Evidence That Intragenic Recombination Contributes to Allelic Diversity of the S-RNase Gene at the Self-Incompatibility (*S*) Locus in *Petunia inflata*[1]

Xi Wang, Austin L. Hughes, Tatsuya Tsukamoto, Toshio Ando, and Teh-Hui Kao*

Department of Biochemistry and Molecular Biology (X.W., T.-H.K.) and Intercollege Graduate Degree Program in Plant Physiology (T.-H.K.), Pennsylvania State University, University Park, Pennsylvania 16802; Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208 (A.L.H.); and Faculty of Horticulture, Chiba University, 648 Matsudo, Chiba 271–8510, Japan (T.T., T.A.)

For Solanaceae type self-incompatibility, discrimination between self and nonself pollen by the pistil is controlled by the highly polymorphic S-RNase gene. To date, the mechanism generating the allelic diversity of this gene is largely unknown. Natural populations offer a good opportunity to address this question because they likely contain different alleles that share recent common progenitors. We identified 19 *S* haplotypes from a natural population of *Petunia inflata* in Argentina, used reverse transcriptase-polymerase chain reaction to obtain cDNAs for 15 alleles of the S-RNase gene, and sequenced all the cDNAs. Phylogenetic studies revealed that five of these alleles and two previously identified alleles form a major clade, and that the 5′ region of S_{19} allele was derived from an ancestor allele closely related to S_2, whereas its 3′ region was derived from an ancestor allele closely related to S_8. A similar evolutionary relationship was found among S_3, S_{12}, and S_{15} alleles. These findings suggest that intragenic recombination contributed to the generation of the allelic diversity of the S-RNase gene. Two additional findings emerged from the sequence comparisons. First, the nucleotide sequence of the S_1 allele identified in this work is completely identical to that of the previously identified S_1 allele of a different origin. Second, in the two hypervariable regions HVa and HVb, thought to be involved in determining *S* allele specificity, S_6 and S_9 alleles differ only by four nucleotides, all in HVb, resulting in two amino acid differences. The implications of these findings are discussed.

Self-incompatibility (SI) in flowering plants is a prezygotic reproductive barrier that enables the pistil of a flower to reject self-pollen, but to accept nonself pollen for fertilization (de Nettancourt, 1977). This inbreeding-prevention strategy is widespread in flowering plants, and recent molecular data suggest that it has evolved independently in different families multiple times during the course of evolution (for review, see McCubbin and Kao, 2000). In the simplest case the specificity of SI interactions is controlled by a multiallelic locus called the *S* locus. SI can be further classified into gametophytic and sporophytic types. For gametophytic (G)SI, the SI phenotype of the pollen is determined by its own *S*-genotype, whereas for sporophytic (S)SI, the SI phenotype of the pollen is determined by the *S* genotype of the pollen-producing plant.

GSI has been extensively studied in the Solanaceae, which includes *Petunia*, potato, tobacco, and tomato. The S-RNase gene, a highly polymorphic pistil-specific gene at the *S* locus, has been shown to encode the *S* haplotype specificity determinant of the pistil (Lee et al., 1994; Murfett et al., 1994). Comparison of the deduced amino acid sequences of Solanaceous S-RNases has revealed five conserved (C1–C5) and two hypervariable (HVa and HVb) regions (Ioerger et al., 1991; Tsai et el., 1992). Because the S-RNase gene does not control pollen function in SI, another polymorphic gene (yet unidentified) at the *S* locus, named the pollen *S* gene, is thought to play this role. For SSI, most research so far has focused on the family Brassicaceae. Three highly polymorphic genes have been identified at the *S* locus, *SLG*, *SRK*, and *SCR*, and the latter two determine the *S* haplotype specificity of the stigma (Takasaki et al., 2000) and pollen (Schopfer et al., 1999; Takayama et al., 2000), respectively.

The molecular mechanisms responsible for the generation of haplotype diversity at the *S* locus have been studied in Brassicaceae and Solanaceae SI systems. The central question is whether point mutation alone or point mutation and recombination/gene conversion has contributed to the generation of allelic diversity. Results from domain-swapping experiments suggest that point mutations in the S-RNase gene can generate new allelic specificities (Matton et al., 1997). However, so far no evidence for recombination/gene conversion at the S-RNase gene has been obtained.

In fact, the following findings have even been interpreted to mean that recombination does not contribute at all to the generation of *S* haplotype diver-

[1] This work was supported by the National Science Foundation (grant nos. IBN–9603993 and IBN–9982659 to T.-H.K.) and by the National Institutes of Health (grant no. GM34940 to A.L.H.).

* Corresponding author; e-mail: txk3@psu.edu; fax 814–863–9416.

sity in the Solanaceae SI system. First, the *S* locus contains the S-RNase gene and the pollen *S* gene, so that recombination would destroy the tight linkage between these two genes, and thus cause breakdown of SI by generating different haplotype specificities for the pistil and pollen. Second, the flanking regions of different alleles of the S-RNase gene share virtually no sequence similarity and contain abundant repetitive sequences (Kaufmann et al., 1991; Coleman and Kao, 1992; Matton et al., 1995), indicative of a lack of recombination in the vicinity of the S-RNase gene. Third, no evidence of intragenic recombination was found when sensitive statistical methods were used to examine cDNA sequences for 12 alleles of the S-RNase gene for clustering of polymorphic sites (Clark and Kao, 1991). Last, the *S* locus of *Lycopersicon peruvianum* has been shown by genetic mapping to be located close to the centromere of chromosome I (Bernatzky, 1993), and the *S* locus of *Petunia hybrida* has been shown by fluorescence in situ hybridization to be located close to the centromere of chromosome III (Entani et al., 1999), the centromere being a region where recombination is known to be suppressed. In the Brassicaceae SI system, extensive sequence divergence and rearrangements were observed when comparing the *S* locus of two different *S* haplotypes (Boyes et al., 1997), suggesting that recombination is likely to be suppressed at the Brassicaceae *S* locus as well. No recombination has been detected between *SLG* and *SRK*, even though the physical distance between these two genes may exceed 200 kb in some haplotypes (Boyes and Nasrallah, 1993; Casselman et al., 2000), nor has recombination between the pollen *S* gene and the S-RNase gene of the Solanaceae SI system ever been observed (de Nettancourt, 1977; Kheyr-Pour and Pernes, 1986).

However, these aforementioned results do not completely rule out the involvement of recombination. First, even though no recombination at the *S* locus has ever been observed in genetic studies, it does remain possible that recombinants are generated, but at a very low frequency, and that those that become self-compatible have been eliminated from the population. Second, the high degree of sequence diversity in the flanking regions of the *S* genes might prevent intergenic, but not intragenic recombination. Moreover, the S-RNases available to Clark and Kao (1991) for examining the possibility of intragenic recombination by statistical methods are too divergent in sequence to allow detection of any ancient recombination events because accumulation of neutral mutations over long evolutionary time would have obscured the exchanges.

In another more recent study of the Brassicaceae SI system, Kusaba et al. (1997) compared the sequences of six class I alleles of the *SLG* gene of *Brassica oleracea* and obtained different phylogenetic tree topologies for different regions of the gene, thus providing the first evidence for the occurrence of intragenic recombination among the alleles of an *S* locus gene. Awadalla and Charlesworth (1999) subsequently analyzed linkage disequilibrium within the *SLG* genes of *B. oleracea* and *Brassica campestris*, and found that it declines with distance within the *SLG* gene, providing further support for the existence of recombination in the Brassicaceae SI system.

