DOI: 10.1534/genetics.108.087825

Recombination at Prunus *S*-Locus Region *SLFL1* Gene

Jorge Vieira, Eliana Teles, Raquel A. M. Santos and Cristina P. Vieira[1]

Instituto de Biologia Celular e Molecular (IBMC), University of Porto, 4150-180 Porto, Portugal

Manuscript received February 4, 2008
Accepted for publication July 12, 2008

ABSTRACT

In Prunus, the self-incompatibility (*S*-) locus region is <70 kb. Two genes—the *S-RNase*, which encodes the functional female recognition component, and the *SFB* gene, which encodes the pollen recognition component—must co-evolve as a genetic unit to maintain functional incompatibility. Therefore, recombination must be severely repressed at the *S*-locus. Levels of recombination at genes in the vicinity of the *S*-locus have not yet been rigorously tested; thus it is unknown whether recombination is also severely repressed at these loci. In this work, we looked at variability levels and patterns at the *Prunus spinosa SLFL1* gene, which is physically close to the *S-RNase* gene. Our results suggest that the recombination level increases near the *SLFL1* coding region. These findings are discussed in the context of theoretical models predicting an effect of linked weakly deleterious mutations on the relatedness of *S*-locus specificities. Moreover, we show that *SLFL1* belongs to a gene family of at least five functional genes and that *SLFL1* pseudogenes are frequently found in the *S*-locus region.

FLOWERS that express self-incompatibility (SI) can achieve fertilization only when receiving pollen grains that express mating specificities different from their own pistils (de Nettancourt 1977). In species of Prunus, two closely linked *S*-locus genes are involved in the incompatibility recognition response: the *S-RNase* (which encodes the pistil component, a basic glycoprotein with ribonuclease activity; see review by Wang *et al.* 2003) and the *SFB* (the pollen component, a protein with an F-box motif; Entani *et al.* 2003; Ushijima *et al.* 2003).

To maintain functional incompatibility, the *S-RNase* and *SFB* genes must be in linkage disequilibrium, since they must co-evolve as a genetic unit. Nunes *et al.* (2006) showed that the evolutionary histories of these two genes are correlated, although not fully correlated. Evidence suggestive of rare recombination has been found at the *S-RNase* (Vieira *et al.*2003; Ortega *et al.* 2006) and *SFB* (Nunes *et al.* 2006; Vieira *et al.* 2008a) genes. Amino acid sites responsible for specificity differences are scattered throughout the *S-RNase* (Vieira *et al.* 2007) and *SFB* genes (Nunes *et al.* 2006; Vieira *et al.* 2008a). It is conceivable that short gene conversion tracts could cause the observed pattern without disrupting specificity recognition.

For a given specificity, the rate of successful fertilization is inversely related to the specificity frequency in the population (frequency-dependent selection; Wright 1939; Vekemans and Slatkin 1994; Schierup *et al.* 1998; Uyenoyama 2000). Under frequency-dependent selection, many specificities are maintained in populations, and in extant Prunus species this number can be as high as 33 (Vieira *et al.* 2008a). Specificities are also predicted to be maintained for long periods of time. In Prunus, the oldest specificities are estimated to be 15–20 million years old (Vieira *et al.* 2008b). The long-term maintenance of specificities is predicted to lead to high diversity at sites closely linked to the amino acid sites where selection acts (Nordborg *et al.* 1996; Charlesworth *et al.* 1997; Schierup *et al.* 2000; Innan and Nordborg 2003; Wiuf *et al.* 2004). Synonymous variability levels are similar at the *S-RNase* and *SFB* genes. The average per site synonymous divergence (K_s) is 0.241 for the *S-RNase* (Vieira *et al.* 2007) and 0.222 for *SFB* (Nunes *et al.* 2006).

The Prunus *S-RNase* gene is flanked by two functional genes, namely the *SFB* and *SLFL1* genes. The latter gene is thought to define one of the boundaries of the *S*-haplotype-specific region (Ushijima *et al.* 2001, 2003). *SLFL1* is expressed in male organs and also in the style, and it seems not directly involved in the gametophytic self-incompatibility specificity reaction since this gene is deleted in the functional *Prunus avium S3* haplotype (Matsumoto *et al.* 2008). On the basis of six Prunus haplotypes, the physical distance between the *S-RNase* and *SLFL1* gene varies from 6.7 kb to >30 kb (Entani *et al.* 2003; Ushijima *et al.* 2003). This variation is due in part to the presence of the *SLFL1*-related pseudogene and transposable element sequences located between the *S-RNase* and *SLFL1* (Ushijima *et al.* 2001, 2003; Entani *et al.* 2003). *SLFL1* seems to always have the same transcriptional orientation as the *S-RNase* gene (Ushijima *et al.* 2001, 2003; Entani *et al.* 2003).

The deduced *SLFL1* amino acid sequence associated with two different *Prunus dulcis* specificities is 95.1%

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. EU876704–EU876742.

[1]*Corresponding author:* IBMC, Molecular Evolution, Group Rua do Campo Alegre 823, 4150-180 Porto, Portugal.
E-mail: cgvieira@ibmc.up.pt

identical (K_s = 0.0795; Ushijima *et al.* 2003; data not shown). The same pattern is observed in *Prunus mume*, where the *SLFL1* amino acid sequence associated with two distinct specificities has been shown to be 92.5% identical (K_s = 0.0722; Entani *et al.* 2003; data not shown). Ten other pairs of genes located in the vicinity of the *S*-locus region show >97.2% amino acid identities (Entani *et al.* 2003). *P. mume* nucleotide sequences are available for two genes in this region that are farther away from the *S-RNase* than *SLFL1*, namely *SLFL2* and *SLFL3*. K_s values are 0.0263 (N = 3) and 0.0039 (N = 2), respectively (data not shown). These values are lower than that for *SLFL1*. Therefore, it is unclear whether there is recombination suppression at the *SLFL1* gene and to what extent.

Suppression of recombination leads to the accumulation of weak deleterious mutations in the *S*-locus region (reviewed in Uyenoyama 2005). Closely related specificities may share the same weak deleterious mutation due to recent common ancestry (Uyenoyama 1997). Therefore, it is predicted that in natural populations there should be a bias against closely related specificities. This effect should be more evident in species where suppression of recombination affects a large region around the *S*-locus. There are, however, other theoretical reasons to expect a bias against closely related specificities. For example, when a new specificity arises, it is expected that it will replace the specificity that gave origin to it, although it is conceivable that the original specificity may be brought back to the population by migration from another population (Uyenoyama *et al.* 2001). Thus far, in Prunus, the evidence for a bias against closely related specificities is ambiguous (Vieira *et al.* 2008b).

