The Plant Cell, Vol. 6, 1923–1932, December 1994 

# Cloning a Putative Self-Incompatibility Gene from the Pollen of the Grass *Phalaris coerulescens*

**Xinmin Li,[a] Jan Nield,[a] David Hayman,[b] and Peter Langridge [a,1]**

[a] Centre for Cereal Biotechnology, Waite Institute, University of Adelaide, South Australia, 5064, Australia
[b] Department of Genetics, University of Adelaide, Adelaide, South Australia, 5005, Australia

**In *Phalaris coerulescens*, gametophytic self-incompatibility is controlled by two unlinked genes: *S* and *Z*. A probable *S* gene has now been isolated and sequenced. This represents a novel self-incompatibility gene isolated from pollen in the multilocus system of a monocotyledonous plant. The gene is ~3 kb long, split by five introns, and exclusively expressed in the mature pollen. The deduced amino acid sequences from the $S_1$, $S_2$, and part of the $S_4$ alleles showed that the protein has a variable N terminus and a conserved C terminus. The sequence of a complete mutant at the *S* locus indicated that mutations in the conserved C terminus, a thioredoxin-like region, led to loss of function. We propose that the gene has two distinct sections, a variable N terminus determining allele specificity and a conserved C terminus with the catalytic function. The gene structure and its deduced protein sequences strongly suggest that this monocotyledon has developed a self-incompatibility system entirely different from those operating in the dicotyledons. The possible interactions between *S* and *Z* genes in both pollen and stigma are discussed.**

## INTRODUCTION

Self-incompatibility (SI) is a widespread natural phenomenon in the plant kingdom. Plants with this genetic device are able to recognize and reject "self"-pollen, consequently preventing "self"-fertilization. This highly specific recognition mechanism has maintained genetic diversity and is therefore believed to have played a vital role in the evolutionary success of the angiosperms. Homomorphic SI falls into two major groups, gametophytic and sporophytic. In the former, the SI phenotype of pollen is determined by the haploid *S* genotype of the pollen grain itself, whereas in the latter, the diploid *S* genotype of the pollen-producing plant determines pollen phenotype. In both systems, SI is controlled by a single gene (*S*) or two genes (*S* and *Z*) with multiple alleles. To date, the *S* genes and their products have been intensively studied in three families: Solanaceae, Brassicaceae, and Papaveraceae (for recent reviews, see Nasrallah et al., 1991; Franklin-Tong and Franklin, 1992; Thompson and Kirch, 1992; Hinata et al., 1993; Nasrallah and Nasrallah, 1993; Newbigin et al., 1993; Charlesworth, 1994; Matton et al., 1994). The *S* genes of the Solanaceae encode ribonucleases (McClure et al., 1989, 1990; Clark et al., 1990; Ioerger et al., 1991; Singh et al., 1991), whereas the *Papaver* *S* gene encodes a small glycoprotein without RNase activity (Franklin-Tong et al., 1991; Foote et al., 1994). In *Brassica*, the *S* locus has a complex organization (Nasrallah et al., 1988; Stein et al., 1991) in which at least two tightly linked multiallelic genes, *S* locus glycoprotein (*SLG*) and *S* receptor kinase (*SRK*), are involved. RNase activity was not detected in the products of these genes. Although S proteins from these families share some common features, most of them are glycoproteins secreted from pistils, and they have totally unrelated sequences. It has been suggested that the SI systems in these families arose independently (Matton et al., 1994).

Despite these major developments in the analysis of the *S* genes in the pistil, limited progress has been made in the identification of pollen genes. Dodds et al. (1993) reported that the *S* RNase gene of *Nicotiana alata* is expressed at low levels in developing pollen, and cDNAs for the $S_2$ RNase have been cloned from an anther-derived library. A similar result has recently been reported in *Petunia hybrida* (Clark and Sims, 1994). In *Brassica*, mRNA homologous to the stigma *S* locus gene has been identified by polymerase chain reaction (PCR) in anthers during early microsporogenesis (Guilluy et al., 1991). However, there is no evidence to suggest that these transcripts expressed at low levels in both gametophytic and sporophytic systems represent the "*S*" genes in the pollen.