These findings of possible intragenic recombination at the Brassicaceae *S* locus prompted us to reexamine this possibility for the *S* locus of the Solanaceae. In this study we used plants derived from seeds collected from a natural population of *Petunia inflata* in Argentina for the cloning of S-RNase sequences. Plants from the same population are more likely to contain alleles that are closely related and share recent progenitor alleles. We identified 19 different *S* haplotypes in this population and used reverse transcription (RT)-PCR to obtain cDNAs corresponding with 15 alleles of the S-RNase gene. Phylogenetic analyses using the deduced amino acid sequences of these alleles and a number of previously identified alleles showed that five of the alleles identified from this natural population and two previously identified alleles of *P. inflata* form a major clade. These closely related alleles allowed us to more precisely assess the occurrence of intragenic recombination between different alleles of the S-RNase gene and also to identify some other interesting features about the Solanaceae SI system.

RESULTS

Identification of *S* Haplotypes and Cloning of cDNAs for S-RNases of *P. inflata* from a Natural Population in Argentina

One hundred SI plants were raised from seeds randomly collected from a natural population of *P. inflata* found in Argentina, and 36 of them were used in this study. Nineteen different *S* haplotypes were identified based on the results of genetic crosses described in the "Materials and Methods." One of these 19 *S* haplotypes was designated S_1 because its SI behavior was identical to that of one of the three previously identified *S* haplotypes of *P. inflata*, S_1, S_2, and S_3 (Ai et al., 1990). All the other 18 *S* haplotypes were novel and thus designated S_4 through S_{21}. Table I lists the *S* genotype of each of the 36 plants and indicates the plants from which each of the 19 *S* homozygotes were obtained by bud-selfing.

Plants homozygous for all except S_{14} and S_{18} haplotypes were used for cloning cDNAs of S-RNases. We designed a degenerate primer based on the amino acid sequences of the conserved region C2 of solanaceous S-RNases (Ioerger et al., 1991; Tsai et al., 1992). Separate RT-PCR was performed using this primer and an oligo $(dT)_{17}$ primer on total pistil RNA isolated from plants homozygous for each of the 17 *S* haplotypes. A major DNA band of approximately 550 to 600 bp in size was obtained for each of the 17 *S*

Table I. *S Genotypes of plants derived from a natural population of P. inflata and S homozygotes of bud-selfed progeny*

Plant Code	*S* Genotype	*S* Homozygotes of Bud-Selfed Progeny
A9-1	S_4S_5	S_4S_4, S_5S_5
A9-2	S_6S_7	S_6S_6, S_7S_7
A9-3	S_4S_5	
A9-4	S_1S_7	S_1S_1
A9-5	S_4S_8	S_8S_8
A9-6	S_9S_{10}	S_9S_9, $S_{10}S_{10}$
A9-7	S_6S_7	
A9-8	S_5S_7	
A9-9	S_4S_{11}	$S_{11}S_{11}$
A9-10	$S_{12}S_{13}$	$S_{12}S_{12}$, $S_{13}S_{13}$
A9-11	S_5S_{12}	
A9-12	S_7S_{14}	$S_{14}S_{14}$
A9-13	S_6S_{11}	
A9-14	S_4S_5	
A9-15	S_7S_{11}	
A9-16	S_7S_{15}	$S_{15}S_{15}$
A9-17	S_7S_{16}	$S_{16}S_{16}$
A9-18	S_7S_{11}	
A9-19	$S_{12}S_{13}$	
A9-20	S_7S_{15}	
A9-21	$S_{16}S_{17}$	$S_{17}S_{17}$
A9-22	S_7S_{16}	
A9-23	S_7S_{11}	
A9-24	S_4S_9	
A9-25	$S_{11}S_{18}$	$S_{11}S_{18}$
A9-26	S_7S_{19}	$S_{19}S_{19}$
A9-27	$S_{13}S_{17}$	
A9-28	S_7S_{11}	
A9-29	S_4S_{16}	
A9-30	S_5S_{13}	
A9-31	S_7S_{11}	
A9-32	S_8S_{13}	
A9-33	S_5S_7	
A9-34	$S_{13}S_{20}$	$S_{20}S_{20}$
A9-35	$S_{11}S_{21}$	$S_{21}S_{21}$
A9-36	S_4S_{11}	

haplotypes (results not shown). This size range was what would be expected of cDNAs containing the coding sequence for S-RNases (except for the region upstream of C2) plus the 3′-non-coding sequence and the poly(A) tail. This major DNA band for each *S* haplotype was cloned and individual colonies from each transformation were analyzed by DNA gel blotting using a radiolabeled probe containing cDNAs for S_1-, S_2-, and S_3-RNases of *P. inflata* (Ai et al., 1990). Under low stringency membrane washing conditions, cDNA clones for eight of the 17 *S* haplotypes (S_1, S_5, S_6, S_8, S_{12}, S_{13}, S_{15}, and S_{19}) hybridized to the probe, whereas for the remaining nine *S* haplotypes (S_4, S_7, S_9, S_{10}, S_{11}, S_{16}, S_{17}, S_{20}, and S_{21}), none of the cDNA clones hybridized to the probe. The failure to hybridize could possibly be due to greater sequence diversity between the cDNAs for the S-RNases of these haplotypes and the cDNAs for S_1-, S_2-, and S_3-RNases.

Genomic DNA-blot analysis was carried out to further ascertain whether the cDNA obtained by RT-PCR for each *S* haplotype encoded an S-RNase. If a cDNA encodes an S-RNase, it should reveal an *S* haplotype specific restriction fragment-length polymorphism (RFLP) and hybridize most strongly to the genomic fragment(s) from the same *S* haplotype. This is because different alleles of the S-RNase gene are divergent in sequence not only in the coding region, but also in the flanking regions. Blots containing *Eco*RI digests of genomic DNA isolated from plants homozygous for different *S* haplotypes were hybridized separately with the cDNAs isolated from the 17 *S* haplotypes. All except the cDNAs for S_4 and S_5 haplotypes revealed a hybridization pattern expected of cDNAs encoding S-RNases. The results of genomic DNA blotting using the cDNA for S_8 haplotype as a probe are shown in Figure 1. The cDNA hybridized to an *Eco*RI fragment of different size in each of the S_8S_8, $S_{12}S_{12}$, $S_{13}S_{13}$, and $S_{20}S_{20}$ homozygotes, and did not hybridize to the genomic DNA digests of the other eight *S* homozygotes on the blot. Moreover, the intensity of the hybridization to the DNA fragment of S_8S_8 homozygote was the strongest.