Within specificities, background selection against deleterious mutations at *S*-locus-linked genes leads to a reduction in effective population size (Charlesworth *et al.* 1993). As a consequence, little variability is expected within specificities, as it is observed in Prunus (Nunes *et al.* 2006; Ortega *et al.* 2006; Vieira *et al.* 2008a). Nevertheless, little variability is expected even in the absence of such an effect (Nunes *et al.* 2006).

Polymorphism data for genes located in the vicinity of the *S*-locus as well as data on reference loci, as reported in this work, can shed light on the size around the *S*-locus region where recombination is suppressed and to what extent it is suppressed. Only then is it possible to understand the impact of weakly deleterious *S*-locus-linked mutations on the evolution of the genes determining gametophytic self-incompatibility specificities in Prunus. Here, we estimate recombination levels at the *SLFL1*, *S-RNase*, and *SFB* genes, as well as in the region between the *S-RNase* and *SLFL1* genes.

MATERIALS AND METHODS

Plant material and DNA extraction: *Prunus spinosa*, a self-incompatible species (Salesses 1973; Nunes *et al.* 2006), is a dense shrub abundant in Europe (Halliday and Beadle 1983). Although *P. spinosa* has been described as being an allotetraploid species (Halliday and Beadle 1983), variations in ploidy levels have been reported ($2n$ = 16, 24, 32, 40, 43, 44, 48, 53, 56, 59, or 64; Flora Iberica, http://www.rjb.csic.es/floraiberica/PHP/cientificos.php) even at the local population level (Baiashvili 1980).

Leaves were collected from the individuals of the *P. spinosa* population Rabal–Bragança (assigned as B) described by Nunes *et al.* (2006) and Vieira *et al.* (2008a). Genomic DNA was extracted from leaves of individual plants using the method of Ingram *et al.* (1997).

***SLFL1* PCR amplification:** On the basis of the available *SLFL* sequences (Ushijima *et al.* 2001, 2003; Entani *et al.* 2003), primers 44F and 800R (supplemental Table 1) were designed to amplify the *SLFL1* gene but not the *SLFL2*, *SLFL3*, and *SFB* genes. Genomic DNA of individuals B8, B10, B15, and B18 was used as template. Standard amplification conditions were 35 cycles of denaturation at 94° for 30 sec, primer annealing at 48° for 30 sec, and primer extension at 72° for 2 min. The 770-bp amplification product (the expected size) was cloned using the TA cloning kit (Invitrogen, Carlsbad, CA). For each individual, on average, the restriction pattern of the insert of 80 colonies was analyzed using *Rsa*I and *Sau*3AI restriction enzymes. For each individual and restriction pattern, three colonies were sequenced to obtain a consensus sequence. The ABI PRISM BigDye cycle-sequencing kit (Perkin Elmer, Foster City, CA) and specific primers or the primers for the M13 forward and reverse priming sites of the pCR2.1 vector were used to prepare the sequencing reactions. Sequencing runs were performed by STABVIDA (Lisbon).

Nucleotide sequence chimeras can be obtained during the PCR reaction, and these could influence the interpretation of the results presented here. Nevertheless, care was taken to eliminate the chimeras by performing the entire procedure (including the PCR reaction) twice. Furthermore, 80 colonies were screened using restriction enzymes. Those colonies showing very rare restriction patterns were discarded. Finally, for each individual, all possible pairs of consensus sequences were inspected to make sure that nucleotide differences were not clustered, as expected, if one of the two sequences used in the comparison is a chimera.

***SLFL4* PCR amplification:** Analysis of the cloned PCR fragments obtained in the previous section revealed two new *SLFL* genes (see RESULTS). Polymorphism studies were performed using genomic DNA from individuals B2, B4, B6, B10, B14, B15, and B19 and the *SLFL4*-specific primers B8-8/B18-4 (supplemental Table 1). We used the same PCR, cloning, and sequencing approaches as for *SLFL1* (see above).

Characterization of extended haplotypes: In both *P. dulcis* and *P. mume*, the *SLFL1* and *S-RNase* genes have the same transcription direction (Entani *et al.* 2003; Ushijima *et al.* 2003). Therefore, we used the *P. spinosa S-RNase* sequences S_1, S_4, S_7, S_8, S_9, S_{10}, and S_{15} (GenBank accession nos. EF36467, EU878543, EU833958, DQ677587, DQ677588, DQ677589, and EF636468) present in individuals B8, B9, B10, B15, B18, and B19 (Nunes *et al.* 2006; Vieira *et al.* 2008a; supplemental Table 2) to design *S-RNase*-specific reverse primers (supplemental Table 1). Each specific *S-RNase* reverse primer was used in combination with the general *SLFL1* forward primer 44F (or 104F in the case of the S_{10} haplotype). These PCR amplifications were performed using system 3 protocol of the Roche Expand long template PCR system (Roche). In all cases, the amplification product was cloned using the Topo XL PCR cloning kit (Invitrogen). To obtain consensus sequences, for each cloning experiment three colonies were sequenced using primers M13F, M13R (for the vector arms), 44F, and 800R (primers designed for the *SLFL1* gene) and 42F and 93F (primers designed for the *S-RNase* gene; Nunes *et al.* 2006).

Testing associations between *SLFL1* and *SFB* alleles: The *SFB* alleles present in 20 different individuals from the *P. spinosa* Rabal–Bragança population have been reported by Vieira *et al.* (2008a). Therefore, to look for co-occurrence of *SFB* and *SLFL1* alleles, specific primers were designed for *SLFL1* alleles B18-2, B8-4, B15-3/B18-3, and B8-6/B15-4/B18-1 (supplemental Table 1). Standard PCR amplification conditions were used (see above).

Summary statistics: DNA sequences were deposited in GenBank (accession nos. EU876704–EU876742). Amino acid 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). Nucleotide sequences were aligned using the amino acid alignment as a guide. Analyses of DNA polymorphism, linkage disequilibrium, and the minimum number of recombination events were performed using DnaSP 4.1 (Rozas *et al.* 2003) and ProSeq version 2.43 software. Standard coalescent simulations constrained on the number of segregating sites and sample size were also performed using DnaSP 4.1 (Rozas *et al.* 2003).