Gametophytic SI in *Phalaris coerulescens* (Poaceae), unlike other well-studied SI systems, is under the control of at least two unlinked multiallelic genes, *S* and *Z*. Self-fertilization is prevented when both the *S* and *Z* alleles present in the pollen are matched in the style. After many years of genetic study, we have collected a large number of *Phalaris* plants with well-characterized *S* and *Z* genotypes (Hayman, 1956), and we have identified several self-compatible mutants, including pollen-only mutants at both the *S* and *Z* locus and a complete mutant at the *S* locus (Hayman and Richter, 1992). Here, we report the identification and characterization of the probable *S* gene from pollen of the grass *P. coerulescens*.

[1] To whom correspondence should be addressed.

## RESULTS

### Restriction Fragments Cosegregating with *S* Genotypes

The strategy adopted to isolate the *S* and *Z* genes from *Phalaris* was to look for pollen-specific cDNA clones that identified restriction fragment length polymorphisms (RFLPs) cosegregating with the *S* and/or *Z* genotype. An extensive collection of plants with well-defined self-incompatibility genotypes was available for this work (Hayman and Richter, 1992). A clone library of 50,000 was prepared from RNA isolated from mature pollen of a plant with the genotype $S_{1.2}Z_{1.1}$. A differential screen was made with $^{32}$P-labeled cDNA from pollen of plants with the $S_{1.2}Z_{2.2}$, $S_{F.F}Z_{1.1}$, or $S_{1.2}Z_{1.1}$ genotype ($S_F$ is used to describe the *S* allele in the complete mutant). Clones that showed no or weak hybridization to $S_{1.2}Z_{2.2}$ or $S_{F.F}Z_{1.1}$ compared with $S_{1.2}Z_{1.1}$ were further characterized. A total of 117 clones was selected and further classified into 18 groups by cross-hybridization. At least two probes were randomly selected from each group, and a total of 54 clones were used for RFLP analysis. As shown in Figure 1A, restriction fragments cosegregating with *S* genotypes are identified by a 907-bp cDNA clone, *Bm2*, after BglII digestion of different genomic DNAs from a known *S* genotype population. *Bm2* hybridized to a single fragment from the $S_{1.1}$ and $S_{2.2}$ genomic DNAs at ~5 and 20 kb, respectively, and to both fragments from the $S_{1.2}$ heterozygotes. The remaining 53 probes tested have not shown cosegregation with either the *S* or *Z* genotype.

Following this observation, two experiments were conducted to test the closeness of linkage between the RFLP identified by the *Bm2* and the *S* genes. In the first experiment, the genotypes of $F_1$ plants obtained from the cross $S_{1.2}Z_{1.1}$ × plant 22 were determined by reciprocally crossing to tester plants. The parental plant 22 was genetically unknown but closely related to the Adelaide population (see Methods). Restriction analysis was also performed on the same plants. Results showed that *S* genotypes derived by both methods coincided perfectly (Figure 1B). Crossing analysis showed that three plants contained a new allele, $S_3$, which must have been present in plant 22. Restriction analysis consistently revealed a different pattern for these plants. In the second experiment, a population of 80 plants was produced from the cross $S_{2.2}Z_{1.2}$ × $S_{1.2}Z_{2.2}$ to test for recombination between the *S* genes and the *Bm2* RFLP. In the cross, only $S_1Z_2$ pollen grains are compatible; therefore, progeny will have the $S_{1.2}$ genotype at the *S* locus. If recombination occurs between the *Bm2* and *S* genes, progeny will show the "$S_{2.2}$"-type pattern on DNA gel blots probed with the *Bm2*. As expected, all plants tested in the $F_1$ population