The cDNAs for all the 17 *S* haplotypes were completely sequenced, and all except the cDNAs for S_4 and S_5 haplotypes showed sequence similarity with the sequences of known solanaceous S-RNases. Thus, cDNAs for a total of 15 alleles of the S-RNase gene identified from a natural population of *P. inflata* were cloned. Because the upstream primer used in RT-PCR corresponded to the C2 region of S-RNases, the cDNAs obtained encoded approximately 167 amino acids of S-RNases, with approximately 35 amino acids at the N-terminal end not included. It is interesting that the nucleotide sequence (both coding and the 3′-non-coding) of the S_1-RNase gene identified in this work is completely identical to that of the previously identified S_1-RNase gene. Moreover, when the

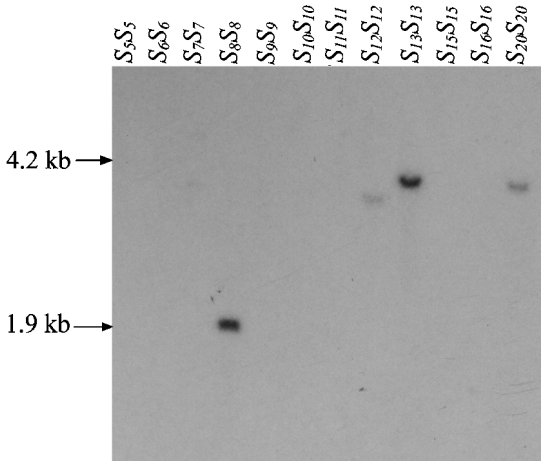

Figure 1. DNA-blot analysis of genomic DNA digests of 12 *S* homozygotes of *P. inflata*. *Eco*RI digests of genomic DNA (5 μg) from each of the *S* homozygotes indicated were hybridized with radiolabeled cDNA that had been obtained by RT-PCR from S_8 haplotype and cloned into pGEM-T.

genomic DNA of these two independent sources of S_1 haplotype was digested with several restriction enzymes and hybridized with the full-length cDNA for S_1-RNase, no RFLP was observed (results not shown), indicating that there is little, if any, sequence difference in the immediate flanking regions of the S-RNase gene. However, when the same genomic DNA blot was hybridized with cDNAs for two pollen-expressed genes that are linked to, but located at as yet undetermined distances from the S-RNase gene (McCubbin et al., 2000), polymorphism was detected (results not shown).

Sequence Comparison among S-RNases

Deduced amino acid sequences of the 15 S-RNases of *P. inflata* identified from this natural population, as well as those of S_2- and S_3-RNases of *P. inflata* were aligned by CLUSTAL X (Thompson et al., 1997), and the alignment is shown in Figure 2A. Three of the five conserved regions and both hypervariable regions are contained in the aligned sequences. The pairwise amino acid sequence identity ranges from 38% to 92%; this wide range of sequence diversity is similar to that found from all previous comparisons of solanaceous S-RNases. Two pairs, S_6-RNase and S_9-RNase, and S_8-RNase and S_{15}-RNase, show the highest degree of sequence identity. The nucleotide sequences of S_6-RNase and S_9-RNase are completely identical in the HVa region, and only differ by four nucleotides in the HVb region, resulting in two different amino acids (Fig. 2B).

Molecular Phylogeny of S-RNases

A phylogenetic tree of the 15 S-RNases of *P. inflata* whose sequences were determined in this work and 19 previously reported solanaceous S-RNases was constructed by the neighbor-joining method based on the proportion of amino acid differences (Fig. 3). The tree was rooted with a fungal RNase (RNase T2) and two S-like RNases (le and lx) from tomato. The tree shows that five of the 15 *P. inflata* S-RNases from this natural population (S_8, S_{12}, S_{13}, S_{15}, and S_{19}) and the previously identified S_2- and S_3-RNases form a major clade, and the other 10 are grouped with one another in the same cluster with S-RNase(s) from other species of the *Petunia* genus, or with S-RNase(s) from another genus of the Solanaceae. This pattern of trans-species polymorphism, first observed by Ioerger et al. (1990), supports the hypothesis that *S* haplotype polymorphism has an ancient origin and has been maintained by balancing selection.

Different Evolutionary Histories of the 5′ and 3′ Regions of the S-RNase Gene

To ascertain the possibility that intragenic recombination could have generated closely related alleles of the S-RNase gene we compared the seven closely related alleles mentioned above to see whether dif-

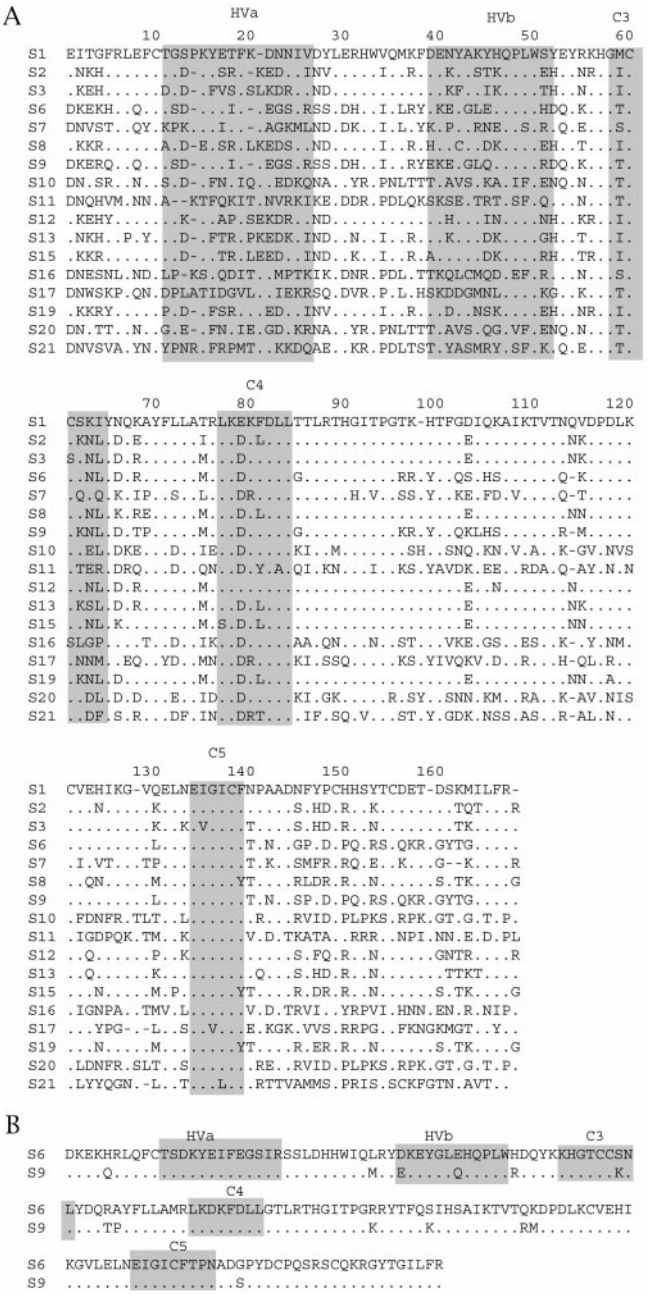

Figure 2. Alignment of deduced amino acid sequences of 17 S-RNases of *P. inflata* (A) and of S_6- and S_9-RNases (B). At each aligned position, amino acids that are identical to the amino acid of the reference sequence, S_1-RNase in A and S_6-RNase in B are indicated with dots, and only those amino acids that are different are labeled. Gaps, indicated with dashes, are introduced to maximize similarity. The two hypervariable regions (HVa and HVb) and three conserved regions (C3, C4, and C5) are shaded. The GenBank accession numbers for these sequences are as follows: S_1, M67990; S_2, AF301533; S_3, M67991; S_6, AF301167; S_7, AF301168; S_8, AF301169; S_9, AF301170; S_{10}, AF301171; S_{11}, AF301172; S_{12}, AF301173; S_{13}, AF301174; S_{15}, AF301175; S_{16}, AF301176; S_{17}, AF301177; S_{19}, AF301178; S_{20}, AF301179; and S_{21}, AF301180.