Phylogenetic analyses of *SLFL* genes: Due to computational burden, the phylogenetic relationship of the 35 *SLFL* sequences used was obtained using minimum evolution. The tree was obtained using complete deletion and Jukes–Cantor-corrected nucleotide distances, as implemented in the MEGA software (Kumar *et al.* 2004). Bootstrap values were obtained using 500 replicates.

Testing for selection at *SLFL1*: Both the codeml software implemented in PAML 3.13 (Yang 1997) and the method of Wilson and McVean (2006) as implemented in the omegaMap v 0.5 software (http://www.danielwilson.me.uk) that uses a population genetics approximation to the coalescent with recombination were used.

When using PAML 3.13, 17 different Prunus *SLFL1* sequences were used. The specified *SLFL1* maximum-likelihood tree was obtained with PAUP (Swofford 2002) after using Modeltest (Posada and Crandall 1998) to find the simplest model of nucleotide sequence evolution that best fit the data, according to the Akaike information criterion (a $GTR + G$ model with nucleotide frequencies $A = 0.27750$, $C = 0.15360$, $G = 0.22540$, and $T = 0.34350$, the substitution model $A \leftrightarrow C = 1.5091$, $A \leftrightarrow G = 2.4386$, $A \leftrightarrow T = 0.6374$, $C \leftrightarrow G = 1.7718$, $C \leftrightarrow T = 1.5168$, $G \leftrightarrow T = 1.0000$, and a gamma distribution with a shape parameter α of 0.5812). Simple models that allow for positive selection [Yang's (1997) M2 (three categories, one of them with a $K_a/K_s > 1$) and M3 models (three categories)] are as likely as a model not allowing for positive selection (model M1).

When using omegaMap, since it uses a population genetics approximation to the coalescent with recombination, we used only *P. spinosa SLFL1* sequences ($N = 15$). A total of 250,000 iterations and a burn-in of 25,000 were used in all analyses. All codons were assumed to be at equal frequencies. Ten random sequence orders were used. The parameters to be estimated were 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 ($\sim$10% of the sequence size) is used when estimating both ω and ρ. One objective and one subjective approach to prior specification were used. First, 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 were 0.01–1000 and for ρ, 0.00000001–1000. Thus the posterior density outside this range should be about zero. When the bound for ω was 0.0001–10 rather than 0.01–1000, similar parameter estimates were obtained (data not shown). Starting values for μ, κ, and φ were, respectively, 0.1, 1, and 1. In the second approach to prior specification, exponential distributions were used for all parameters (starting values were $\mu = 0.1$, $\kappa = 1$, $\omega = 1$, $\rho = 0.001$, and $\varphi = 1$). Similar parameter estimates were obtained regardless of the approach used. Strong evidence for positively selected sites (posterior probability values >95%) was never obtained.

Amino acid variability levels along the SLFL1 protein: Normed variability indices for each site of the *SLFL1*gene were calculated as in Kheyr-Pour *et al.* (1990). Thus, for each site, information on both the number of different amino acids as well as their frequencies was used.

Estimating the population recombination rate at the *SLFL1* gene: The Wilson and McVean (2006) method allows coestimation of the ratio K_a/K_s and the population recombination rate. Estimates are given for every codon, but we used information on codon positions 31–217 only. The first 30 and last 30 codons (the size of the block used) were discarded since estimates at the edges of the sequence can be inaccurate when using a block approach (Wilson and McVean 2006). For the interval considered, the two runs using different prior specification approaches (see above) converged to the same estimate of the population recombination rate. Therefore we performed a joint analysis of the two runs to obtain a point estimate for the population recombination rate at *SLFL1* as well as the 95% credibility intervals.

Estimating the relative importance of recombination and mutation at the *SLFL1*, S-*RNase*, and *SFB* genes: The following Prunus data sets were used (every DNA sequence in the data sets is unique): 17 *SLFL1* sequences, the 88 *S-RNase* sequences used by Vieira *et al.* (2007), and the 70 *SFB* sequences used by Vieira *et al.* (2008a). To infer the number of independent recombination events implied by each DNA sequence data set, the recombination detection program RDP (Martin *et al.* 2005) was used. The following methods, with default options, were selected: RDP, Chimaera, BootScan, 3Seq, GeneConv, MaxChi, and SiScan. A sequence was taken as recombinant if at least one of the methods used identified a recombination tract in that sequence with a probability <0.05. The number of inferred independent recombination events was often smaller than the number of sequences showing evidence for recombination tracts, since inferred recombination events may be old and thus may be apparent in several descendant sequences. For each data set, the total number of synonymous mutations implied by the data was inferred using Yang's (1997) methodology under the appropriate model [M0 (this work), M3 (Vieira *et al.* 2007), and M3 (Vieira *et al.* 2008a) for the *SLFL1*, *S-RNase*, and *SFB* data sets, respectively].

Estimating the population recombination rate between *SLFL1* and the *S*-locus: Estimates of the population recombination rate between *SLFL1* and the *S*-locus were obtained using the formulas given by Kamau and Charlesworth (2005).

RESULTS

***P. spinosa SLFL1*-like genes:** For individuals B8, B10, B15, and B18, a 770-bp amplification product obtained with primers 44F and 800R that support the amplification of the *SLFL1* gene (covering 63% of this gene) was cloned. For these individuals, the amplification product revealed 9, 4, 4, and 7 *SLFL1*-like sequences, respectively. Although ploidy levels vary in *P. spinosa*, no more than 6 alleles at the *S*-locus have been described for these individuals (supplemental Table 2; Nunes *et al.* 2006;

TABLE 1

Nucleotide identity of the 16 different *P. spinosa SLF1*-like sequences and *P. avium SLFL1-S4* allele

Sequence types	% nucleotide identity with the *P. avium SLFL1-S4* haplotype (AB280953)
B8-1/B10-3	97
B15-1	95
B8-2	95
B18-6	95
B8-6/B15-4/B18-1	97
B8-9	96
B15-3/B18-3	96
B10-1	95
B8-5/B10-2	89
B18-7/B15-2	94
B8-4	88
B18-2	88
B8-3	78
B8-7/B18-5	78
B10-4	78
B8-8/B18-4	78

