Plant Molecular Biology **53:** 565–580, 2003.
© 2003 *Kluwer Academic Publishers. Printed in the Netherlands.*

Genetic mapping and molecular characterization of the self-incompatibility (*S*) locus in *Petunia inflata*

Yan Wang[1,+], Xi Wang[2,3,+], Andrew G. McCubbin[2,4] and Teh-hui Kao[1,2,*]

[1]*Intercollege Graduate Degree Program in Plant Physiology and* [2]*Department of Biochemistry and Molecular Biology, 403 Althouse Laboratory, Pennsylvania State University, University Park, PA 16802, USA (*author for correspondence; e-mail txk3@psu.edu); present addresses:* [3]*Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA;* [4]*School of Biological Sciences, Washington State University, Pullman, WA 99164, USA;* [+]*these authors contributed equally to the work*

Received 25 August 2003; accepted in revised form 23 October 2003

Key words: genetic mapping, *Petunia*, self-incompatibility, *S*-locus-linked genes

Abstract

Gametophytic self-incompatibility (SI) possessed by the Solanaceae is controlled by a highly polymorphic locus called the *S* locus. The *S* locus contains two linked genes, *S-RNase*, which determines female specificity, and the as yet unidentified pollen *S* gene, which determines male specificity in SI interactions. To identify the pollen *S* gene of *Petunia inflata*, we had previously used mRNA differential display and subtractive hybridization to identify 13 pollen-expressed genes that showed *S*-haplotype-specific RFLP. Here, we carried out recombination analysis of 1205 F_2 plants to determine the genetic distance between each of these *S*-linked genes and *S-RNase*. Recombination was observed between four of the genes (*3.16*, *G211*, *G212*, and *G221*) and *S-RNase*, whereas no recombination was observed for the other nine genes (*3.2*, *3.15*, *A113*, *A134*, *A181*, *A301*, *G261*, *X9*, and *X11*). A genetic map of the *S* locus was constructed, with *3.16* and *G221* delimiting the outer limits. None of the observed crossovers disrupted SI, suggesting that all the genes required for SI are contained in the chromosomal region defined by *3.16* and *G221*. These results and our preliminary chromosome walking results suggest that the *S* locus is a huge multi-gene complex. Allelic sequence diversity of *G221* and *3.16*, as well as of *3.2*, *3.15*, *A113*, *A134* and *G261*, was determined by comparing two or three alleles of their cDNA and/or genomic sequences. In contrast to *S-RNase*, all these genes showed very low degrees of allelic sequence diversity in the coding regions, introns, and flanking regions.

Abbreviations: GSI, gametophytic self-incompatibility; PFGE, pulsed-field gel electrophoresis; RFLP, restriction fragment length polymorphism; RT-PCR, reverse transcriptase PCR; SI, self-incompatibility

Introduction

Self-incompatibility (SI) is a reproductive strategy adopted by many flowering plants to prevent inbreeding and promote out-crosses. For most of the families that have been studied at the molecular level so far, a single polymorphic locus, named the *S* locus, determines the specificity of SI interactions between pollen and the pistil (de Nettancourt, 2001). The Solanaceae possess a major type of SI, termed gametophytic SI (GSI), so named because the SI behavior of pollen is determined by its own *S* genotype, rather than by the *S* genotype of the pollen-producing parent, as is the case for sporophytic SI. For GSI, self-pollen, which carries an *S* haplotype identical to one of the two *S* haplotypes carried by a pistil, is rejected during its tube growth in the pistil.

The solanaceous *S* locus contains the *S-RNase* gene, which determines the *S*-haplotype specificity of the pistil (Lee *et al.*, 1994; Murfett *et al.*, 1994), and the as yet unidentified pollen *S* gene, which determines the *S*-haplotype specificity of the pollen.

We previously used mRNA differential display and high-stringency subtractive hybridization to identify pollen-expressed genes of *Petunia inflata* that showed sequence differences between different *S* haplotypes (Dowd *et al.*, 2000; McCubbin *et al.*, 2000a). The rationale for using these two approaches is that the pollen *S* gene is expected to be polymorphic and show a significant degree of allelic sequence diversity, as is the case for *S-RNase*. All the primary candidates identified were further analyzed by genomic DNA blotting to determine whether they showed *S*-haplotype-specific restriction fragment length polymorphism (RFLP). For the 21 pollen-expressed genes that did, recombination analysis was carried out in 100 F_2 plants segregating for different *S* haplotypes to examine whether the RFLP co-segregated with the *S* haplotypes. Twelve genes were found to be linked to the *S* locus based on the lack of recombination observed between each of them and *S-RNase*.

The *S*-linked pollen-expressed genes are potential candidates for the pollen *S* gene. However, because recombination is thought to be suppressed at the *S* locus (de Nettancourt, 2001), recombination analysis using larger numbers of plants would be needed to assess how tightly it is that these genes are linked to the *S* locus. The results would eliminate any incompletely linked genes and narrow the number of potential candidates for the pollen *S* gene. Moreover, the degree of allelic sequence diversity of each of these genes would have to be determined.

Here we report the results of extensive recombination analysis of a total of 1205 F_2 plants segregating for S_1 and S_2 haplotypes. Various degrees of recombination were observed between three of the 12 marker genes and *S-RNase*, but no recombination was observed for the other nine markers. Another gene (*G221*) that had previously been found not to be completely linked to the *S* locus (McCubbin *et al.*, 2000a) was also used in the recombination analysis. We constructed a genetic map of the *S* locus based on the results of the recombination analysis, and determined the allelic sequence diversity of five of the nine *S*-linked genes that were mapped within the *S* locus and two of the four *S*-linked genes that were mapped outside the *S* locus.

Materials and methods

Plant material

Petunia inflata plants of S_1S_1 and S_2S_2 genotypes (Ai *et al.*, 1990) were used in generating populations segregating for S_1 and S_2 haplotypes used in this study. To increase the rate of recombination in the sub-centromeric region, some of the F_1 plants obtained from crosses between S_1S_1 and S_2S_2 plants were subjected to chemical or physical treatment. For the chemical treatment, plants were grown in 2.5 × 2.5 cm square pots and watered with 20 μM *n*-butyric acid. The flower-bearing stems were immersed in the solution as well. For the physical treatment, plants were either exposed to 350 nm UV light for 10 s or incubated at 42 °C for 2 h (heat shock). Pollen from these F_1 plants was used to pollinate stigmas of untreated S_1S_2 plants at immature flower stages (bud-selfing) to overcome SI. The untreated F_1 plants were also bud self-pollinated. A total of 1105 F_2 plants were generated.

Isolation of cDNA clones and cDNA fragments

S_1, S_2 and S_3 pollen cDNA libraries previously constructed from poly(A)$^+$ RNA of S_1, S_2 and S_3 pollen (Skirpan *et al.*, 2001), respectively, were used for screening. The probes were partial cDNAs for the 13 *S*-linked genes, which had been previously isolated from mRNA differential display and subtractive hybridization (Dowd *et al.*, 2000; McCubbin *et al.*, 2000a). Library screening followed the procedure described by Skirpan *et al.* (2001). Reverse transcriptase PCR (RT-PCR) was used to isolate cDNA fragments for the S_1, S_2 and S_3 alleles of *A113*, for the S_1 and S_2 alleles of *3.15*, and for the S_2 allele of *G221*. Total RNA was separately isolated from S_1, S_2 and S_3 pollen with TRIzol reagent (Life Technologies) following the manufacturer's manual. First-strand cDNA synthesis and PCR reactions were performed according to the procedures of Wang *et al.* (2001) except for the following modifications. For the S_1 and S_2 alleles of *3.15* and the S_1, S_2 and S_3 alleles of *A113*, the reaction mixture was denatured at 94 °C for 3 min and subjected to 30 cycles of PCR. Each cycle consisted of denaturation at 94 °C for 30 s, annealing at 54 °C for 30 s and extension at 72 °C for 2 min. After the final cycle, the extension was allowed to continue for an additional 10 min at 72 °C. For the S_2 allele of *G221*, the cycle conditions were the same as those described above except that the extension was carried out

at 72 °C for 30 s. The primers for the S_1 and S_2 alleles of *3.15* were: 5′-TTGTCATGCACGCCAATTATG-3′ and 5′-TTGTCTAGGTGGTCCTGCG-3′; the primers for the S_1, S_2 and S_3 alleles of *A113* were: 5′-TTGGATGGGACCATGAAGGA-3′ and 5′-GAATAT GGCTTCCATTGATC-3′; and the primers for the S_2 allele of *G221* were: 5′-TTTGATCCAAATGGAAGA CAA-3′ and 5′-GGAAAACATAACTGCCCCTC-3′.