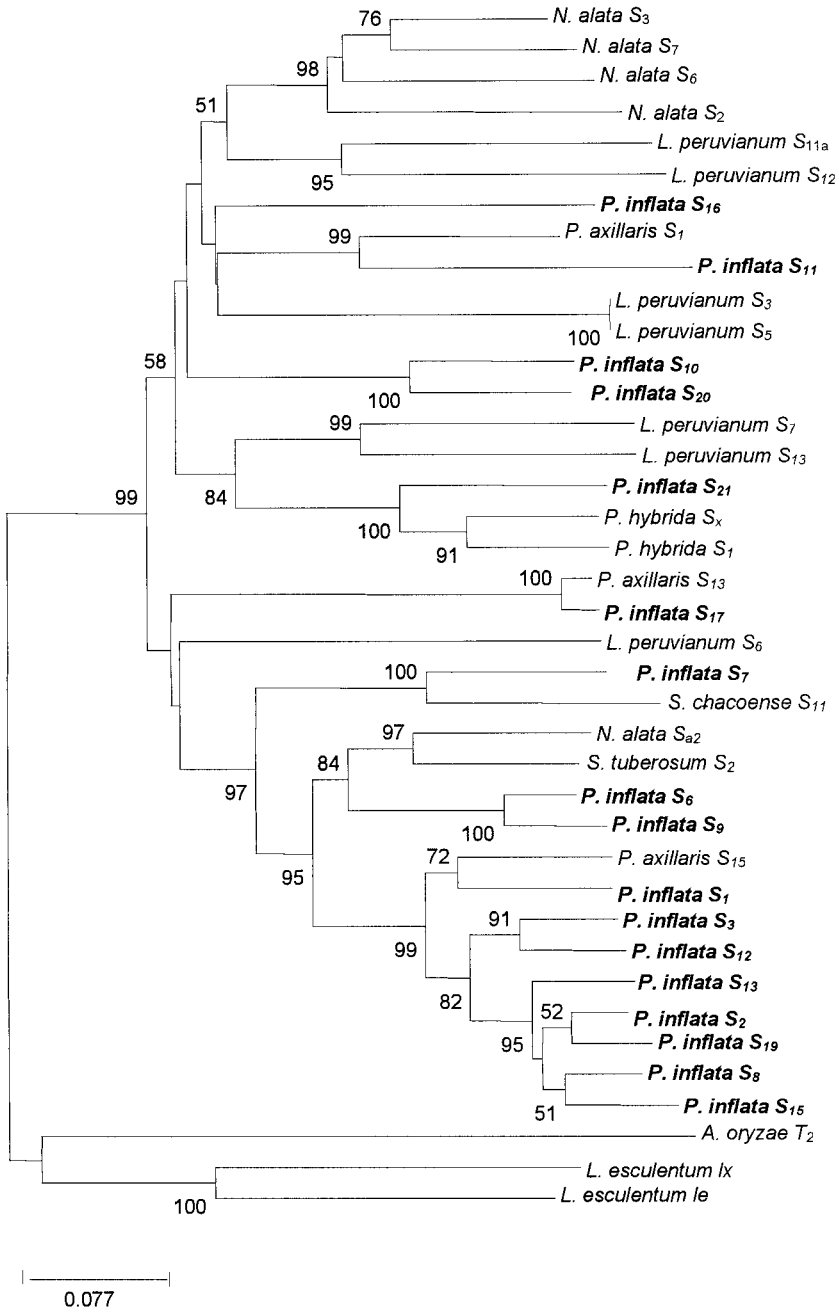

Figure 3. A phylogenetic tree of solanaceous S-RNases. The GenBank accession numbers for 17 S-RNases of *P. inflata* (shown in bold) are listed in Figure 2. The accession numbers for the other S-RNases are as follows: *P. hybrida*, S_1, U07362; S_x, M81685. *Petunia axillaris*, S_1, AF239908; S_{13}, AF239909; S_{15}, AF239910. *L. peruvianum*, S_3, X76065; S_5, S61768; S_6, Z26583; S_7, Z26582; S_{11a}, S65047; S_{12}, D17324; S_{13}, D17325. *Nicotiana alata*, S_2, U08860; S_3, U66427; S_6, U08861; and S_7, U13255; S_{a2}, U45957. *Solanum chacoense*, S_{11}, L36464. *Solanum tuberosum*, S_2, X62727. The tree was constructed by the neighbor-joining method based on the proportion of amino acid differences. It is rooted by RNase T2 from *Aspergillus oryzae* (accession no. X61086) and two S-like RNases of tomato, le (accession no. X79337), and lx (accession no. X79338). Numbers on the internal branches indicate bootstrap percentages that are higher than 50, and the scale below the tree is the measure of branch length.

ferent regions of the S-RNase gene have taken the same evolutionary pathway. If allelic diversity of the S-RNase gene is entirely due to the accumulation of point mutations over evolutionary time, different regions of the gene would be expected to have consistent genealogies because the entire gene would have evolved as a unit. On the other hand, if recombination among different alleles of the S-RNase gene has occurred, it would shuffle polymorphic sequences among alleles, causing different regions of the S-RNase gene to have different evolutionary histories.

As an exploratory tool we plotted p_s (the proportion of synonymous differences per synonymous site) and p_n (the proportion of non-synonymous differences per non-synonymous site) in a sliding window of 30 codons for all possible pairwise comparisons among the aforementioned seven alleles of the S-RNase gene. The plots for the pair of the S_2-RNase gene and the S_{19}-RNase gene (Fig. 4A) show that they are more similar to each other at synonymous and non-synonymous sites in the 5′ region than in the 3′ region of the gene. The plots for the pair of the S_8-RNase gene and the S_{19}-RNase gene (Fig. 4B) show that they are more similar to each other at synonymous and non-synonymous sites in the 3′ region than in the 5′ region of the gene. The maximum chi-square