Vieira *et al.* 2008a). Thus the amplification of >6 *SLFL1*-like sequences in two individuals (B8 and B18) indicates that other *SLFL* genes are being amplified. The 24 sequences define 16 different nucleotide sequences (Table 1). Blastn searches revealed that 12 of them show >87% nucleotide identity with Prunus *SLFL1* sequences. On the other hand, the remaining four sequences (B8-3, B8-7/B18-5, B8-8/B18-4, and B10-4) showed ~78% nucleotide identity with previously described Prunus *SLFL1* sequences. Specific primers were designed for sequences B8-3 and B8-8/B18-4 (supplemental Table 1). For both primer sets, an amplification product with the expected size was obtained when genomic DNA from each of the 20 different *P. spinosa* individuals of the Bragança natural population (for which *SFB* alleles are known; Vieira *et al.* 2008a) was used. Specificity of the PCR reaction was confirmed by sequencing the amplification products (data not shown). Since these *SLFL1*-related sequences are amplified in all individuals, it is unlikely that they are *SLFL1* alleles, as this would imply that all individuals have the same two divergent *SLFL1* alleles. Since both sequences are present in all individuals analyzed, and the synonymous nucleotide divergence between both types is high ($K_s = 0.2901$), it is likely that they represent two different genes. Therefore, we named *SLFL4* the B8-8/B18-4 type of sequence and *SLFL5* the B8-3 type of sequence.

***S*-locus region *SLFL1* pseudogenes:** Two *SLFL1* sequence types (B18-6 and B18-7/B15-2) have multiple in-frame stop codons. *SLFL1* pseudogenes have been described in the *P. mume S*1 haplotype (Entani *et al.* 2003), although the corresponding nucleotide sequences are not available in the public databases. One of the pseudogenes is located between the *SLFL1* and the *S-RNase* gene (Entani *et al.* 2003). To estimate how frequent this situation is in Prunus, we determined the sequence of the *SLFL1*-like gene that is closest to the *S-RNase*, as well as a fragment of the *S-RNase* gene (data not shown), for seven *P. spinosa S* haplotypes (S_1, S_4, S_7, S_8, S_9, S_{10}, and S_{15}; see materials and methods and supplemental Table 2).

For all *S* haplotypes analyzed here, with the exception of the *S*9 haplotype, the SLFL1 primers used allow amplification of a region that is ~770 bp long. The *SLFL1*-like sequence amplified from the *S*9 haplotype is longer (1183 bp). Putative splicing sites can be found around the 413-bp insertion. When the putative intron is removed, the protein encoded by this gene is six amino acids longer than all other SLFL1 proteins. Nevertheless, *SLFL1* has been described as being an intronless gene (Entani *et al.* 2003; Ushijima *et al.* 2003). Therefore, it is likely a pseudogene. The *SLFL1* sequence obtained from the S_4, S_{10}, and S_{15} haplotypes show multiple in-frame stop codons. Thus, in four of the seven cases here studied, the neighbor of the *S-RNase* gene is an *SLFL1* pseudogene rather than the *SLFL1* functional gene.

The *SLFL1* sequence from the S_4 haplotype is identical to one of the sequences obtained from B15 and B18 individuals. These individuals have been shown to have the *S*4 haplotype (supplemental Table 2).

Phylogenetic analyses: The phylogenetic relationship of Prunus *SLFL1*, *SLFL2*, *SLFL3*, *SLFL4*, and *SLFL5* gene sequences is presented in Figure 1. All Prunus *SLFL1* sequences cluster together with high bootstrap value (Figure 1). *SLFL1* pseudogenes are found mingled with functional *SLFL1* sequences. In two cases, *SLFL1* pseudogenes cluster with functional *SLFL1* sequences with high bootstrap value (>91%). The *SLFL4* and *SLFL5* genes are more closely related than either to the other *SLFL* genes. *SLFL4*/*SLFL5* are more closely related to *SLFL1* than to *SLFL2* and *SLFL3* (Figure 1). *SLFL2* is shown as an out-group to the other *SLFL* genes.

Individuals B8 and B10 have two identical *SLFL1* sequences (B8-1/B10-3 and B8-5/B10-2), but they share allele *SFB22* only. Although the two types of sequences do not show in-frame stop codons in the region analyzed, it is likely that one of the sequence types is from a pseudogene (see *S-locus region SLFL1 pseudogenes*). Pseudogenes are known to evolve faster than functional copies. To test this hypothesis, we performed Tajima's relative rate tests using as the out-group one *SLFL4* sequence and, as one of the in-groups, the *P. avium SLFL1* sequence. Significant results were obtained for sequences B8-5/B10-2. We extended these analyses to all other *SLFL1* sequences with no in-frame stop codons. Only when sequence B10-1 was used was a significant result obtained. B10-1 and B8-5/B10-2 sequences cluster together with 100% bootstrap support (Figure 1).

Evidence for historical recombination at *SLFL1*: Table 2 shows the per-site synonymous polymorphism