Recombination analysis

For recombination analysis using the genomic DNA blotting approach, genomic DNA was extracted from young leaves of F_2 plants segregating for S_1 and S_2 haplotypes with Plant DNAzol Reagent (Life Technologies) according to the manufacturer's instructions. Genomic DNA (15 μg) was digested with 40 units of *Eco*RI (for hybridization with *3.2*, *3.16*, *A181*, *A301*, *G211*, *G221*, *X9* and *S*-RNase cDNA probes) or with 10 units of *Xba*I (for hybridization with *3.15*, *A113*, *A134*, *G212*, *G261*, *X11* and *S-RNase* cDNA probes) at 37 °C overnight. The digested DNA samples were fractionated on 0.6% agarose gels, and transferred onto positively charged Biodyne B nylon membranes (Life Technologies). The DNA blots were baked at 80 °C for 2 h, and hybridized with a radiolabeled probe in 1 M NaCl, 1% SDS, 10% dextran sulfate at 65 °C overnight. The blots were washed twice with 1× SSC, 0.1% SDS at 65 °C for 30 min each before being exposed to X-ray film. The *S-RNase* and *S*-linked cDNAs were labeled with α-^{32}P-dCTP with the RTS RadPrime DNA Labeling System (Life Technologies). The blots were stripped of a probe by rinsing them in boiling 0.1% SDS solution three times before being used for hybridization with another probe.

For recombination analysis using a PCR-based approach, genomic DNA was prepared from one small young leaf of each F_2 plant as described above. The DNA was dissolved in 30 μl H_2O, and 2 μl was added to an 8 μl reaction buffer (20 mM Tris-HCl pH 8.4, 2 mM $MgCl_2$, 50 mM KCl) containing 0.25 μM of each primer, 0.2 mM dNTPs, and 1 unit of *Taq* DNA polymerase (Life Technologies). The reaction mixture was denatured at 93 °C for 2 min and then cycled for 35 times as follows: denaturation at 93 °C for 30 s, annealing (at 54 °C, 52 °C and 50 °C for *S-RNase*, *G211*, and *G221*, respectively) for 45 s, and extension at 72 °C for 30 s. After the final cycle, the sample was kept at 72 °C for an additional 5 min. The primers used were: 5′-ACGCACTTGAGGTCAGGATCT-3′ and 5′-CGTCTAGAGTTCTGCACTGGC-3′ for *S-RNase*; 5′-CATTGAGACTACTTTCAACAATTCCA-3′ and 5′-GACGCAGCACAGTTAGCCAGG-3′ for *G211*; and 5′-TAACTGCCCCTCACTAACAAACT-3′ and 5′-TGGAAGACAATGCTTGACGATGGA-3′ for *G221*. The PCR products were then subjected to complete restriction enzyme digestion, and the digests were fractionated on 1.5% agarose gels to reveal genotype-specific digestion patterns. The recombinant plants were further confirmed by Southern blotting as mentioned above. The genetic distance (in cM) between each marker and *S-RNase* was determined using the following formula: number of recombinant plants/number of total plants analyzed × 1/2 × 100.

Construction and screening of a P. inflata *BAC library of the* S_1S_1 *genotype*

Isolation of megabase (Mb) genomic DNA from leaves of S_1S_1 plants, partial *Bam*HI digestion of genomic DNA, purification of BIBAC2 vector (Hamilton, 1997), and ligation of digested DNA to *Bam*HI digested BIBAC2 were all carried out as described by McCubbin *et al.* (2000b). To estimate the quality of the BAC library, BAC DNA was isolated from 120 randomly chosen clones, digested with *Not*I, and subjected to pulsed-field gel electrophoresis (PFGE), all following the procedures described by McCubbin *et al.* (2000b). The library was then screened using as probes cDNAs for S_1*-RNase* and the 13 *S*-linked genes as described by McCubbin *et al.* (2000b).

To confirm the authenticity of each BAC clone isolated, BAC DNA was prepared, digested with *Eco*RI or *Bam*HI, and separated by PFGE on 1% Seakem Gold agarose gel (BioWhittaker), along with S_1S_1 and S_2S_2 genomic DNA that had been digested with the same restriction enzyme. The conditions for PGFE were optimized for separation of 1–50 kb DNA fragments (McCubbin *et al.*, 2000b). The DNA digests were transferred to a nylon membrane, and the DNA blot was hybridized with cDNA probes for *S-RNase* and the *S*-linked gene contained in the BAC clone. Each BAC clone was confirmed by comparing its hybridization pattern with that of the genomic DNA.

Sequence analysis of cDNA and BAC clones

BAC DNA was isolated for sequencing with the NucleoBond Plasmid Purification Kit (Clontech) following the manufacturer's manual. cDNA clones and BAC clones were sequenced at the Nucleic Acid Facility of Pennsylvania State University. Cycle sequencing reactions were performed with 3′ BigDye-labeled

dideoxynucleotide triphosphates (v 3.0 dye terminators), and samples were run on an ABI Hitachi 3100 Genetic Analyzer (Applied Biosystems). Data were analyzed with the ABI 3100 Data Collection Program (v 1.0.1). Allelic cDNA sequences and their deduced amino acid sequences were compared by means of global ALIGN (Pearson *et al.*, 1997) with gap penalties of $-12/-2$. Sequences of large ($\geq$24 nucleotides) insertions/deletions were eliminated, and thus only small ($\leq$15 nucleotides) insertions/deletions were taken into account when performing comparisons of allelic genomic sequences with global ALIGN. Multiple sequences were aligned with CLUSTALW (www.ebi.ac.uk/clustalw). The proportion of synonymous differences per synonymous site (p_s) and the proportion of non-synonymous differences per non-synonymous site (p_n) were determined with *MEGA* version 2.1 (Kumar *et al.*, 2001) according to the method of Nei and Gojobori (1986).

Results

Genetic mapping of the P. inflata S *locus*

We used genomic DNA blotting and a PCR-based method to assess the genetic linkage between *S-RNase* and each of the 12 pollen-expressed genes (*3.2*, *3.15*, *3.16*, *A113*, *A134*, *A301*, *A181*, *G211*, *G212*, *G261*, *X9* and *X11*) that had previously been found to be linked to the *S* locus (Dowd *et al.*, 2000; McCubbin *et al.*, 2000a). We also included another *S*-linked marker, *G221*, for which we had previously found one recombinant from the analysis of 100 F_2 plants (McCubbin *et al.*, 2000a). A total of 1105 F_2 plants segregating for S_1 and S_2 haplotypes were generated from F_1 plants obtained from crosses between an S_1S_1 plant and an S_2S_2 plant. The *S* locus of another species of *Petunia*, *P. hybrida*, was found, by fluorescence *in situ* hybridization, to be located near the centromere (Entani *et al.*, 1999), where the recombination rate is greatly reduced. Thus, we subjected some of the F_1 plants to chemical and physical treatments (see Materials and methods) that have been shown to increase recombination frequencies near the centromere in *Arabidopsis* (Copenhaver *et al.*, 1999).

We first used genomic DNA blotting to analyze 300 of the F_2 plants. Genomic DNA isolated from these plants was digested with ether *Eco*RI or *Xba*I and the digests were separately hybridized with cDNAs for each of the 13 *S*-linked genes and for the S_1- and S_2-*RNase* genes. The results for nine of these plants, as

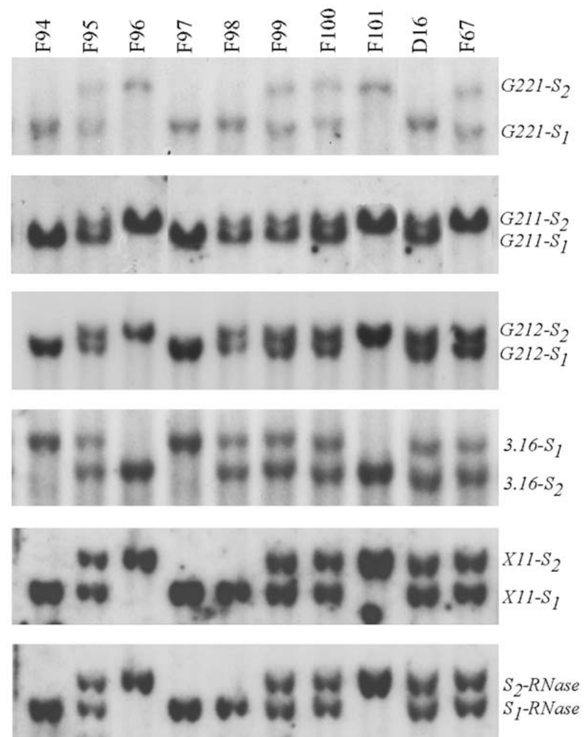

Figure 1. Genomic DNA blot analysis of recombination between *S-RNase* and five of the *S*-linked genes. The blot was first hybridized with a probe containing cDNAs for S_1-*RNase* and S_2-*RNase*. After hybridization and autoradiography, the blot was stripped of the hybridizing probe, and hybridized sequentially with cDNAs for each of the five *S*-linked genes, *G221*, *G211*, *G212*, *3.16* and *X11*. F67, and F94 to F101 were 9 of the 300 F_2 plants analyzed by genomic DNA blotting; D16 was the recombinant plant previously identified by genomic blot analysis of 100 F_2 plants. The identity of each hybridization band is indicated to the right of the figure, with the name of the gene followed by a hyphen and the allele number.

well as for the previously identified recombinant plant, D16, are shown in Figure 1. The *S* genotype of each plant was determined by the hybridization pattern of the two *S-RNase* genes. For example, the *S* genotype of F94 was determined to be S_1S_1, because its genomic DNA only contained a fragment specific to the S_1-*RNase* gene. For this plant, the hybridization patterns of all the five *S*-linked genes shown in Figure 1, as well as of the other eight *S*-linked genes (results not shown), were consistent with those expected of the S_1S_1 genotype. That is, only the fragment corresponding to the S_1 allele of each gene was observed. Thus, no recombination occurred between any of the 13 *S*-linked genes and *S-RNase* in F94.