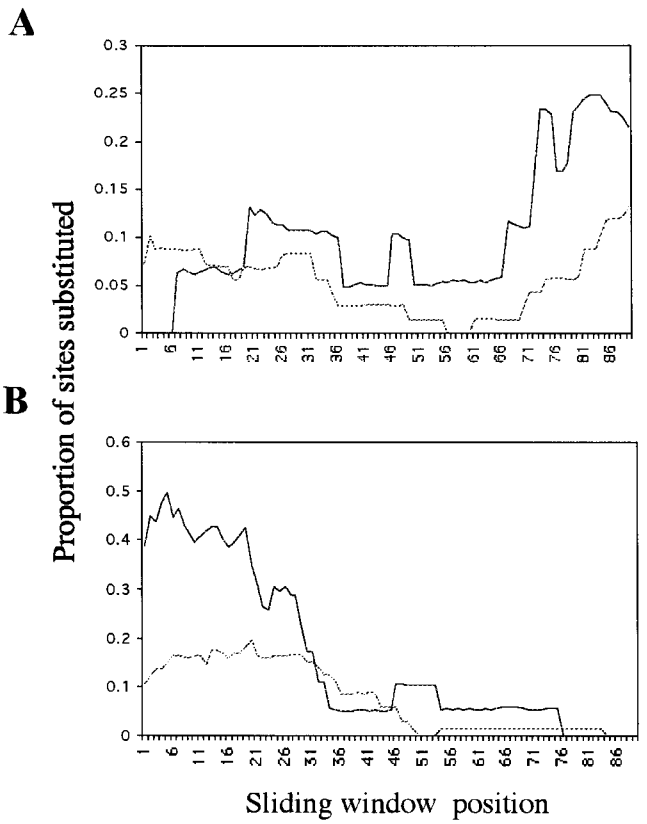

Figure 4. Plots of p_s and p_n for pairwise comparison between S_2 and S_{19} alleles of the S-RNase gene (A) and between S_8 and S_{19} alleles of the S-RNase gene (B). Solid lines are plots of p_s and dotted lines are plots of p_n. Both plots were obtained using a sliding window of 30 codons.

method gave statistically significant evidence ($P = 0.007$) of a recombination event giving rise to different evolutionary histories in the 5′ region (codons nos. 1–68; refer to Fig. 2A for the numbering of amino acids) and the 3′ region (codons nos. 69–168) of the S-RNase gene.

However, a significant result of the maximum chi-square method alone is not definitive evidence of recombination because several alternative hypotheses might explain the observation of greater similarity between alleles in one region of a gene than in another region. To decide between the hypothesis of recombination and alternative explanations we constructed separate phylogenetic trees for the 5′ (codons nos. 1–68) and 3′ (codons nos. 69–168) regions of the S_2-, S_8-, S_{13}-, and S_{19}-RNase genes. For each region we constructed two separate trees on the basis of p_s and p_n. Recombination is expected to produce similarity at synonymous and non-synonymous sites, whereas certain other factors can yield similarity at synonymous sites alone or at non-synonymous sites alone. For example, a shared G + C content bias can produce an unexpectedly high similarity at synonymous sites, whereas a shared functional constraint can produce an unexpectedly high similarity at non-synonymous sites (Hughes, 1991). The phylogenetic trees show identical topologies with high bootstrap values (Fig. 5, A and B), strongly supporting the conclusion drawn from the results shown in Figure 4, A and B that the 5′ region of the S_{19}-RNase gene is closely related to the S_2-RNase gene, but the 3′ region is closely related to the S_8-RNase gene.

Another group of alleles in which we detected the existence of recombination are S_3, S_{12}, and S_{15}. The maximum chi-square test supported the hypothesis of recombination ($P = 0.048$) between the 5′ region (codon nos. 1–62) and the 3′ region (codons nos. 63–168) of these alleles. Phylogenetic analysis based on p_s and p_n showed that the 5′ region of the S_3-RNase gene is closely related to that of the S_{12}-RNase gene, whereas the 3′ region is closely related to that of the S_{15}-RNase gene (Fig. 5, C and D).

For both sets of alleles, because the regions examined are large (encoding at least 62 amino acids) and the bootstrap values are high, the observed topological differences are unlikely to be due to stochastic error. Thus, our results suggest that intragenic recombination has played a role in the generation of new allelic specificities of the S-RNase gene.

DISCUSSION

Use of RT-PCR in Cloning cDNAs for S-RNases from Natural Populations

Although cDNAs for a large number of S-RNases have been cloned and sequenced from a number of solanaceous species, only a few studies so far have used plants derived from natural populations as the source of *S* haplotypes. Richman et al. (1995, 1996) were the first to use RT-PCR to clone cDNAs for S-RNases from natural populations of solanaceous species; they determined the number of *S* haplotypes and the frequency of each *S* haplotype in individual populations. In these studies *S* genotypes of plants used were not determined genetically, and because the PCR primers were designed based on the C2 and C3 (or C4) regions of S-RNases, the cDNAs obtained encode approximately 60 or 120 amino acids out of a total of approximately 200 amino acids for full-size S-RNases.

In our study of a natural population of *P. inflata* identified in Argentina we first carried out genetic crosses to identify the *S* genotype of each of the plants used and then generated progeny plants homozygous for each of the *S* haplotypes identified. Using *S* homozygotes for RT-PCR has allowed unambiguous assignment of the resulting cDNAs to a particular *S* allele of the S-RNase gene. To obtain longer cDNAs for sequence comparison we chose a pair of PCR primers based on the sequence of the C2 region and the poly(A)$^+$ tail; in this way, the cDNAs obtained lack only approximately 100 bp of the sequence upstream from the C2 region, and the de-