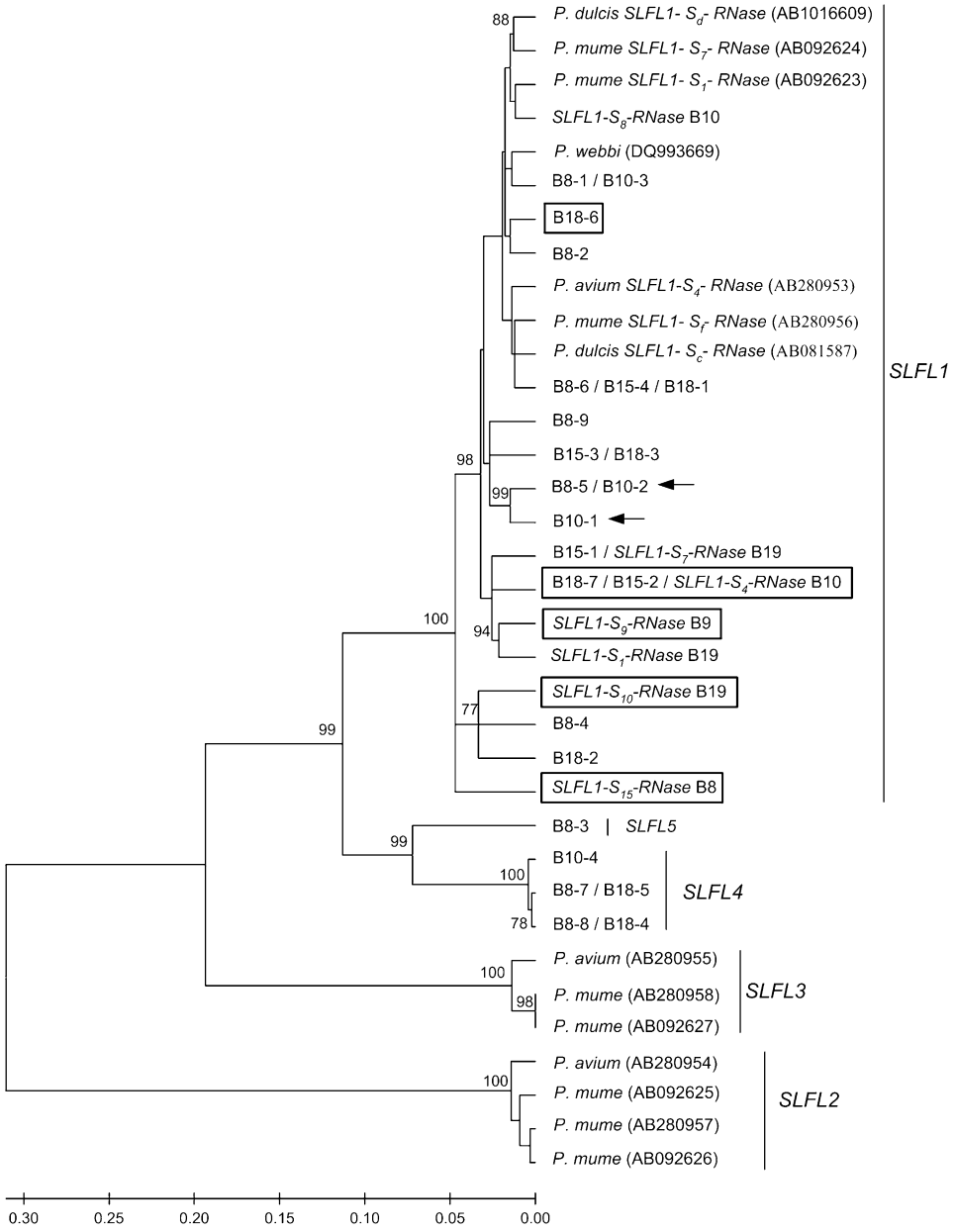

FIGURE 1.—Linearized rooted minimum evolution tree showing the relationship of Prunus *SLFL* genes. Bootstrap values >70% are shown. Sequences that present in-frame stop codons or that have intron-like insertions are boxed. These sequences are believed to be pseudogenes; arrows indicate sequences that show a significant Tajima's relative rate test.

for Prunus *SLFL1* gene sequences as well as for *P. spinosa*, *P. mume*, and *P. dulcis* sequences. All putative *SLFL1* pseudogene sequences have been removed from the analyses. For comparison, levels of polymorphism at the *P. spinosa SLFL4* gene are also shown. There is no evidence to support the view that patterns of polymorphism at *SLFL4* are influenced by the *S*-locus, and thus this gene is used here as a reference locus. In *P. mume*, ENTANI *et al.* (2003) did not find this gene when sequencing a region of ~32 and 22 kb at the left and right of the *S*-locus, respectively. For *P. spinosa*, the synonymous variability level is about four times higher at *SLFL1* than at *SLFL4*. This observation suggests that variability patterns at *SLFL1* are being influenced by the neighboring *S-RNase–SFB* genes. When using all individuals of the Bragança population and specific primers for four *SLFL1* alleles (B18-2, B8-4, B15-3/B18-3, and B8-6/B15-4/B18-1; supplemental Table 1), complete co-occurrence of *SFB* and *SLFL1* alleles was indeed observed (Table 3). Furthermore, when using long PCR, a specific primer for a given *S-RNase* allele and a *SLFL1* general primer, a fragment with the same size was obtained from all individuals of the Bragança population known to have that particular *S-RNase* allele (Table 3).

Variability levels at *P. spinosa SLFL1* (the larger sample) are on the same order as divergence between Prunus species from the Prunus and Amygdalus subgenera (Table 4). Species of these two subgenera shared a common ancestor ~2.5 million years ago (VIEIRA *et al.* 2008b). Therefore, this observation indicates that *SLFL1* alleles are on average ~2.5 million years old. That most variability predates Prunus speciation is indicated by the

TABLE 2

DNA sequence variation summary

	SLFL1				*SLFL4*:
	All Prunus ($N = 22$)	*P. spinosa* ($N = 15$)	*P. mume* ($N = 3$)	*P. dulcis* ($N = 2$)	*P. spinosa* ($N = 12$)
Silent π JC	0.09340	0.10532	0.05893	0.07880	0.02639
R_m	16	14	—	—	2
4GT	673/15931	599/11325	—	—	6/72
LD	1108 (1)/15931	663 (0)/11325	—	—	5 (1)/15
ZnS	0.1113	0.1744	—	—	0.3466
ϕ-test	$P < 5 \times 10^{-7}$	$P < 5 \times 10^{-9}$	—	—	$P = 0.085$

N, number of sequences used; π, average number of pairwise nucleotide differences per base pair Jukes–Cantor corrected (Nei 1987); R_m, minimum number of recombination events (Hudson and Kaplan 1985); 4GT, number of pairwise comparisons presenting the four gametic types over the total number of all pairwise comparisons; LD, pairs of sites showing significant linkage disequilibrium using Fisher's exact test (in parentheses after Bonferroni correction for multiple comparisons) over the total number of all pairwise comparisons ; *ZnS*, average of R^2 values over all pairwise comparisons (Kelly 1997); ϕ-test, probability of observing the inferred nucleotide homoplasies under the assumption of no recombination, as implemented in SplitsTree4 (Huson and Bryant 2006).

high number of shared variants and the low number of fixed differences between Prunus species (Table 4).

The high variability at the *SLFL1* gene compared to *SLFL4*, taken above as evidence for restricted recombination between *SLFL1* and the *S*-locus, could be due to diversifying selection acting on the *SLFL1* gene. We tested for this possibility using two different approaches (Yang 1997; Wilson and McVean 2006; see materials and methods). 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).