From the 300 F_2 plants analyzed, we identified two recombinant plants, designated F67 and F98. Both F67 (S_1S_2) and F98 (S_1S_1) showed recombination between

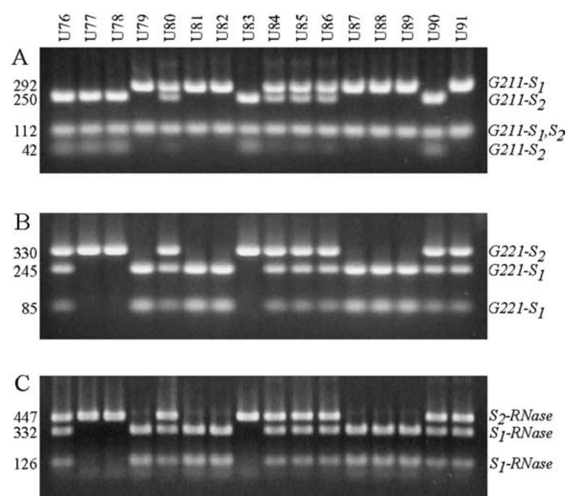

Figure 2. PCR-based analysis of recombination between *S-RNase* and two *S*-linked genes, *G211* and *G221*. A. PCR products of *G211* digested with *Hae*III. The product of the S_1 allele (*G211-*S_1) was cut into two fragments (112 bp and 292 bp), whereas the product of the S_2 allele (*G211-*S_2) was cut into three fragments (42 bp, 112 bp and 250 bp). B. PCR products of *G221* digested with *Pvu*II. The product of the S_1 allele of *G221* (*G221-*S_1) was cut into two fragments (85 bp and 245 bp), whereas the product of the S_2 allele (*G221-*S_2) was not cut. C. PCR products of *S-RNase* digested with *Hin*dIII for genotyping *S* haplotypes. The product of the S_1*-RNase* gene (458 bp) was cut into two fragments (126 bp and 332 bp), whereas the product of the S_2*-RNase* gene (447 bp) was not cut. U76 to U91 were 16 of the 805 F_2 plants analyzed by the PCR-based method. The sizes (in bp) of the DNA fragments are indicated to the left of the figure, and the identity of each fragment is indicated to the right.

S-RNase and *G211*, because only the S_2-specific fragment of *G211* was observed in F67 and because both S_1- and S_2-specific fragments of *G211* were observed in F98 (Figure 1). F67 did not show recombination between any of the other 12 *S*-linked genes and *S-RNase* (Figure 1 and results not shown), whereas F98 also showed recombination between *S-RNase* and *G212*, and between *S-RNase* and *3.16* (Figure 1). The previously identified D16 showed recombination between *S-RNase* and *G221*, but not between *S-RNase* and any of the other 12 *S*-linked genes examined (Figure 1; results not shown). The results, taken together, suggest that *G221* is on one side of *S-RNase*, whereas *3.16*, *G211* and *G212* are on the opposite side of *S-RNase*, with *G211* being farthest from *S-RNase*.

Having identified *G221* and *G211* as two *S*-linked genes that are located on opposite sides of *S-RNase*, we switched to the use of a PCR-based method and developed cleaved amplifiable polymorphism (CAP) markers to facilitate recombination analysis of the remaining 805 F_2 plants. For *G221*, a partial sequence of the S_1 allele was determined from the cDNA fragment obtained by mRNA differential display (McCubbin *et al.*, 2000a). A cDNA fragment for the S_2 allele of *G221* was then isolated by RT-PCR with two primers designed on the basis of the sequence of the cDNA fragment for the S_1 allele (See Materials and methods). Comparison of the sequences of these two cDNA fragments of *G221* revealed a region that contained a *Pvu*II site in the S_1 allele but none in the S_2 allele. For *G211*, cDNA clones for S_1 and S_2 alleles were isolated from pollen cDNA libraries of S_1S_1 and S_2S_2 genotypes, respectively (see next section), by using as a probe the corresponding partial cDNA obtained from mRNA differential display (McCubbin *et al.*, 2000a). Comparison of the sequences of S_1- and S_2-cDNA clones of *G211* (GenBank accession numbers AY283239 and AY283240) revealed a region that contained one *Hae*III site in the S_1 allele but two *Hae*III sites in the S_2 allele. Comparison of the genomic sequences of the S_1 and S_2 alleles of *S-RNase* (see the last section of Results) revealed a region that contained a *Hin*dIII site in the S_1 allele but none in the S_2 allele.

For each of *G221*, *G211* and *S-RNase*, PCR primers were designed on the basis of two sequences conserved between the S_1 and S_2 alleles, such that the PCR products could be distinguished by the restriction digestion patterns described above (see Materials and methods for the primer sequences). For each of the 805 F_2 plants, three separate PCRs were then carried out on genomic DNA using the primer pairs for *G211*, *G221* and *S-RNase*, respectively. The amplified fragment from each reaction was then digested with the appropriate restriction enzyme.

Of the 805 F_2 plants analyzed by the PCR-based method, six (N43, N124, P85, U76, U90, and U91) were found to show recombination between *S-RNase* and *G211*, and one (P85) of the six plants also showed recombination between *S-RNase* and *G221*. Figure 2 shows the results for representative plants. For example, the *S* genotype of U76 was determined to be S_1S_2, because a 447 bp PCR fragment of the S_2-*RNase* gene and two fragments (126 bp and 332 bp) resulting from *Hin*dIII digestion of the PCR product of the S_1*-RNase* gene were obtained. When *G211* was analyzed, only three fragments, 42, 112 and 250 bp, resulting from *Hae*III digestion of the PCR product of the S_2 allele were obtained, suggesting that a crossover occurred between *G211* and *S-RNase* in U76. For *G221*, the fragments expected of both S_1 and S_2 alleles were obtained, thus no recombination occurred between *S-RNase* and *G221* in U76.

Table 1. Summary of recombination analysis of 1205 F_2 plants.

S-linked gene	Number of recombinant plants	Plant number	Recombination frequency (%)
3.2	0		0
3.15	0		0
3.16	4	F98, N43, N124, P85	0.17
A113	0		0
A134	0		0
A181	0		0
A301	0		0
G211	8	F67, F98, N43, N124, P85, U76, U90, U91	0.33
G212	7	F98, N43, N124, P85, U76, U90, U91	0.29
G221	2	D16, P85	0.08
G261	0		0
X9	0		0
X11	0		0

Table 2. Allelic sequence identity of *S-RNase* and seven *S*-linked genes

Gene	Alleles sequenced[a]	Allelic sequence identity in coding region		p_s[b]	p_n[b]	Putative homologue[c] (accession number)
		nucleotide sequence (number of base pairs compared)	amino acid sequence (number of amino acids compared)			
3.2	S_1, S_2, S_3	98.0–98.6% (648)	96.8–97.7% (216)	0.0285 ± 0.0120	0.0093 ± 0.0034	rose petal EST (BQ106518)
3.15	S_1[d], S_2[d], S_3	98.5–99.0% (1830)	97.4–98.5% (610)	0.0265 ± 0.0058	0.0081 ± 0.0019	*Arabidopsis* protein (BAC43446)
3.16	S_1, S_2	98.4% (984)	99.7% (328)	0.0633 ± 0.0158	0.0013 ± 0.0013	tomato fructokinase 2 (AAB51108)
A113	S_1, S_2, S_3	96.5–97.1% (1128)	94.9–96.8% (376)	0.0833 ± 0.0146	0.0182 ± 0.0039	*Antirrhinum* SLF-S_2 (CAC33202)
A134	S_1, S_2, S_3	98.7–99.3% (1137)	96.6–98.2% (379)	0.0054 ± 0.0035	0.0105 ± 0.0026	*Antirrhinum* SLF-S_2 (CAC33202)
G221	S_1[d], S_2[d]	98.3% (300)	98.0% (100)	0.0339 ± 0.0211	0.0111 ± 0.0070	*Arabidopsis* protein (NP_563817)
G261	S_1, S_2	99.6% (255)	100% (85)	0.0180 ± 0.0179	0	rice NOI protein (T02212)
S-RNase	S_1, S_2, S_3	80.3–85.4% (666)	70.9–79.4% (222)	0.3086 ± 0.0317	0.1417 ± 0.0145	

[a]GenBank accession numbers: AF530586 (*3.2-*S_1), AF530587 (*3.2-*S_2), AF530588 (*3.2-*S_3), AF530590 (*3.15-*S_1), AF530591 (*3.15-*S_2), AF530589 (*3.15-*S_3), AF530592 (*3.16-*S_1), AF530593 (*3.16-*S_2), AF530594 (*G261-*S_1), AF530595 (*G261-*S_2), AF530596 (*G221-*S_1), AF530597 (*G221-*S_2), AY363970 (*A113-*S_1), AY363971 (*A113-*S_2), AY363972 (*A113-*S_3), AY363973 (*A134-*S_1), AY363974 (*A134-*S_2), AY363975 (*A134-*S_3), M67990 (S_1*-RNase*), AY136628 (S_2*-RNase*) and M67991 (S_3*-RNase*).
[b]Value ± SE; overall mean ± SE shown for *3.2*, *3.15*, *A113*, *A134* and *S-RNase*.
[c]Expect value $\geq e^{-10}$.
[d]Partial cDNA sequences.