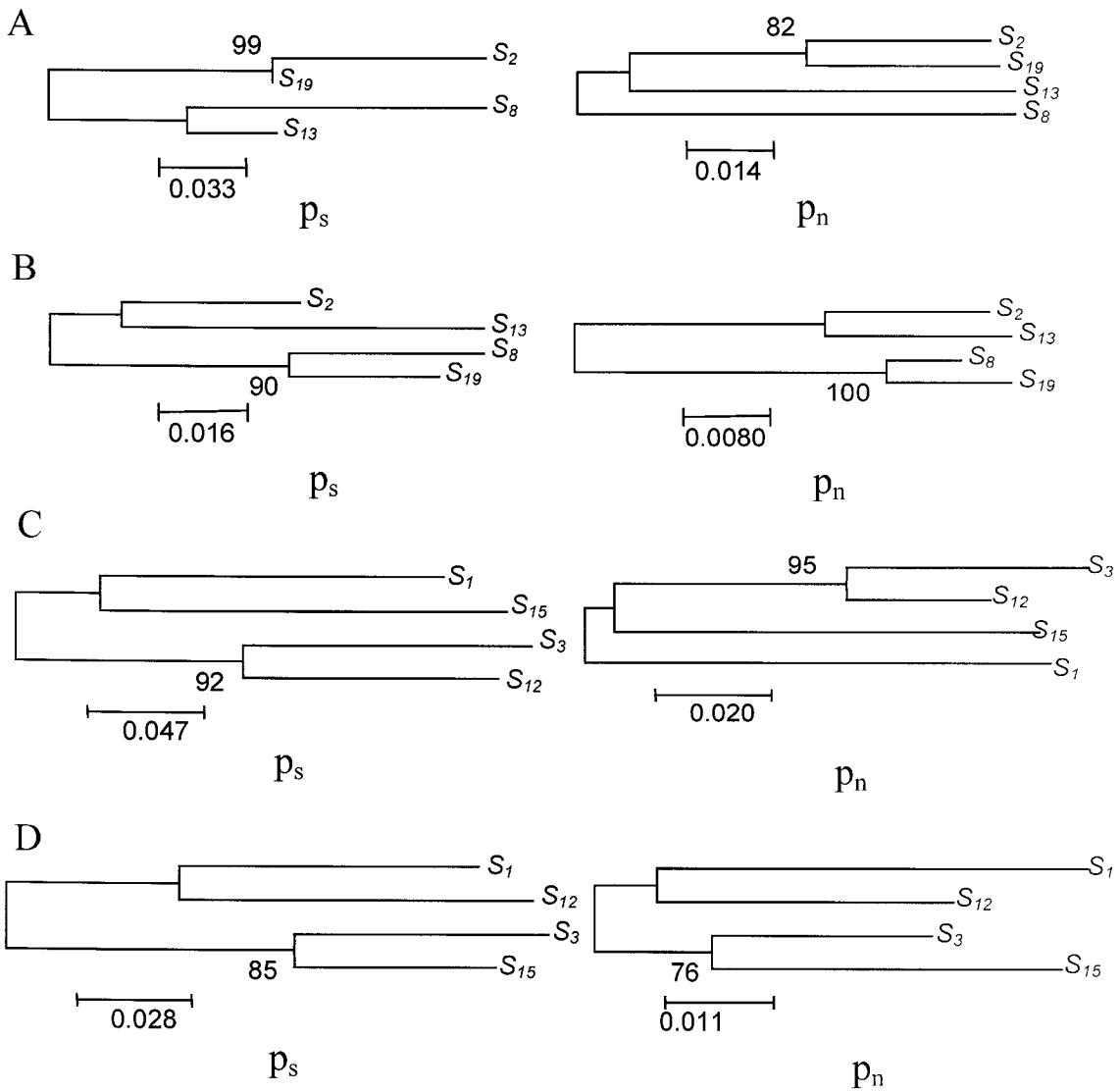

Figure 5. Neighbor-joining trees for the 5′ and 3′ regions of two sets of alleles of the S-RNase gene. For each set of alleles, separate trees were constructed based on the p_s and p_n. Numbers on the branches are bootstrap percentages, and the scale below each tree is the measure of branch length for that tree. For S_2, S_8, S_{13}, and S_{19} alleles, the 5′ region (A) spans amino acid numbers 1 to 68, and the 3′ region (B) spans the rest of the protein (amino acid nos. 69–168). For S_1, S_3, S_{12}, and S_{15} alleles, the 5′ region (C) spans amino acid numbers 1 to 62, and the 3′ region (D) spans the rest of the protein (amino acid nos. 63–168). Refer to Figure 2 for the numbering of amino acids.

duced amino acid sequences downstream from the C2 region account for approximately 80% of the full-size S-RNases. We also used genomic-blotting analysis to examine the cDNAs obtained by RT-PCR to see whether each revealed an *S* haplotype specific RFLP. Subsequent sequencing of all the cDNAs showed that those cDNAs that revealed this pattern encode S-RNases and those that did not encode proteins unrelated to S-RNases.

Genetically Identical *S* Haplotypes Identified from Different Natural Populations

When studying natural populations of SI species, questions can be asked as to whether genetically identical *S* haplotypes are contained in different populations and if so, whether the deduced amino acid sequences of their corresponding *S* locus genes are identical. If differences are found, those amino acids that are different can be eliminated from consideration as playing a role in the determinant of *S* allelic specificity.

Prior to this study only three *S* haplotypes, S_1, S_2, and S_3, had been identified from *P. inflata*. Although the geographical origin of the plants from which these haplotypes were identified is not known, it is unlikely to be the same as the natural population studied here. It is thus interesting to find that one of the haplotypes, S_1, was also present in this natural

population. In fact, no sequence differences were found between these two S_1 alleles of the S-RNase gene in the coding region or the 3′-non-coding region. Preliminary results suggest that the sequence conservation in the *S* locus of these two S_1 haplotypes likely extends beyond the S-RNase gene, because no RFLP was observed when the genomic DNA digests from these two S_1 haplotypes were hybridized with the cDNA for S_1-RNase. However, the conservation appears to break down in other parts of the *S* locus because RFLP was revealed when cDNAs for two *S*-linked pollen-expressed genes were used as hybridization probes. These findings suggest that these two S_1 haplotypes may have evolved independently for some time and that the S-RNase gene may be under strict functional constraint to maintain allelic identity. Perfect amino acid sequence conservation has also been found between pistil S proteins of two genetically identical *S* haplotypes identified in a British and a Spanish population of *Papaver rhoeas*, a species that possesses a different type of GSI from solanaceous species (Walker et al., 1996).

Role of Two Hypervariable Regions in *S* Specificity

Among all the solanaceous S-RNases that have been reported so far, the pair of S_6- and S_9-RNases identified in this work are most similar to each other in the two hypervariable regions; they have an identical sequence in HVa and only differ by two amino acids in HVb. These two regions have recently been suggested to be the sole determinant of the *S* allele specificity of S-RNases (Matton et al., 1997) based on the study of hybrid S-RNases between S_{11}- and S_{13}-RNases of *S. chacoense*. S_{11}- and S_{13}-RNases differ by 10 amino acids, three in HVa, and one in HVb, and the *S* allele specificity of the S_{11}-RNase was found to be converted to that of the S_{13}-RNase simply by swapping the HVa and HVb regions. However, as pointed out by Verica et al. (1998), these results do not necessarily mean that amino acids outside the two hypervariable regions are not involved in *S* allele specificity, because a domain-swapping experiment can only address the role of the regions swapped, but not those that are not swapped. It remains to be determined whether the two amino acids that are different between S_6- and S_9-RNases of *P. inflata* are sufficient for distinguishing the *S* allele specificity of one S-RNase from the other; however, considering that both differences are conservative replacements, this does not seem likely.

Intragenic Recombination As a Mechanism of Generation of Allelic Diversity

The closely related alleles of the S-RNase gene that we identified in a natural population of *P. inflata* have allowed us to obtain the first evidence that apart from accumulation of point mutations, allelic diversity of the S-RNase gene has also been generated by intragenic recombination. This finding is quite interesting, considering that recombination has previously been thought to be absent at the *S* locus to maintain tight linkage between the S-RNase gene and the pollen *S* gene.

One can envision that point mutations would first produce highly divergent polymorphic sequence motifs, and then intragenic recombination would shuffle these polymorphic sequence motifs into different allelic combinations. If accumulation of point mutations were solely responsible for allelic diversity of the S-RNase gene, favorable mutations would have to occur consecutively in the same allelic lineage. However, the additional involvement of intragenic recombination could bring together mutations that have occurred in different allelic lineages into new allelic forms and thereby speed up the process of allelic diversity generation. This evolutionary process is very similar to the diversification of alleles of the major histocompatibility (MHC) loci, which are also involved in self/nonself recognition (Hughes et al., 1993; Hughes and Yeager, 1998). Because balancing selection has maintained polymorphism for a long time at the MHC loci and the *S* locus, substantial allelic diversity is available for the process of recombination. Note that although there is evidence of recombination among alleles at MHC loci of mammals, there are certain MHC loci at which recombinants are rare (Hughes and Yeager, 1998). Thus, the extent to which recombination has contributed to the polymorphism of the S-RNase gene needs further investigation. Considering the potential deleterious effect of recombination on the SI system, we think recombination does occur, but at a low rate. Deleterious recombinants most likely have been eliminated from the population by natural selection. In the case of our observed recombination events that led to the generation of the S_3 and S_{19} alleles of the S-RNase gene, presumably they had selectively neutral or even advantageous effects, thus allowing plants harboring the recombinant alleles to remain in the population.