When using Yang's (1997) approach for detecting amino acid sites under positive selection, of all models tested, the simplest model that fits the data is model M1 that does not consider a positively selected class (see materials and methods). Two amino acid positions (47 and 247; at these sites there are four and five different amino acids, respectively; Figure 2) have, on average, posterior probabilities of selection >50% when using omegaMap. Nevertheless, under no condition did any of these sites show strong evidence for positive selection (posterior probability values >95%). Therefore, the hypothesis that diversifying selection is acting on the *SLFL1* gene itself cannot be ruled out, although it seems very unlikely.

Despite the evidence for specific associations between *SLFL1* and *S-RNase–SFB* genes as well as the old age of *SLFL1* alleles, there is ample evidence suggestive of recombination at the *SLFL1* gene (Table 2). For example, the minimum number of recombination events (Hudson and Kaplan 1985) implied by the 22 *SLFL1* sequences is 16. Furthermore, despite the relatively small sample size, 4.2% of all pairwise comparisons show all four gametic types and only 6.9% of all possible pairs of sites show significant linkage disequilibrium (only one pair gives a significant result if the sequential Bonferroni correction is applied). The overall linkage disequilibrium, as measured by Kelly's (1997) *ZnS* statistic, is relatively low (varying from 0.1113 to 0.1744). Nevertheless, for *P. spinosa* (the larger sample), standard coalescent simulations show that it is likely to obtain the observed value under the assumption of no recombination ($P > 0.05$) although the simulations performed do not incorporate

TABLE 3

Amplification products that show associations with a particular *SFB* allele

SLFL1 allele or haplotype	Individuals that show the expected amplification product	*SFB* allele
B18-2	B13, B14, B16, B18, B26	SFB_2
B8-4	B8	SFB_{15} or SFB_{16} or SFB_{17} or SFB_{18}[a]
B15-3/B18-3	B14, B15, B18, B22, B26	SFB_5
B8-6/B15-4/B18-1	B5, B7, B8, B13, B15, B17, B26	SFB_{24}
SLFL1-S_8-*RNase*B10	B6, B10, B16, B19, B24	SFB_8
SLFL1-S_1-*RNase*B19	B15, B16, B19, B21, B22, B24, B25, B28	SFB_1

[a] These four *SFB* alleles appear only in the B8 individual (Nunes *et al.* 2006; Vieira *et al.* 2008a).

TABLE 4

Synonymous divergence (Jukes–Cantor corrected) at the *SLFL1* gene (above the diagonal) and number of fixed/shared polymorphisms (below the diagonal)

	P. spinosa	*P. mume*	*P. dulcis*
P. spinosa	—	0.08665	0.08850
P. mume	0/10	—	0.06474
P. dulcis	1/12	0/7	—

the effect of the neighboring *S-RNase–SFB* genes. The phylogenetic ϕ-test for recombination (Huson and Bryant 2006) gives a strong indication for recombination (Table 2), as do other tests for recombination (see results, *The relative importance of recombination and mutation at the SLFL1, S-RNase, and SFB genes*).

The 10 Prunus *SLFL1* alleles known to be associated with a given *S-RNase* allele (6 in *P. spinosa*, 2 in *P. mume*, and 2 in *P. dulcis*) can be used to test whether the evolutionary histories of the two genes are correlated. To test this prediction, per-site synonymous (K_s) values were calculated for the *S-RNase* pairwise comparisons and for the corresponding *SLFL1* pairwise comparisons. A nonsignificant correlation was obtained between synonymous divergence values at the two genes ($r = 0.026$; $P > 0.05$ Spearman nonparametric correlation). Therefore, it seems likely that, historically, the region where the *SLFL1* gene is located has experienced non-negligible levels of recombination. Nevertheless, when the partition-homogeneity test was performed using only variable sites (to correct for a possible effect of different variability levels at the *SLFL1* and *S-RNase* genes; Cunningham 1997), as implemented in PAUP (1000 replicates), a *P*-value of 0.049 was obtained. This value is too high to safely conclude that the *SLFL1* and *S-RNase* tree topologies are significantly different (Cunningham 1997). This, however, can be the result of a small sample size and lack of definition of the two topologies.

When using the Wilson and McVean (2006) approach and the *P. spinosa* random sample of 15 sequences, an average (and standard deviation) per-codon point estimate of 0.198 ± 0.063 is obtained for the population recombination rate at *SLFL1*. The lower and higher 95% credible intervals are, respectively, 0.082 ± 0.043 and 0.505 ± 0.150. The SLFL1 protein is ~409 amino acids long. Therefore, the *SLFL1* gene is expected to experience ~81 recombination events per generation (0.07 recombination events between adjacent nucleotides per generation). Nevertheless, the approach here used assumes a panmictic population. Patterns of variability at the *SLFL1* gene may be influenced by the *S*-locus. Thus, patterns of variability at *SLFL1* may look like those expected for a subdivided population. Therefore, the recombination estimate presented here may be an overestimate.

The relative importance of recombination and mutation at the *SLFL1*, *S-RNase*, and *SFB* genes: The number of inferred independent recombination events is 3, 9, and 15 for the *SLFL1*, *S-RNase*, and *SFB* data sets used here, respectively. Furthermore, 136, 679.2, and 1367.9 synonymous mutations are implied by the *SLFL1*, *S-RNase*, and *SFB* data sets, respectively. Therefore, there are 0.022, 0.013, and 0.011 recombination events per synonymous mutation for the *SLFL1*, *S-RNase*, and *SFB* genes, respectively. This calculation suggests that the recombination rate at the *SLFL1* gene may be only twofold higher than that at the *S-RNase* and *SFB* genes. Nevertheless, the power to detect recombinant sequences may depend on sample size and variability levels.