Genomic DNA blot analysis was carried out on the six recombinant plants with cDNAs for the other 11 *S*-linked genes as probes to determine whether any of them also showed recombination with *S-RNase* (results not shown). All these six recombinant plants showed recombination between *S-RNase* and *G212*, and three of them (N43, N124, and P85) showed recombination between *S-RNase* and *3.16*.

Table 1 summarizes the results of the recombination analysis of the 1105 F_2 plants generated in this work and the 100 F_2 plants previously generated. In all, nine recombinant plants were found. Among the 12 *S*-linked genes for which recombination was not observed in our previous work (Dowd *et al.*, 2000; McCubbin *et al.*, 2000a), three (*3.16*, *G211* and *G212*) were found to show recombination with *S-RNase*. Recombination for *G221*, *3.16*, *G212*, and *G211* occurred in 2, 4, 7 and 8 plants, respectively. Five of the recombinant plants (D16, F67, F98, N43 and N124) were identified among 764 F_2 plants derived from untreated F_1 plants (0.65%). UV treatment appeared to increase recombination frequency, as four of the recombinant plants (P85, U76, U90 and U91) were identified from among 290 F_2 plants derived from UV-treated F_1 plants (1.38%). No recombinants were identified among 108 F_2 plants derived from *n*-butyric acid-treated F_1 plants, or among 43 F_2 plants derived from heat-shock-treated F_1 plants. However, the number of plants derived from these two latter treatments might be too small for us to assess their effectiveness.

No recombination was observed for any of the remaining nine *S*-linked genes (*3.2*, *3.15*, *A113*, *A134*, *A301*, *A181*, *G261*, *X9* and *X11*), suggesting that these genes are separated from each other and from *S-RNase* by less than 0.08 cM. However, the relative genetic distance of these nine genes with respect to *S-RNase* could not be determined. A genetic map of the *S* locus was constructed and is shown in Figure 3. Two markers, *3.16* and *G221*, which delimit the maximum physical size of the *S* locus, were mapped at 0.17 cM and 0.08 cM from the *S* locus, respectively, but the exact position of the break points is as yet unknown. The 13 markers collectively span a genomic region of 0.41 cM.

In constructing the genetic map shown in Figure 3, we assumed that a double crossover occurred in plant P85, one between *S-RNase* and *G221*, and the other between *S-RNase* and *3.16*. This is consistent with the fact that P85 was derived from an UV-treated F_1 plant because, as described above, this treatment increased the recombination frequency by two-fold over that of

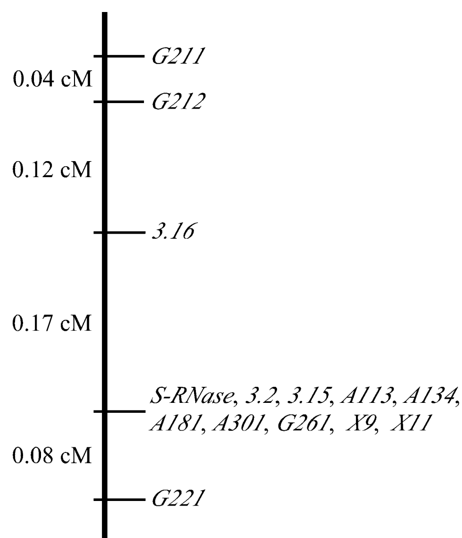

Figure 3. A genetic map of the *S* locus of *P. inflata*. The genetic distance between adjacent makers is in centimorgan (cM). The order of *S-RNase* and the markers *3.2*, *3.15*, *A113*, *A134*, *A181*, *A301*, *G261*, *X9* and *X11* is arbitrary.

untreated F_1 plants. An alternative map would be to place *G221* between *S-RNase* and 3.16, assuming that a double crossover occurred in D16 (derived from an untreated F_1 plant) rather than in P85. However, even though this scenario cannot be completely ruled out, it would seem unlikely considering the very low frequency of recombination observed among the plants derived from the untreated F_1 plants.

To examine whether the crossovers that occurred in the nine recombinant plants disrupted the SI behavior, these plants were self-pollinated and reciprocally crossed with tester plants of S_1S_2 genotype. All the nine plants were found to be self-incompatible and exhibited the SI behavior expected of their respective *S* genotypes (results not shown). For example, F67 (S_1S_2 genotype) was reciprocally incompatible with the S_1S_2 tester plants; N43 (S_2S_2 genotype) accepted pollen of the S_1S_2 tester plants, whereas the S_1S_2 tester plants rejected pollen of N43. Thus, none of the crossovers disrupted the linkage between *S-RNase* and the pollen *S* gene, suggesting that the pollen *S* gene must be located within the region delimited by *G221* and *3.16*.

We also examined two additional markers, *CP100* of *Solanum tuberosum* and *48A* of *Nicotiana alata*, for their linkage to the *S* locus of *P. inflata*. This is because *CP100* has previously been found to be linked to the *S* locus of *P. hybrida* (ten Hoopen *et al.*, 1998) and the pollen *S* gene has previously been mapped to the re-

gion between *48A* and *S-RNase* of *N. alata* (Golz *et al.*, 2001). We found that *CP100* strongly hybridized to the genomic DNA of both S_1S_1 and S_2S_2 genotypes of *P. inflata* under high stringency and that it exhibited the same RFLP pattern as *X11* (data not shown). This is consistent with our previous finding that both *X11* and *CP100* hybridized strongly to the same BAC clone (McCubbin *et al.*, 2000b). *48A*, however, did not hybridize to the genomic DNA of either S_1S_1 or S_2S_2 genotype even under low stringency of hybridization (results not shown).

Isolation and sequence analysis of cDNA clones for S*-linked genes*

To examine the degrees of allelic sequence diversity in the coding regions of the 13 *S*-linked genes, pollen cDNA libraries of S_1, S_2 and S_3 haplotypes were screened by using as probes the corresponding partial cDNAs isolated from mRNA differential display and subtractive hybridization. cDNA clones for all three alleles of *3.2* (*3.2-S_1*, *-S_2*, *-S_3*) and *A134* (*A134-S_1*, *-S_2*, *-S_3*), for S_1 and S$_2$ alleles of *3.16* (*3.16-S_1*, *-S_2*) and *G261* (*G261-S_1*, *-S_2*), and for the S_3 allele of *3.15* (*3.15-S_3*), were isolated (Table 2). The sequences of all these cDNA clones were full-length, or nearly full-length, because they contained a poly(A)$^+$ tail at their 3′ end and, except for the three alleles of *A134*, they contained at least one stop codon before the putative start codon. Moreover, the sizes of these cDNA clones were similar to those of their corresponding transcripts as determined by RNA gel blot analysis (results not shown). For example, cDNAs for *3.2-S_1*, *3.2-S_2* and *3.2-S_3* were 1355, 1341 and 1358 bp in size, respectively, and the size of the *3.2* transcript was determined to be ca. 1.4 kb.

For those alleles of the marker genes for which cDNA library screening failed to yield positive clones, RT-PCR was used to isolate cDNA fragments with total RNA of S_1, S_2 , and/or S_3 pollen. cDNA fragments for *3.15-S_1* and *3.15-S_2* were obtained with primers designed based on the sequence of *3.15-S_3*. The sequence of the cDNA fragment for the S_3 allele of *A113* obtained from differential display (McCubbin *et al.*, 2000a) is very similar to the sequences of the cDNA clones for S_1, S_2 and S_3 alleles of *A134*. Thus, cDNA fragments for these three alleles of *A113* were isolated with primers designed on the basis of a conserved sequence at the 5′ end of *A134-S_1*, *-S_2* and *-S_3* cDNAs, and the 3′ non-coding sequence of the *A113-S_3* cDNA fragment. For *G221*, as described earlier, the sequences of the S_1 and S_2 alleles were determined from the cDNA fragment previously isolated from mRNA differential display (McCubbin *et al.*, 2000a) and from the cDNA fragment obtained by RT-PCR, respectively. The sequence of *G221-S_1* (which was longer than that of *G221-S_2*) had a complete 3′ end with a poly(A)$^+$, and BLAST searches of the non-redundant databases revealed that its deduced amino acid sequence (101 residues) was most similar (63% identity) to that of the C-terminal half of an *Arabidopsis* protein (377 residues) of unknown function (GenBank accession number NP_563817), with the first eight amino acids identical to amino acids residues 277–284 of the latter.

The deduced amino acid sequence of *3.16* (328 residues) was most similar to that of fructokinase 2 of tomato (accession number AAB51108; Kanayama *et al.*, 1997; Martinez-Barajas *et al.*, 1997), also with 328 amino acid residues. The deduced amino acid sequence of *G261* (85 residues) was most similar to that of a putative anion transporter, NOI, of rice (80 amino acids in size; accession number T02212). The deduced amino acid sequences of *3.2* and *3.15* were most similar to those of a rose petal EST (accession number BQ106518) and a putative protein of *Arabidopsis* (accession number BAC43446), respectively. The deduced amino acid sequences of all three alleles of *A113* and *A134* were most similar (ca. 32% identical with E values ranging from $8e^{-37}$ to $2e^{-40}$) to that of the S_2 allele of the *S*-locus F-box gene (*AhSLF-S_2*) of *Antirrhinum hispanicum*. *AhSLF-S_2* was identified from sequence analysis of a 69 kb *S*-locus region that contains the *S_2-RNase* gene (Lai *et al.*, 2002). An alignment of the deduced amino acids of *A113-S_1*, *-S_2*, *-S_3*, *A134-S_1*, *-S_2*, *-S_3*, and *AhSLF-S_2* is shown in Figure 4. Pairwise comparisons of the three alleles of *A113* with the three alleles of *A134* showed that they shared 91.8–93.0% nucleotide sequence identity in the coding region and 86.8–89.2% amino acid sequence identity.

