rare recombination at the gametophytic self-incompatibility locus. *Heredity* **91**: 262–267.

Vieira J, Morales-Hojas R, Santos RAM, Vieira CP (2007). Different positively selected sites at the gametophytic self-incompatibility pistil *S-RNase* gene in the Solanaceae and Rosaceae (*Prunus*, *Pyrus* and *Malus*). *J Mol Evol* **65**: 175–185.

Vieira J, Fonseca NA, Santos RAM, Habu T, Tão R, Vieira CP (2008a). The number, age, sharing and relatedness of S-locus specificities in Prunus. *Genet Res* **90**: 17–26.

Vieira J, Teles E, Santos RAM, Vieira CP (2008b). Recombination at Prunus S-locus region *SLFL1* gene. *Genetics*, submitted.

Wilson DJ, McVean G (2006). Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics* **172**: 1411–1425.

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

Yamane H, Ikeda K, Ushijima K, Sassa H, Tao R (2003). A pollen-expressed gene for a novel protein with an F-box motif that is very tightly linked to a gene for *S-RNase* in two species of cherry, *Prunus cerasus* and *P. avium*. *Plant Cell Physiol* **44**: 764–769.

Yang Z (1997). PAML a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**: 555–556.

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