If recombination and/or point mutations result in a new *S* allele specificity of the S-RNase gene, there must be complementary changes in the pollen *S* gene for an SI plant to remain SI while acquiring a new haplotype specificity. How the pistil and pollen genes "communicate" with each other during the mutational process is one of the most perplexing questions in the study of SI. A solution has recently been proposed based on the finding of a hybrid S-RNase that possesses dual *S* allele specificity (Matton et al., 1999; see discussion in McCubbin and Kao, 2000). According to this model, mutations in the S-RNase gene could result in a protein that retains the original specificity, as well as acquires a new specificity. Subsequent mutations in the pollen *S* gene could render its protein product able to recog-

nize the amino acids of the mutated S-RNase that specify the new specificity, but not those that specify the original specificity. Further mutation in the S-RNase gene with dual specificity would result in the loss of the original specificity. This process would preserve SI during the entire process of the generation of a new *S* haplotype specificity. However, since there have never been any reports of plants in natural populations whose pistils or pollen possess dual specificity, the validity of this model remains to be determined.

MATERIALS AND METHODS

Plant Material

The natural population of *Petunia inflata* R. E. Fr. (*Petunia integrifolia* [Hook.] Schinz & Thell. subsp. *inflata* [R. E. Fr.] Wijsman) selected for this study was designated A9. It was located 1.6 km east of the junction of Ruta Nacional 4 and Ruta Provincial 103, Municipal Candelaria, Provincial Misiones, Argentina (27 23′13″S, 55 34′00″W). The population was comprised of approximately 100 individuals and was established at a gentle load-side slope poorly covered with vegetation. Seed capsules were randomly collected on November 28, 1990, and the seeds were mixed together and stored in a refrigerator. Thirty-six plants were raised from the seeds, and they were coded from A9-1 to A9-36. All of them were SI as judged by their failure to set seed upon selfing and by the arrest of pollen tube growth in the upper one-third segment of the style.

Determination of *S* Genotypes

Pollen from plants homozygous for the previously identified S_1, S_2, and S_3 haplotypes of *P. inflata* (Ai et al., 1990) was used to pollinate plants A9-1 through A9-36. The only pollination that did not set any capsules was between A9-4 and the S_1S_1 homozygote, indicating that one of the *S* haplotypes carried by A9-4 was S_1. The *S* genotype of A9-1 was arbitrarily assigned S_4S_5. The scheme of genetic crossing used to determine the *S* genotypes of plants A9-2 through A9-36 was similar to that described by Tsukamoto et al. (1999).

Isolation of Total RNA from Pistils

Pistils (without ovaries) were collected 1 or 2 d before anthesis from plants homozygous for a particular *S* haplotype, frozen in liquid nitrogen, and were ground to fine powder. Total RNA was extracted from the powder using TRIzol reagent (Life Technologies, Rockville, MD) according to the protocol recommended by the manufacturer. RNA was recovered by precipitation with isopropyl alcohol. After washing with 75% (v/v) ethanol, the RNA pellet was dissolved in water and stored at −80°C.

Isolation of Genomic DNA

For each *S* genotype, 0.2 to 0.3 g of freshly collected young leaves were pulverized in liquid nitrogen, and genomic DNA was extracted from the homogenate with Plant DNAzol Reagent (Life Technologies). Chloroform was then used to remove polysaccharides. Following chloroform extraction, plant debris was removed by centrifugation and DNA was precipitated from the supernatant with ethanol. The resulting DNA pellet was washed with ethanol and dissolved in water.

RT-PCR, Cloning, and Sequencing of cDNAs Encoding S-RNases

To amplify cDNA for each S-RNase, first-strand cDNA was synthesized at 42°C for 60 min in a 20-μL reaction buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, and 10 mM dithiothreitol) that contained 1 to 5 μg of total RNA, 500 ng of oligo $(dT)_{17}$, 0.5 mM dNTPs, and 200 units of SuperScript II reverse transcriptase (Life Technologies). After the reaction, a 0.5-μL aliquot of the single-stranded cDNA was used as a template for PCR amplification in a 20-μL reaction buffer (20 mM Tris-HCl, pH 8.4, 2 mM $MgCl_2$, and 50 mM KCl) containing 2.5 μM upstream primer, 2.5 μM oligo $(dT)_{17}$, 0.1 mM dNTPs, and 2 units of Display *Taq* DNA polymerase (PGC Scientific, Gaithersburg, MD). The upstream primer was a degenerate primer, 5′-ATTCAYGGGCTTTGGCCC-3′ (Y indicating C or T), which was designed based on the sequence of the conserved region C2 of S-RNases. The reaction mixture was denatured at 94°C for 2 min and then cycled as follows: 94°C for 30 s, 40°C for 2 min, and extension at 72°C for 30 s. After the first five cycles the annealing condition was changed to 45°C for 90 s. The reaction was continued for 25 more cycles with an additional 5-min extension after the final cycle. Aliquots of all PCR reaction mixtures were analyzed on 1% (w/v) agarose gels. The remainder of each PCR reaction mixture was purified, ligated into pGEM-T vector (Promega, Madison, WI), and transformed into *Escherichia coli*. Plasmid DNA was prepared from multiple colonies of each transformation using a NucleoSpin Plasmid Miniprep Kit (CLONTECH, Palo Alto, CA). Sequencing was carried out on an ABI 60 sequencer at the Pennsylvania State University Life Sciences Consortium's Nucleic Acid facility. To eliminate potential errors introduced by PCR, two or more clones obtained from independent runs of RT-PCR were sequenced for each *S* haplotype.

DNA-Blot Analysis

Plasmid DNA (100 ng) was digested with *Eco*RI (1 unit) and genomic DNA (5 μg) was digested with *Eco*RI (20 units); the digests were separately electrophoresed on 0.7% (w/v) agarose gels at 25 V overnight. The gels were blotted to Biodyne B nylon membranes (Life Technologies). The DNA fragments were fixed on the membranes by baking at 80°C for 2 h. The membranes were pre-hybridized at 65°C for 1 h in a solution containing 10% (w/v) Dextran sulfate, 1 M NaCl, and 1% (w/v) SDS. Radiolabeled DNA probes were prepared by using an RTS RadPrime DNA Labeling kit (Life Technologies). Hybridization was performed at 65°C overnight in the pre-hybridization solution plus ra-

diolabeled probe. The membranes were washed twice in 0.1× SSC, 0.1% (w/v) SDS for 30 min each (high stringency) or in 2× SSC, 0.1% (w/v) SDS for 30 min each (low stringency), and exposed to phosphoscreen for 8 h. The screen was scanned by a phosphorimager (Molecular Dynamics, Sunnyvale, CA).

Statistical Methods

Nucleotide and deduced amino acid sequences of cDNAs for S-RNases were aligned using CLUSTAL X (Thompson et al., 1997). The alignment of nucleotide sequences was further modified according to the corresponding amino acid sequence alignment. Phylogenetic trees were constructed by the neighbor-joining method (Saitou and Nei, 1987) using the Molecular Evolution Genetic Analysis package (Kumar et al., 1993). The method of Nei and Gojobori (1986) was used to estimate the p_s and the p_n. The reliability of clustering patterns of the phylogenetic trees was tested by bootstrapping (Felsenstein, 1985). A test for the occurrence of patterns of polymorphic sites suggestive of recombination was conducted using the MAXCHI program (McGraw et al., 1999), which implements a version of Smith's (1992) maximum chi-square method.