Allelic diversity of the nucleotide and deduced amino acid sequences was determined for *3.2*, *3.15*, *3.16*, *A113*, *A134*, *G221*, and *G261* (Table 2). These seven genes exhibited very low degrees of allelic sequence diversity in their coding regions, with nucleotide sequence identity ranging from 96.5% to 99.6% and amino acid sequence identity ranging from 94.9% to 100%. In contrast, the S_1, S_2 and S_3 alleles of *S-RNase* show 14.6–19.7% nucleotide sequence diversity in the coding region and 20.6–29.1% amino acid sequence diversity. We also calculated p_s (propor-

tion of synonymous differences per synonymous site) and p_n (proportion of non-synonymous differences per non-synonymous site) for each pairwise allelic sequence comparison of these seven genes and *S-RNase* (Table 2). For *3.2*, *3.15*, *A113*, *A134* and *S-RNase*, where three alleles were compared, the average values of p_s and p_n were calculated and are shown. Both the p_s and p_n values of all these seven *S*-linked genes were much lower than the corresponding values of *S-RNase*.

Construction and screening of an S_1S_1 *BAC library*

To examine the degree of allelic sequence diversity in the non-coding regions (i.e., introns and flanking regions) of the *S*-linked genes and to prepare for comparative studies of the *S* locus, a long-term goal, we first constructed an S_1S_1 BAC library with the procedure described by McCubbin *et al.* (2000b). This library contained 84 892 clones, and analysis of 120 randomly chosen clones showed that 87% of the clones contained genomic DNA inserts, with an average insert size of 98 kb. This library represented a 6.2-fold genome coverage. Screening of the library yielded positive clones for the S_1*-RNase* gene and all the 13 *S*-linked genes analyzed in this work (Table 3). The BAC clone(s) identified for each gene was (were) confirmed by comparing the hybridization patterns of restriction digests of BAC DNA with those of genomic DNA of S_1S_1 and S_2S_2 genotypes (data not shown).

Sequence analysis of S_1 *and* S_2 *BAC clones containing* 3.16*,* G221*,* G261 *and* S-RNase

We chose to examine the degree of allelic sequence diversity of *G221* and *3.16* (both of which were mapped outside the *S* locus), *G261* (which was completely linked to the *S* locus), and *S-RNase*. The clones previously isolated from the S_2S_2 BAC library (McCubbin *et al.*, 2000b) and the clones isolated from the S_1S_1 BAC library were used for sequencing analysis. The extent of sequencing for the S_1 and S_2 alleles of each of these genes and the results of allelic sequence comparison are shown in Figure 5. Comparison of the cDNA sequences with their corresponding genomic sequences revealed that *3.16* contained four exons, *S-RNase* contained two exons (as had been shown previously; Coleman and Kao, 1992), and *G261* contained one exon. The genomic sequences of both alleles of *G221* lacked the first codon (GCT) of the partial *G221-*S_1 cDNA, but contained the remaining cDNA sequence except for the poly(A)$^+$ tail. This suggests that the *G221* genomic sequences contained only the last exon, as well as an intron and the 3′-flanking sequence.

The sequences of the S_1 and S_2 alleles of *S-RNase* only shared 48.2% and 48.1% identity in their 5′-flanking region (2 kb) and 3′-flanking region (2 kb), respectively. Thus, these two regions were even more divergent than the coding region (73.0% and 68.7% identity in the two exons) and the intron (66.7% identity). These results confirm our previous finding from comparison of more limited 5′-flanking and 3′-flanking sequences of the S_1 and S_3 alleles of *S-RNase* (Coleman and Kao, 1992). In contrast, *3.16*, *G221*, and *G261* all showed very low degrees of allelic sequence diversity in their flanking regions. For *G221* and *G261*, the allelic sequence divergence was due to small ($\leq$ 15 bp) insertions/deletions and nucleotide substitutions, except for a 30 bp insertion/deletion in the 5′-flanking region of *G261*. For *3.16*, a number of large (24–646 bp) insertions/deletions were found in the introns and flanking regions. When these large insertions/deletions were excluded from the calculation, the allelic nucleotide sequence identity of all the flanking and intron regions of *3.16*, *G221*, and *G261* ranged from 93.1% (for the 5′-flanking region of *3.16*) to 98.5% (for the 3′-flanking region of both *3.16* and *G261*).

Discussion

In this work, we used 13 pollen-expressed genes of *P. inflata* which we had previously identified as being linked to *S-RNase* to characterize the *S* locus that controls SI. The linkage of these genes to the *S* locus had previously been established based on recombination analysis of 100 plants. Here, we expanded the analysis to a total of 1205 plants segregating for S_1 and S_2 haplotypes, and the results have allowed us to construct a genetic map of the *S* locus.

Nine of the S*-linked marker genes lie in a large chromosomal region where recombination is suppressed*

No recombination with *S-RNase* was observed for nine of the 13 *S*-linked genes, whereas recombination was observed for the remaining four *S*-linked genes, *3.16*, *G211*, *G212* and *G221*. In all, recombination was found in nine of the 1205 F_2 plants analyzed. Since

```
                                   F-Box motif
A113-S1     MKE--LP-QDVVIYILVMLPVKSLLRFKCSCKTFCNIIKSSTFINLHLNHTTNVKDELVL 57
A113-S2     ***--**-***********************************V********K******* 57
A113-S3     ***--**-************************************Y*************** 57
A134-S1     ***--**-*******F*************T****YH*****************F****** 57
A134-S2     ***--**-*******F*************T*****H*****************F****** 57
A134-S3     ***--**-*******F**********S**T*****H*****************F****** 57
Ah-SLF-S2   *MDRRF*R***ISE**LFSS*******R*VS*SW*SL***ND**DN**LRR-QTNGNVMV 60

A113-S1     LKRSFKTDE---YNFYK----SIISFLSSKEDYDFKLISADVEIPHLTTTSACVFHQLIG 110
A113-S2     *********---*****----**L**F********MPM*P************R******* 110
A113-S3     ********D---*****----**L*******G****S**P******************** 110
A134-S1     *****E***---*****----**L***FA*******P**P**********A**IC*R*** 110
A134-S2     *****E***---*K***----**L***FA*******P**P**********A***C*R*** 110
A134-S3     *****E***---*****----**L***FA*K*****P**P**K*******A**IC*R*** 110
Ah-SLF-S2   V**YVR*P*RDMFS**NINSPELDEL*PDLPNPY**N*KFDYDYFY*PQRV-----N*M* 114

A113-S1     PCNGLIALTDSLTTIVFNPATRKYRLIPPCPFGIPRGFRRSISGIGFGFNSDAND-YKVV 169
A113-S2     ****************************************************V**-**** 169
A113-S3     *******************************************************-**** 169
A134-S1     ******V**************L*********************************-**** 169
A134-S2     ******V**************L*********************************-**** 169
A134-S3     ******V**************L*********************************-**** 169
Ah-SLF-S2   ******C*AYGDCVLLS***L*EIKRL**T**AN*E*HCTD*I*Y***--NTC**C**** 172

A113-S1     RLSEVYKEPCDKEMKVDIYDFSVDSWRELLGQEVPIVY--WLPCAEILYKRNFHWFAFAD 227
A113-S2     **************************************--*****D**F*********** 227
A113-S3     **************************************--******************** 227
A134-S1     ***********************************F*F--*F****************** 227
A134-S2     ***********************************F*F--*F****************** 227
A134-S3     *******GT***K**********************F*F--*F****************** 227
Ah-SLF-S2   LIES*--G*E*HHINIYV*YSDTN**KHIEDDST**K*ICHF**N*LFF*GA***N*NST 230

A113-S1     DVV----ILCFDMNTEKFHNMGMPDACHFDDGKCY-GLVILCKCMTLICYPDPMPSSPTE 282
A113-S2     ***----****************************-*********S************** 282
A113-S3     ***----*************L**************-************************ 282
A134-S1     ***----**************************S*-*****F****************** 282
A134-S2     V**----****************************-*****F***********K****** 282
A134-S3     ***----**************************S*-*****F****************** 282
Ah-SLF-S2   *IFYADF**T**II**V*KE*AY*H-*LAQFSNSFLS*MS*NE*LAMVR*KEW*EDP--- 286

A113-S1     KLTDIWIMKEYGEKESWIKRCSI-RLLPESPLAVWK-DEILLLQSKMGHLIAYDHNSDEV 340
A113-S2     ***********************-***********N-*********************** 340
A113-S3     ***********************-***********M-*********I************* 340
A134-S1     ***********************-************-******H**********L**N** 340
A134-S2     *****************M*****-************-******H******M***L**N** 340
A134-S3     ***********************-************-******H**********L**N** 340
Ah-SLF-S2   E*F***V*NQ**VR***T*QYV*GPQVVVCSHVC**N**C*IVEDGN*Q*VSCAFRTNKI 346

A113-S1     KELDLHGLPTSLRVIIYRESLTPIPRSKDSIELEQF     376
A113-S2     ************************************     376
A113-S3     **********************L********D****     376
A134-S1     Q******Y*E***I********A***NN*C***QN*RCN  379
A134-S2     Q******Y*E***I********A***NN*C***QN*RCN  379
A134-S3     Q******Y*E***I********A***NN*C***QN*RCN  379
Ah-SLF-S2   EK*PIYAVEET***L*VD***ISLN*V---LN*        376
```

Figure 4. Alignment of the deduced amino acid sequences of *A113*, *A134* and *Antirrhinum hispanicum SLF-S_2* (*Ah-SLF-S_2*). Three alleles of *A113* (*A113-S_1*, -S_2, -S_3) and *A134* (*A134-S_1*, -S_2, -S_3) are included in the sequence comparison. Asterisks represent residues identical to those of *A113-S_1*. Gaps are introduced to optimize the alignment. The F-box domain is overlined.