Estimating the population recombination rate between *SLFL1* and the *S-RNase*: For *P. spinosa*, using a simplified model that does not explicitly incorporate the effect of selection at the neighboring *S*-locus but rather approximates it by assuming two alleles held at intermediate frequencies (Kamau and Charlesworth 2005), an estimate of 0.33 was obtained for the population recombination rate between *SLFL1* and the *S-RNase*. Nevertheless, for *P. spinosa*, when a model that explicitly models selection at the neighboring *S*-locus is used, a much higher estimate (9.04) is obtained for the population recombination rate between *SLFL1* and the *S-RNase*. For the calculations, we used *SLFL4* as a reference

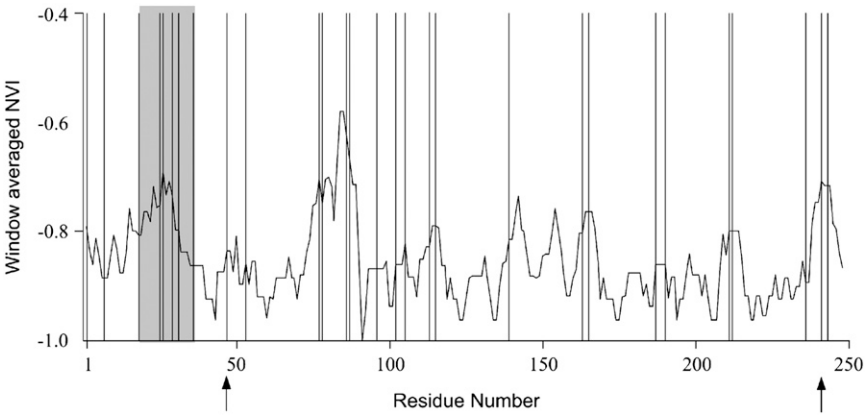

Figure 2.—Window-averaged plot of normed variability index along the *SLFL1* gene. The shadowed region indicates a variable region. Vertical lines indicate the location of the 10% most variable amino acid sites. Amino acid sites 47 and 241 showing four and five different amino acids, respectively, are indicated with arrows.

locus (Table 2), assumed 33 specificities in the *P. spinosa* Bragança population (Vieira *et al.* 2008a) and an *f*-value (the multiple by which variability at the selected locus is increased compared with the population average variability value) of 9.13 (obtained considering that the average synonymous variability level at the Prunus *S-RNase* gene is 0.241; Vieira *et al.* 2007). Ideally, the average of synonymous variability levels observed at several reference loci should be used but such data are not available for any Prunus species. Standard coalescent simulations suggest that when a per-site synonymous variability value of 0.026 is observed at *SLFL4* (Table 2), the true value could be as low as 0.018. When this value is used rather than the 0.026 value, an estimate of 6.15 is obtained for the population recombination rate between *SLFL1* and the *S-RNase* (the *f*-value is now 13.4). The physical distance between the *SLFL1* and the *S-RNase* gene varies (Entani *et al.* 2003; Ushijima *et al.* 2003). Assuming an average distance of 20 kb between the two genes, between 1.65×10^{-5} to 4.57×10^{-4} recombination events between adjacent nucleotides are expected per generation, depending on the method used.

The approach used here assumes that each specificity is mutually exchangeable and that there is no dominance among specificities and similar frequencies for all specificities. There is no theoretical expectation regarding isoplethy in polyploids. Recently, it has been shown (Vieira *et al.* 2008a) that in the *P. spinosa* population studied here, specificity frequencies may be unequal. The deviation is due to an apparent excess of both high- and low-frequency specificities. This type of deviation is similar to that observed in wild cherry populations, where a significant departure from the isoplethic distribution is also observed when standard tests are used (Stoeckel *et al.* 2008). These authors have, however, shown that the observed allele-frequency distribution is compatible with genetic drift and a model of subdivided populations and moderate migration between demes. Furthermore, for polyploid species, such as *P. spinosa*, it is conceivable that all chromosome pairings are not equally likely during meiosis. Nevertheless, in the polyploid *Prunus cerasus*, this is not the case (Hauck *et al.* 2006). Moreover, in Prunus, heteroallelic pollen retains its SI phenotype (Hauck *et al.* 2006). Therefore, it may be appropriate to use the formula of Kamau and Charlesworth (2005).

DISCUSSION

The *SLFL4* and *SLFL5* genes identified here are more closely related to *SLFL1* than to *SLFL2* and *SLFL3*. Both *SLFL2* and *SLFL3* are found in the vicinity of the *S*-locus (Entani *et al.* 2003; Ushijima *et al.* 2003), but the genomic location of *SLFL4* and *SLFL5* genes is unknown. Thus, it is not possible to determine whether the different genes originated through a series of regional duplications. Nevertheless, in *P. spinosa*, *SLFL1* pseudogenes are commonly found between *SLFL1* and the *S-RNase* gene. *SLFL1* pseudogenes also have been found in *P. mume* (Entani *et al.* 2003). These pseudogene sequences do not form a monophyletic group (Figure 1). Thus, it is inferred that they have multiple origins. It is thus conceivable that *SLFL* genes were frequently duplicated during evolution.

In *P. spinosa*, the *SLFL1* synonymous variability level is about four times higher than that found for the reference locus *SLFL4*. This suggests that variability patterns at *SLFL1* are influenced by the neighboring *S*-locus. Nevertheless, the *SLFL1* synonymous variability level is 2.3 times lower than those found for the *S-RNase* and *SFB* genes (Vieira *et al.* 2007, 2008a). Therefore, it is unlikely that the evolutionary histories of *SLFL1* and the *S*-locus are completely correlated. Indeed, a nonsignificant correlation is obtained between synonymous divergence values at the *SLFL1* and *S-RNase* genes.

Fewer than 10 recombinants per generation are expected for the *SLFL1*–*S*-locus intergenic region. In contrast, the *SLFL1* gene is expected to experience ~81 recombination events per generation, although, as noted (see results), this number may be an overestimate. Overall, the analyses performed here suggest that recombination levels increase near the *SLFL1* coding region (see results). Under this scenario, the observed associations between *SLFL1* alleles and *S*-locus specificities are expected, because to create an association between a given *SLFL1* sequence and two different specificities, one of the recombination breakpoints must be located in the *SLFL1*–*S*-locus intergenic region and this is a rare event. Recombination events affecting the *SLFL1* gene will uncouple the evolutionary histories of the two genes, as it is observed (see results).

Recombination seems to be severely repressed at the *S*-locus only. This region varies in size from 2.6 to ~50 kb (Nunes *et al.* 2006; Tao *et al.* 2007). Evidence suggestive of rare recombination has been reported at the *S-RNase* (Ortega *et al.* 2006; Vieira *et al.* 2007) and *SFB* (Nunes *et al.* 2006; Vieira *et al.* 2008a) genes. Given the evidence for severely restricted recombination at the *S*-locus only, the accumulation of weak deleterious mutations in the *S*-locus region is unlikely. Therefore, in Prunus there should be no selection against closely related allele pairs, in contrast with what has been predicted by Uyenoyama (1997). Such an effect may be restricted to Solanaceae species showing gametophytic self-incompatibility. In these species, the *S*-locus has been shown to have a centromeric location (see review by Wang *et al.* 2003). Thus it is conceivable that in Solanaceae species recombination is severely repressed in a large region around the *S*-locus.