ACKNOWLEDGMENTS

We thank Andrew G. McCubbin for helpful discussions and critical reading of the manuscript and Jiong Wang for helping to maintain plants in the greenhouse and collecting tissues.

Received September 8, 2000; returned for revision October 11, 2000; accepted November 11, 2000.

LITERATURE CITED

Ai Y, Singh A, Coleman CE, Ioerger TR, Kheyr-Pour A, Kao TH (1990) Self-incompatibility in *Petunia inflata*: isolation and characterization of cDNAs encoding three *S* allele-associated proteins. Sex Plant Reprod **3:** 130–138

Awadalla P, Charlesworth D (1999) Recombination and selection at *Brassica* self-incompatibility loci. Genetics **152:** 413–425

Bernatzky R (1993) Genetic mapping and protein product diversity of the self-incompatibility locus in wild tomato (*Lycopersicon peruvianum*). Biochem Genet **31:** 173–184

Boyes DC, Nasrallah JB (1993) Physical linkage of the *SLG* and *SRK* genes at the self-incompatibility locus of *Brassica oleracea*. Mol Gen Genet **236:** 369–373

Boyes DC, Nasrallah ME, Vrebalov J, Nasrallah HB (1997) The self-incompatibility (*S*) haplotypes of *Brassica* contain highly divergent and rearranged sequences of ancient origin. Plant Cell **9:** 237–247

Casselman AL, Vrebalov J, Conner JA, Singhal A, Giovannoni J, Nasrallah ME, Nasrallah JB (2000) Determining the physical limits of the *Brassica S* locus by recombinational analysis. Plant Cell **12:** 23–33

Clark AG, Kao TH (1991) Excess non-synonymous substitution at shared polymorphic sites among self-incompatibility alleles of Solanaceae. Proc Natl Acad Sci USA **88:** 9823–9827

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

de Nettancourt D (1977) Incompatibility in Angiosperms. Springer-Verlag, New York

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

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791

Hughes AL (1991) Independent gene duplications, not concerted evolution, explain relationships among class I MHC genes of murine rodents. Immunogenetics **34**: 367–373

Hughes AL, Hughes MK, Watkins DI (1993) Contrasting roles of interallelic re-combination at the *HLA-A* and *HLA-B* loci. Genetics **133:** 669–680

Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet **32**: 415–435

Ioerger TR, Clark AG, Kao TH (1990) Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc Natl Acad Sci USA **87:** 9732–9735

Ioerger TR, Gohlke JR, Xu B, Kao TH (1991) Primary structural features of the self-incompatibility protein in Solanaceae. Sex Plant Reprod **4:** 81–87

Kaufmann H, Salamini F, Thompson RD (1991) Sequence variability and gene structure at the self-incompatibility locus of *Solanum tuberosum*. Mol Gen Genet **226:** 457–466

Kheyr-Pour A, Pernes J (1986) A new *S*-allele and specific S-proteins associated with two *S*-alleles in *Nicotiana alata*. *In* DL Mulcahy, GB Mulcahy, E Ottaviano, eds, Biotechnology and Ecology of Pollen. Springer-Verlag, Berlin, pp 191–196

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular Evolutionary Genetics Analysis, Version 1.02. The Pennsylvania State University, University Park, PA

Kusaba M, Nishio T, Satta Y, Hinata K, Ockendon D (1997) Striking sequence similarity in inter-and intraspecific comparisons of class I *SLG* alleles from *Brassica oleracea* and *Brassica campestris*: implications for the evolution and recognition mechanism. Proc Natl Acad Sci USA **94:** 7673–7678

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

Matton DP, Luu DT, Xike Q, Laublin G, O'Brien M, Maes O, Morse D, Cappadocia M (1999) Production of an S-RNase with dual specificity suggests a novel hypothesis for the generation of new *S* alleles. Plant Cell **11:** 2087–2097

Matton DP, Mases O, Laublin G, Xike Q, Bertrand C, Morse D, Cappadocia M (1997) Hypervariable domains of self-incompatibility RNases mediate allele-specific pollen recognition. Plant Cell **9:** 1757–1766

Matton DP, Mau SL, Okamoto S, Clarke AE, Newbigin E (1995) The *S*-locus of *Nicotiana alata*: genomic organiza-

tion and sequence analysis of two *S*-RNase alleles. Plant Mol Biol **28:** 847–858

McCubbin AG, Kao TH (2000) Molecular recognition and response in pollen and pistil interactions. Annu Rev Cell Dev Biol **16:** 333–364

McCubbin AG, Wang X, Kao TH (2000) Identification of self-incompatibility (*S*-) locus linked pollen cDNA markers in *Petunia inflata*. Genome **43:** 619–627

McGraw ER, Selander RK, Whittam TS (1999) Molecular evolution and mosaic structure of α, β, and γ intimins of pathogenic *E. coli*. Mol Biol Evol **16:** 12–22

Murfett J, Atherton TL, Mou B, Gasser CS, McClure BA (1994) S-RNase expressed in transgenic *Nicotiana* causes *S*-allele-specific pollen rejection. Nature **367:** 563–566

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

Richman AD, Kao TH, Schaeffer SW, Uyenoyama MK (1995) *S*-allele sequence diversity in natural populations of *Solanum carolinense* (Horsenettle). Heredity **75:** 405–415

Richman AD, Uyenoyama MK, Kohn JR (1996) *S*-allele diversity in a natural population of *Physalis crassifolia* (Solanaceae) (ground cherry) assessed by RT-PCR. Heredity **76:** 497–505

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:** 406–425

Schopfer CR, Nasrallah ME, Nasrallah JB (1999) The male determinant of self-incompatibility in *Brassica*. Science **266:** 1697–1700

Smith JM (1992) Analyzing the mosaic structure of genes. J Mol Evol **34:** 126–129

Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K (2000) The *S* receptor kinase determines self-incompatibility in *Brassica stigma*. Nature **403:** 913–916

Takayama S, Shiba H, Iwano M, Shimosato H, Che F-S, Kai N, Watanabe M, Suzuki G, Hinata K, Isogai A (2000) The pollen determinant of self-incompatibility in *Brassica campestris*. Proc Natl Acad Sci USA **97:** 1920–1925

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **25:** 4876–4882

Tsai D-S, Lee H-S, Post LC, Kreiling KM, Kao TH (1992) Sequence of an S-protein of *Lycopersicon peruvianum* and comparison with other solanaceous S-proteins. Sex Plant Reprod **5:** 256–263

Tsukamoto T, Ando T, Kokubun H, Watanabe H, Masada M, Zhu X, Marchesi E, Kao TH (1999) Breakdown of self-incompatibility in a natural population of *Petunia axillaris* (Solanaceae) in Uruguay containing both self-incompatible and self-compatible plants. Sex Plant Reprod **12:** 6–13

Verica JA, McCubbin AG, Kao TH (1998) Are the hypervariable regions of S RNases sufficient for allele-specific recognition of pollen? Plant Cell **10:** 314–316

Walker EA, Ride JP, Kurup S, Franklin-Tong TE, Lawrence MJ, Franklin FCH (1996) Molecular analysis of two functional homologs of the S3 allele of the *Papaver rhoeas* self-incompatibility gene isolated from different populations. Plant Mol Biol **30**: 983–994