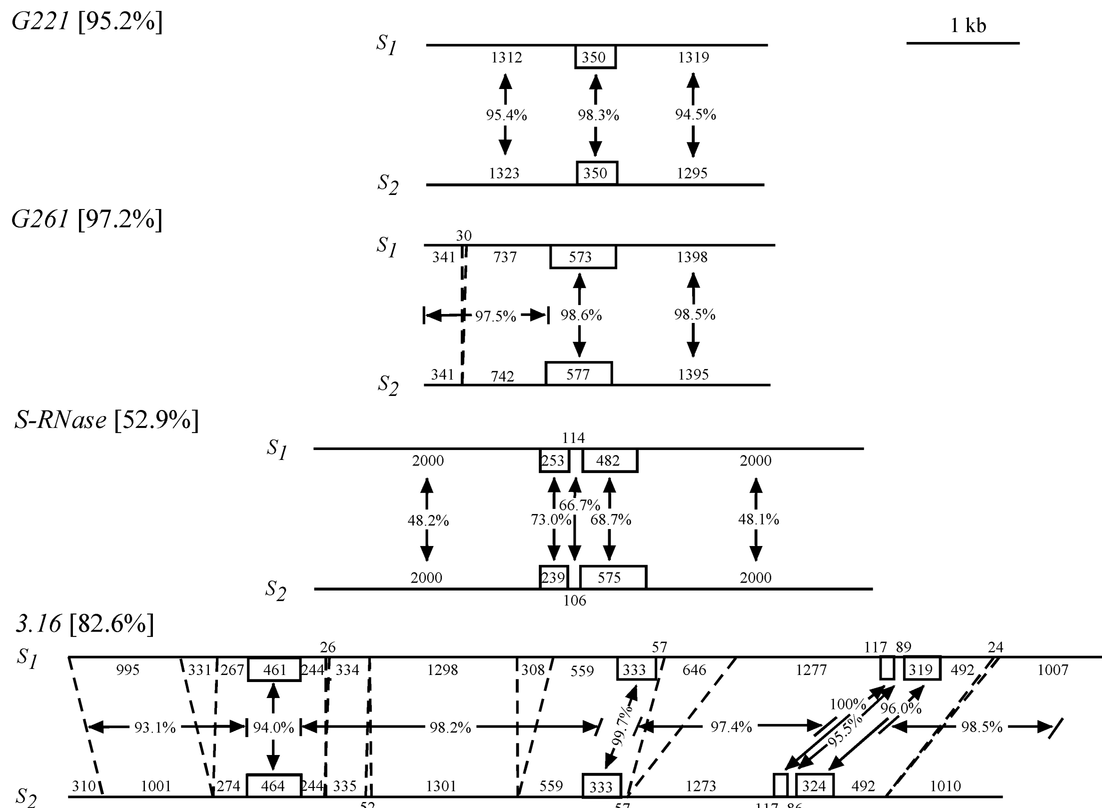

Figure 5. Schematic representation of allelic sequence identity of *G221*, *G261*, *3.16* and *S-RNase*. The horizontal lines (drawn to scale) represent the sequenced regions (in 5′ to 3′ orientation) of S_1 and S_2 alleles of these four genes. The boxed regions represent the exons. The double-headed arrows indicate the corresponding regions of S_1 and S_2 alleles compared; the size (in bp) of each region compared is shown above or below the horizontal line and the numbers (as a percentage) indicate nucleotide sequence identities. Dash lines were introduced to show large (≥ 24 bp) insertions/deletions; insertions/deletions smaller than 15 bp are not shown. The overall allelic sequence identity for each gene is shown in brackets. See Table 2 for the GenBank accession numbers.

all these nine plants remained self-incompatible and showed normal SI behavior, none of the recombination events disrupted the function of SI. Thus, all the genes required for SI specificity must reside within the chromosomal region that contains the nine genes tightly linked to *S-RNase*. The *S* locus of *P. inflata* was mapped to a 0.25 cM region defined by *3.16* and *G221*.

Recombination analysis was previously carried out on the *S* locus of *N. alata*. Li *et al.* (2000) carried out genetic mapping of the *S* locus in *N. alata* by analyzing 154 to 215 F_2 plants using five *S*-linked genes: 48*A*, 133*G* and 167*A* of *N. alata* identified by differential display, and *CP100* and *CP108* of *S. tuberosum* (Gebhardt *et al.*, 1991). *133G* and *167A* were mapped to the same side of *S-RNase* at 2.7 cM and 0.9 cM from *S-RNase*, respectively, whereas the other markers were found tightly linked to *S-RNase*. Since no markers on the opposite side of *S-RNase* were found, the maximum size (in cM) of the *S* locus is unknown.

To date, for the solanaceous species that possess the S-RNase type of GSI, there is no report of the physical size of the *S* locus, or the ratio of physical distance to genetic distance in the *S*-locus region. These nine genes that are tightly linked to *S-RNase* provide useful markers for molecular cloning and characterization of the *S*-locus region. We have begun chromosomal walking from multiple points of the *S* locus as defined by these nine genes and *S-RNase*, using the BAC clones isolated from the S_2S_2 BAC library (McCubbin *et al.*, 2000b). The construction of a BAC contig of the *S* locus will allow us to make an integrated genetic and physical map and perform comparative studies of the *S* locus. We have obtained 10 separate S_2S_2 BAC contigs, each containing one of the nine marker genes or S_2-*RNase*. The 10 contigs

Table 3. Number and insert sizes of BAC clones containing *S*-linked genes.

S-linked gene	Number of positive clones	Insert size (kb)
S_1-*RNase*	1	79
3.2	2	110, 122
3.15	1	121
3.16	1	100
A113	1	118
A134	2	117, 130
A181	2	105[a], 110[a]
A301	1	65
X9	3	69, 105[a], 110[a]
G211	1	110
G212	1	153
G221	1	117
G261	1	110
X11	5	98, 100, 100, 102, 121

[a]Two BAC clones contain both *A181* and *X9*.

collectively span a region of 4.4 Mb, suggesting that the chromosomal region where recombination was not observed is at least 4.4 Mb in size (T. Tsukamoto, Y. Wang, K.-W. Yi, A.G. McCubbin, and T.-h. Kao, unpublished results). The ratio of the physical size to the genetic distance in the region delimited by *3.16* and *G221* is thus at least 17.6 Mb/cM. *P. inflata* has a haploid genome size of 1158 Mb (Bennet and Leitch, 1995), slightly larger than that of tomato (954 Mb). If we assume that these two species have the same size linkage maps (1300 cM; Tanksley *et al.*, 1992), 1 cM in *P. inflata* would be equivalent to 0.9 Mb on average. The large deviation from this ratio in the *S*-locus region is consistent with the notion that recombination is suppressed as a result of its sub-centromeric location (Entani *et al.*, 1999). Because two separate genes at the *S* locus control the pollen and pistil specificities in SI, suppression of recombination allows the maintenance of SI without disrupting the linkage between these two genes.

The genes that are tightly linked to the *S* locus can also be used to map the pollen *S* gene at the *S* locus. Golz *et al.* (2001) examined pollen-part mutants of *N. alata* that contained duplication of part or the entire *S* locus, for the presence or absence of several *S*-linked markers on the duplicated *S*-locus fragments. They were able to determine the order of these markers and to locate the pollen *S* gene between marker *48A* and *S-RNase*. However, the physical distances between *48A* and the pollen *S* gene, and between *S-RNase* and the pollen *S* gene are unknown.

It is interesting that a number of pollen-expressed genes are tightly linked to the *S* locus of *P. inflata*. We are currently using the approach of cDNA selection to identify any additional *S*-linked genes that are expressed in pollen and/or other tissues. In *Lycopersicon hirsutum*, genes controlling several reproductive traits (e.g., floral size) have been mapped to the *S* locus (Bernacchi and Tanksley, 1997). Whether any of the genes we have identified is involved in these traits or in any other reproductive development remains to be determined.