We thank the anonymous reviewers for the constructive criticisms of earlier versions of the manuscript. This work has been funded by Fundação para a Ciência e Tecnologia [research project Programa Operacional Ciência e Inovação (POCI)/BIA-BDE/59887/2004 funded by POCI 2010, cofunded by Fundo Europeu de Desenvolvimento Regional funds].

LITERATURE CITED

Baiashvili, E. I., 1980 Karyological study of *Prunus spinosa* L. Bull. Georgian Acad. Sci. **100:** 645–647.

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

Cunningham, C. W., 1997 Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. **14:** 733–740.

de Nettancourt, D., 1997 *Incompatibility in Angiosperms*. Springer-Verlag, Berlin.

Entani, T., M. Iwano, H. Shiba, F. S. Che, A. Isogai *et al.*, 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.

Halliday, G., and M. Beadle, 1983 *Flora Europaea*. Cambridge University Press, Cambridge, UK.

Hauck, N. R., H. Yamane, R. Tao and A. F. Iezzoni, 2006 Accumulation of nonfunctional *S*-haplotypes results in the breakdown of gametophytic self-incompatibility in tetraploid Prunus. Genetics **172:** 1191–1198.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

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

Ingram, G. C., S. Doyle, R. Carpenter, E. A. Schultz, R. Simon *et al.*, 1997 Dual role for *fimbriata* in regulating floral homeotic genes and cell division in *Antirrhinum*. EMBO J. **16:** 6521–6534.

Innan, H., and M. Nordborg, 2003 The extent of linkage disequilibrium and haplotype sharing around a polymorphic site. Genetics **165:** 437–444.

Kelly, J. K., 1997 A test of neutrality based on interlocus associations. Genetics **146:** 1197–1206.

Kheyr-Pour, A., S. C. Bintrim, T. R. Ioerger, R. Remy, S. A. Hammond *et al.*, 1990 Sequence diversity of pistil *S*-proteins associated with gametophytic self-incompatibility in *Nicotiana alata*. Sex. Plant Reprod. **3:** 88–97.

Kamau, E., and D. Charlesworth, 2005 Balancing selection and low recombination affects diversity near the self-incompatibility loci of the plant *Arabidopsis lyrata*. Curr. Biol. **15:** 1773–1778.

Kumar, S., K. Tamura and M. Nei, 2004 MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinformatics **5:** 150–163.

Martin, D. P., C. Williamson and D. Posada, 2005 RDP2: recombination detection and analysis from sequence alignments. Bioinformatics **21:** 260–262.

Matsumoto, D., H. Yamane and R. Tao, 2008 Characterization of *SLFL1*, a pollen-expressed F-box gene located in the Prunus S locus. Sex. Plant Reprod. **21:** 113–121.

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

Nordborg, M., B. Charlesworth and D. Charlesworth, 1996 Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. Proc. R. Soc. Lond. B Biol. Sci. **163:** 1033–1039.

Nunes, M. D. S., R. A. M. Santos, S. M. Ferreira, J. Vieira and C. P. Vieira, 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.

Ortega, E., R. I. Boskovic, D. J. Sargent and K. T. Tobutt, 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., and K. A. Crandall, 1998 MODELTEST: testing the model of DNA substitution. Bioinformatics **14:** 817–818.

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

Salesses, G., 1973 Études cytologiques chez les *Prunus* II. Hybrides interspécifiques impliquant *P. cerasifera*, *P. spinosa*, *P. domestica* et *P. insititia*. Ann. Améloriement Plantes **23:** 145–161.

Schierup, M. H., X. Vekemans and F. B. Christiansen, 1998 Allelic genealogies in sporophytic self-incompatibility systems in plants. Genetics **150:** 1187–1198.

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

Stoeckel, S., V. Castric, S. Mariette and V. Vekemans, 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, D. L., 2002 *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods)*, Version 4.0b10. Sinauer, Sunderland, MA.

Tao, R., A. Watari, T. Hanada, T. Habu, H. Yaegaki *et al.*, 2007 Self-compatible peach (*Prunus persica*) has mutant versions of the *S* haplotypes found in self-incompatible *Prunus* species. Plant Mol. Biol. **63:** 109–123.

Thompson, J., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins, 1997 The ClustalX window interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:** 4876–4882.

Ushijima, K., H. Sassa, M. Tamura, M. Kusaba, R. Tao *et al.*, 2001 Characterization of the *S*-locus region of almond (*Prunus dulcis*): analysis of a somaclonal mutant and a cosmid contig for an *S* haplotype. Genetics **158:** 379–386.

Ushijima, K., H. Sassa, A. M. Dandekar, T. M. Gradziel, R. Tao *et al.*, 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, M. K., 1997 Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. Genetics **147:** 1389–1400.

Uyenoyama, M. K., 2000 Mutational origin of new mating type specificities in flowering plants. Genes Genet. Syst. **75:** 305–311.

Uyenoyama, M. K., 2005 Evolution under tight linkage to mating type. New Phytol. **165:** 63–70.

Uyenoyama, M. K., Y. Zhang and E. Newbigin, 2001 On the origin of self-incompatibility haplotypes: transition through self-compatible intermediates. Genetics **157:** 1805–1817.

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

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

Vieira, J., R. Morales-Hojas, R. A. M. Santos and C. P. Vieira, 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., R. A. M. Santos, S. M. Ferreira and C. P. Vieira, 2008a Molecular evolution at the *Prunus spinosa SFB*: allele diversity, population structure and amino acid sites under positive selection. Heredity (in press).

Vieira, J., N. A. Fonseca, R. A. M. Santos, T. Habu, R. Tao *et al.*, 2008b The number, age, sharing and relatedness of S-locus specificities in *Prunus*. Genet. Res. **90:** 17–26.

Wang, Y., X. Wang, A. L. Skirpan and T. H. Kao, 2003 *S-RNase*-mediated self-incompatibility. J. Exp. Bot. **54:** 115–122.

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

Wiuf, C., K. Zhao, H. Innan and M. Nordborg, 2004 The probability and chromosomal extent of trans-specific polymorphism. Genetics **168:** 2363–2372.

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

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

Communicating editor: M. K. Uyenoyama