S-*linked marker genes exhibit low degrees of allelic sequence diversity*

Sequence analysis of the cDNAs for seven marker genes, *3.2*, *3.15*, *3.16*, *A113*, *A134*, *G221* and *G261* has revealed that these genes exhibit very low degrees of allelic sequence diversity in their coding regions (0.4–3.5%). Further genomic sequencing analysis of three of these genes (*3.16*, *G221* and *G261*) has shown that the low degrees of allelic sequence diversity extend into their flanking regions and introns (1.5–6.9%). This is in sharp contrast to *S-RNase*, whose S_1, S_2 and S_3 alleles show 14.6–19.7% allelic sequence diversity in the coding region and 33.3–53.5% sequence diversity in the flanking regions and introns (this work; Coleman and Kao, 1992). Thus, although *3.2*, *3.15*, *A113*, *A134* and *G261* are tightly linked to *S-RNase* and expressed in pollen, none of them are likely to be the pollen *S* gene.

One explanation for the presence of the genes with very low degrees of allelic sequence diversity at the *S* locus is that recombination has occurred in some regions of the *S* locus, even though intergenic recombination between the pollen *S* gene and *S-RNase* is suppressed to prevent the breakdown of SI. That is, the extent of recombination suppression in different regions of the *S* locus may be different. The possible occurrence of recombination is also revealed from the finding of much lower p_s and p_n values of *3.2*, *3.15*, *3.16*, *A113*, *A134*, *G221* and *G261* than the corresponding values of *S-RNase*. Synonymous substitutions are likely to be subject to little purifying selection, whereas the majority of non-synonymous substitutions would be eliminated by purifying selection because of the deleterious effect the mutations have on the function of a protein. Thus, theoretically, the synonymous substitution rate of a gene reflects

the mutation rate. The much lower p_s values of these *S*-linked genes suggest that they have different evolutionary histories from *S-RNase*, and that they are derived from more recent ancestor genes, due to more frequent recombination events. Since recombination is expected to produce similarity at both synonymous and non-synonymous sites, the much lower p_n values of these genes than that of *S-RNase* further supports this notion. Our previous phylogenetic study of 15 alleles of *S-RNase* of *P. inflata* also revealed evidence of intragenic recombination in *S-RNase*. However, in this case, we hypothesized that intragenic recombination would shuffle polymorphic sequences resulting from point mutations into different allelic combinations to facilitate allelic diversity of *S-RNase* (Wang *et al.*, 2001).

For the *S* locus of *Brassica*, two non-polymorphic genes, *SLL1* and *SLL2*, are located near three highly polymorphic genes, *SLG* (encoding the *S*-locus glycoprotein), *SRK* (encoding the *S*-locus receptor kinase, the female determinant of SI) and *SCR* (encoding the *S*-locus cysteine-rich protein, the male determinant) (Cui *et al.*, 1999; Schopfer *et al.*, 1999; Suzuki *et al.*, 1999). It has been proposed that the region containing *SLL1* and *SLL2* was transposed into the *S* locus of a certain *S* haplotype after the sequence polymorphism of the SI genes had been established, and *SLL1* was subsequently acquired by other *S* haplotypes through recombination and/or gene conversion (Yu *et al.*, 1996). A similar mechanism may account for the existence of the low polymorphic genes at the *S* locus of *P. inflata*. Since, as stated above, the nine genes that are tightly linked to *S-RNase* could be located as far as 4 Mb away from *S-RNase*, if transposition did occur, a large chromosomal region would have to be involved.

Multiple F-box genes are linked to the S *locus*

From genomic DNA blot analysis of the *S* locus of *Prunus dulcis* (almond, a rosaceous species) which possesses the S-RNase type of GSI, Ushijima *et al.* (2001) also suggested that recombination might have occurred in the regions flanking a ca. 70 kb region that contains *S-RNase*. This was based on the finding that the sequences within the ca. 70 kb region were highly divergent between different *S* haplotypes, whereas the sequences outside this region were more similar between different *S* haplotypes. However, the extent of the *S*-locus region where recombination is suppressed has not yet been determined by genetic analysis. Sequence analysis of the ca. 70 kb region has uncovered two pollen-expressed genes, both of which encode F-box-containing proteins (Ushijima *et al.*, 2003). *SFB* (for *S*-haplotype-specific F-box gene) exhibits a high level of allelic sequence diversity, with 69.0–76.5% identity at the amino acid sequence level among four *S* alleles compared. *SLF* (for *S*-locus F-box gene) exhibits a much lower level of allelic sequence diversity, with 95.1% identity at the amino acid sequence level between two alleles compared.

Sequence analysis of two *S*-locus regions (a 64 kb region containing the S_1*-RNase* gene and a 62.5 kb region containing the S_7*-RNase* gene) of another rosaceous species, *P. mume* (Japanese apricot), has also revealed the presence of multiple F-box genes (Entani *et al.*, 2003). *SLF*, like *SFB* of *P. dulcis*, shows a high level of allelic sequence diversity (77.8 to 81.3% amino acid sequence identity among three alleles compared). *SLFL1* and *SLFL2* (for *SLF*-like gene 1 and 2), like *SLF* of *P. dulcis*, show a much lower level of allelic sequence diversity (92.5% and 99.2% amino acid sequence identity between S_1 and S_7 alleles, respectively). Although only one F-box gene, *AhSLF*, was initially identified in a 69 kb chromosomal region of *Antirrhinum* that contains the S_2*-RNase* gene (Lai *et al.*, 2002), subsequent sequencing of a more extended region of the S_2 haplotype, as well as of the *S*-locus regions of additional *S* haplotypes, has shown that *Antirrhinum* also has multiple F-box genes at the *S* locus (Zhou *et al.*, 2003). Contrary to *SFB* of *P. dulcis* and *SLF* of *P. mume*, all of these F-box genes show higher than 90% allelic sequence diversity.

The level of allelic sequence diversity of *A113* and *A134* is comparable to that of *SLF* of *P. dulcis*, *SLFL1* and *SLFL2* of *P. mume*, and the F-box genes of *Antirrhinum*. However, *A113* and *A134* appear to be located at a much greater physical distance from the *S-RNase* gene than these other F-box genes. Our chromosome walking through the S_2-locus region has shown that none of the nine genes that are tightly linked to *S-RNase* is located within an ca. 881 kb contig that contains the S_2*-RNase* gene (with ca. 180 kb upstream and ca. 700 kb downstream regions), and that some of these nine genes could be 4 Mb away from *S-RNase* (Y. Wang, T. Tsukamoto, A.G. McCubbin and T.-h. Kao, unpublished data). Although none of the *S*-linked marker genes of *P. inflata* identified by mRNA differential display and subtractive hybridization appears to be a homologue of the highly polymorphic F-box genes identified in the rosaceous species, a putative homologue has been identified from the sequence ana-

lysis of a 328 kb region of the *S* locus that contains the S_2-*RNase* gene (X. Wang, A. G. McCubbin, S. Huang, Y. Wang and T.-h. Kao, unpublished results). Whether this F-box gene is the long sought-after pollen *S* gene is currently being investigated.

No correlation between degree of allelic sequence diversity and genetic distance from S-RNase

Sliding-window analysis has been widely used to study sequence polymorphism at a locus linked to a second locus that is under balancing selection (e.g., Hudson and Kaplan, 1988). The extent of polymorphism at such a locus is a function of the extent of recombination between the locus and the one under selection. High pairwise differences at sites absolutely linked to the selected site reflect the expansion of divergence time induced by balancing selection, with progressively lower sequence differences expected as linkage loosens. Among the seven *S*-linked genes whose allelic sequence diversity we have examined in this work, there does not appear to be a correlation between the degree of allelic sequence diversity and the genetic distance from *S-RNase*. *3.2*, *3.15*, *A113*, *A134* and *G261*, all of which are tightly linked to *S-RNase*, exhibit similar low degrees of allelic sequence diversity as do *3.16* and *G221*, both of which are located at a much greater genetic distance from *S-RNase* (Figure 3). In fact, among all the allelic comparisons, the S_1 and S_2 alleles of *3.16* exhibit the highest degree of overall nucleotide sequence diversity, with several large blocks of insertion/deletion in the non-coding regions, whereas the S_1 and S_2 alleles of *G261* exhibit the highest degree of sequence identity (Figure 5). Furthermore, the deduced amino acid sequences of *G261-*S_1 and *G261-*S_2 are completely identical, whereas those of *3.16-*S_1 and *3.16-*S_2 differ in one residue.

It is possible that the tightly linked genes we have examined for allelic sequence diversity are sufficiently distant from *S-RNase* that the level of diversity is virtually indistinguishable from that of genes at unlinked loci. Strobeck (1980) showed that the increase in allelic sequence diversity due to linkage to an SI locus extends only over a very restricted region around the incompatibility region. Moreover, as discussed earlier, low rates of recombination (undetectable by direct observation) may have occurred between these genes and *S-RNase*. Takebayashi *et al.* (2003) analyzed the magnitude and nature of nucleotide variation of the *S*-linked gene, *48A*, and *S-RNase* of *N. alata*, and concluded that recombination may have occurred between these two genes, even though none was observed from the analysis of 215 plants. Given the ancient divergence of the *S* locus in the solanaceous species (in excess of 30 million years; Ioerger *et al.*, 1990), even very low rates of recombination are sufficient to render the level of diversity of the *S*-linked genes virtually indistinguishable from that of genes unlinked to the *S* locus.

Conclusion

In conclusion, the recombination analysis of 1205 F_2 plants has mapped the *P. inflata S* locus in a 0.25 cM region defined by *3.16* and *G221*. This result along with our preliminary results from chromosome walking of the *S* locus has shown that the *S* locus of *P. inflata* is a huge multigene complex (> 4.4 Mb in size) with the ratio of physical size to genetic distance being at least 17.6 Mb/cM. Allelic sequence comparison showed that this complex contains both highly polymorphic genes involved in SI and a number of pollen-expressed genes with very low degrees of allelic diversity. There is no obvious correlation between the degree of allelic sequence diversity and the genetic distance from *S-RNase*. It will be of interest to identify additional genes that are tightly linked to *S-RNase*, determine the physical distance of all the tightly linked genes to *S-RNase*, and determine the physiological functions of all the genes that are linked to the highly polymorphic *S* locus.

Acknowledgements

We thank Marcy Uyenoyama (Duke University) for valuable comments, the Cornell Research Foundation and Carol Hamilton for providing BIBAC2 vector, Christiane Gebhardt (Max Planck Institute for Plant Breeding Research, Cologne, Germany) for providing *CP100*, and Ed Newbigin (University of Melbourne, Melbourne, Australia) for providing *48A*. We also thank Anthony H. Omeis for the greenhouse work and Jiong Wang for routine lab assistance. This work was supported by grants from the National Science Foundation (grants IBN-9982659 and IBN-0235176 to T.-h.K).

References

Ai, Y., Singh, A., Coleman, C.E., Ioerger, T.R., Kheyr-Pour, A. and Kao, T.-h. 1990. Self-incompatibility in *Petunia inflata*: isolation and characterization of cDNAs encoding three *S*-allele-associated proteins. Sex. Plant Reprod. 3: 130–138.

Bennet, M.D. and Leitch, I.J. 1995. Nuclear DNA amounts in angiosperms. Ann. Bot. (Lond.) 76: 113–176.

Bernacchi, D. and Tanksley, S.D. 1997. An interspecific backcross of *Lycopersicon esculentum* × *L. hirsutum*: linkage analysis and a QTL study of sexual compatibility factors and floral traits. Genetics 147: 861–877.

Coleman, C.E. and Kao, T.-h. 1992. The flanking regions of two *Petunia inflata S*-alleles are heterogeneous and contain repetitive sequences. Plant Mol. Biol. 18: 725–737.

Copenhaver, G.P., Nichel, K., Kuromori, T., Benito, M.-I., Kaul, S., Liu, X., Bevan, M., Murphy, G., Harris, B., Parnell, L.D., McCombie, W.R., Martienssen, R.A., Marra, M. and Preuss, D. 1999. Genetic definition and sequence analysis of *Arabidopsis* centromeres. Science 24: 2468–2474.

Cui, Y., Brugiere, N., Jackman, L., Bi, Y.-M. and Rothstein, S.J. 1999. Structural and transcriptional comparative analysis of the *S* locus regions in two self-incompatible *Brassica napus* lines. Plant Cell 11: 2217–2231.

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

Dowd, P.E., McCubbin, A.G., Wang, X., Verica, J.A., Tsukamoto, T., Ando, T. and Kao, T.-h. 2000. Use of *Petunia* as a model for the study of Solanaceous type self-incompatibility. Ann. Bot. 85 (Suppl A): 87–93.

Entani, T., Iwano, M., Shiba, H., Che, F.-S., Isogai, A. and 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.

Entani, T., Iwano, M., Shiba, H., Takayama, S., Fukui, K. and Isogai, A. 1999. Centromeric localization of an S-RNase gene in *Petunia hybrida* Vilm. Theor. Appl. Genet. 99: 391–397.

Gebhardt, C., Ritter, E., Barone, A., Debener, T., Walkemeier, B., Schachtschbel, U., Kaufmann, H., Thompson, R.D., Bonierbale, M.W., Ganal, M.W., Tanksley, S.D. and Salamini, F. 1991. RFLP maps of potato and their alignment with the homeologous tomato genome. Theor. Appl. Genet. 83: 49–57.

Golz, J.F., Oh, H.-Y., Su, V., Kusaba, M. and Newbigin, E. 2001. Genetic analysis of *Nicotiana* pollen-part mutants is consistent with the presence of an S-ribonuclease inhibitor at the *S*-locus. Proc. Natl. Acad. Sci. USA 98: 15372–15376.

Hamilton, C.M. 1997. A binary-BAC system for plant transformation with high-molecular-weight DNA. Gene 200: 107–116.

Hudson, R.R. and Kaplan, N.L. 1988. The coalescent process in models with selection and recombination. Genetics 120: 831–840.

Ioerger, T.R., Clark, A.G. and Kao, T.-h. 1990. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc. Natl. Acad. Sci. USA 87: 9732–9735.

Kanayama, Y., Dai, N., Granot, D., Petreikov, M., Schaffer, A. and Bennett, A.B. 1997. Divergent fructokinase genes are differentially expressed in tomato. Plant Physiol. 113: 1379–1384.

Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 17: 1244–1245.

Lai, Z., Ma, W., Han, B., Liang, L., Zhang, Y., Hong, G. and Xue, Y. 2002. An F-box gene linked to the self-incompatibility (*S*) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. Plant Mol. Biol. 50: 29–42.

Lee, H.-S., Huang, S. and Kao, T.-h. 1994. S proteins control rejection of incompatible pollen in *Petunia inflata*. Nature 367: 560–563.

Li, J.-H., Nass, N., Kusaba, M., Dodds, P.N., Treloar, N., Clarke, A.E. and Newbigin, E. 2000. A genetic map of the *Nicotiana alata S* locus that includes three pollen-expressed genes. Theor. Appl. Genet. 100: 956–964.

Martinez-Barajas, E., Luethy, M.H. and Randall, D.D. 1997. Molecular cloning and analysis of fructokinase expression in tomato (*Lycopersicon esculentum* Mill.) Plant Sci. 125: 13–20.

McCubbin, A.G., Wang, X. and Kao, T.-h. 2000a. Identification of self-incompatibility (*S*-) locus linked pollen cDNA markers in *Petunia inflata*. Genome 43: 619–627.

McCubbin, A.G., Zuniga, C. and Kao, T.-h. 2000b. Construction of a binary bacterial artificial chromosome library of *Petunia inflata* and the isolation of large genomic fragments linked to the self-incompatibility (*S*-) locus. Genome 43: 820–826.

Murfett, J., Atherton, T.L., Mou, B., Gasser, C.S. and McClure, B.A, 1994. S-RNase expressed in transgenic *Nicotiana* causes *S*-allele-specific pollen rejection. Nature 367: 563–566.

Nei, M. and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3: 418–426.

Pearson, W.R., Wood, T., Zhang, Z. and Miller, W. 1997. Comparison of DNA sequences with protein sequences. Genomics 46: 24–36.

Schopfer, C.R., Nasrallah, M.E. and Nasrallah, J.B. 1999. The male determinant of self-incompatibility in *Brassica*. Science 286: 1697–1700.

Skirpan, A.L., McCubbin, A.G., Ishimizu, T., Wang, X., Hu, Y., Dowd, P.E., Ma, H. and Kao, T.-h. 2001. Isolation and characterization of kinase interacting protein 1, a pollen protein that interacts with the kinase domain of PRK1, a receptor kinase of *Petunia*. Plant Physiol. 126: 1480–1492.

Strobeck, C. 1980. Heterozygosity of a neutral locus linked to a self-incompatibility locus or a balanced lethal. Evolution 34: 779–788.

Suzuki, G., Kai, N., Hirose, T., Fukui, K., Nishio, T., Takayama, S., Isogai, A., Watanabe, M. and Hinata, K. 1999. Genomic organization of the *S* locus: identification and characterization of genes in *SLG*/*SRK* region of S^9 haplotype of *Brassica campestris* (syn. *rapa*). Genetics 153: 391–400.

Takebayashi, N., Brewer, P.B., Newbigin, E. and Uyenoyama, M.K. 2003. Patterns of variation within self-incompatibility loci. Mol. Biol. Evol., 20: 1778–1794.

Tanksley, S.D., Ganal, M.W., Prince, J.P., de Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., Messerguer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Röder, M.S., Wing, R.A., Wu, W. and Young, N.D. 1992. High-density molecular linkage maps of the tomato and potato genomes. Genetics 132: 1141–1160.

ten Hoopen, R., Harbord, R.M., Maes, T., Nanninga, N. and Robbins, T.P. 1998. The self-incompatibility locus in *Petunia hybrida* is located on chromosome III in a region syntenic for the Solanaceae. Plant J. 16: 729–734.

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

Ushijima, K., Sassa, H., Dandekar, A.M., Gradziel, T.M., Tao, R. and Hirano, H. 2003. Structural and transcriptional ana-

lysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. Plant Cell 15: 771–781.

Wang, X., Hughes, A.L., Tsukamoto, T., Ando, T. and Kao, T.-h. 2001. Evidence that intragenic recombination contributes to allelic diversity of the S-RNase gene at the self-incompatibility (*S*) locus in *Petunia inflata*. Plant Physiol. 125: 1012–1022.

Yu, K., Schafer, U., Glavin, T.L., Goring, D.R. and Rothstein, S.J. 1996. Molecular characterization of the *S* locus in two self-incompatible *Brassica napus* lines. Plant Cell 8: 2369–2380.

Zhou, J., Wang, F., Ma, W., Zhang, Y., Han, B. and Xue, Y. 2003. Structural and transcriptional analysis of *S*-locus F-box (*SLF*) genes in *Antirrhinum*. Sex. Plant Reprod. 16: 165–177.