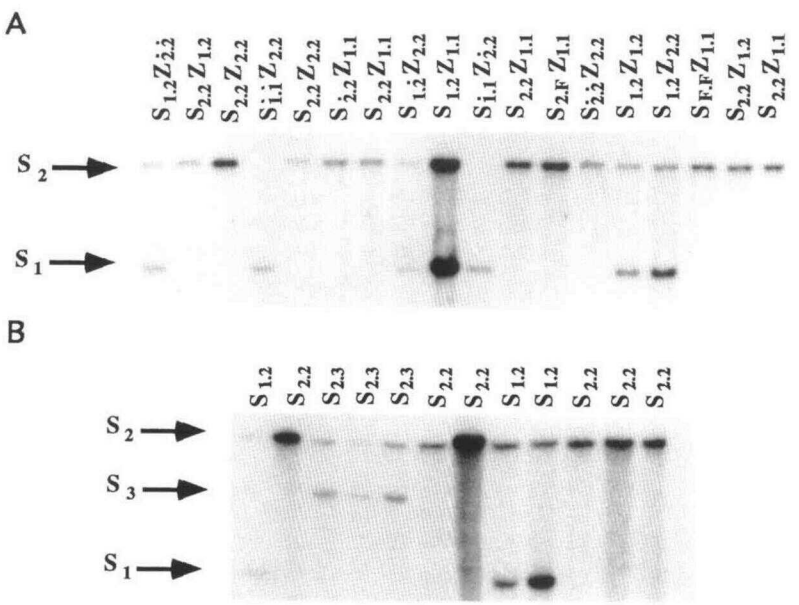


**Figure 1.** Cosegregation of *Bm2* RFLP with the *S* Genotype in *P. coerulescens*.

**(A)** In a collected Adelaide population. The BglII RFLP segregated with the *S* genotype using the clone *Bm2* as a probe. Each lane contains genomic DNA from a different plant and is labeled with the genotype determined by genetic studies. Dots above numbers represent the pollen-only mutants, and F denotes the complete mutant.

**(B)** In an $F_1$ population. The BglII RFLP segregated with the *S* genotype using the clone *Bm2* as a probe. Each lane contains genomic DNA from an $F_1$ plant and is labeled with the genotype determined by reciprocally crossing to tester plants.

were heterozygotes at the *S* locus and no recombinant genotypes were found. So far, 120 plants from the Adelaide *Phalaris* collection and two crossed populations have been tested and showed absolute correlation between the *S* genotype and the *Bm2* RFLP. This strongly suggests that *Bm2* represents either the *S* gene itself or a closely linked gene ($r \leqslant 0.038$ at $P = 0.01$).

### The Gene Is Expressed Only in Mature Pollen

The temporal and spatial expression of the *Bm2* gene was studied by RNA gel blot hybridization. Total RNA was prepared from mature pollen of different *S* genotypes and the *S* complete mutant, as well as from mature stigmas, young leaves, roots, and stems of a plant with the genotype $S_{1.2}Z_{1.2}$. Results showed that *Bm2* strongly hybridized to a 1.1-kb pollen mRNA (Figure 2A) regardless of the *S* genotype. This indicates that the mRNAs of the putative $S_1$ and $S_2$ genes share extensive homology. Strong hybridization to a transcript of similar size in the *S* complete mutant implies that the mutation neither blocked the expression pathway of the *Bm2* gene nor significantly changed the length of the transcripts. There was no detectable hybridization to RNAs from leaves, roots, stems, or stigmas under these hybridization conditions.

The expression of the *Bm2* sequence in the anthers at different stages of development was also investigated. Anther development was divided into five stages, from very young transparent anthers just after meiosis to anthers with mature pollen. Total RNA from different stages was subjected to RNA gel blot analysis with the *Bm2* probe (Figure 2B). The *Bm2* mRNA was absent in stages 1, 2, and 3, present at stage 4 but at a very low level, and abundant at stage 5, when the pollen was fully matured and being shed. Thus, the *Bm2* gene is expressed only in mature pollen. This agrees with the expected pattern of *S* gene expression.

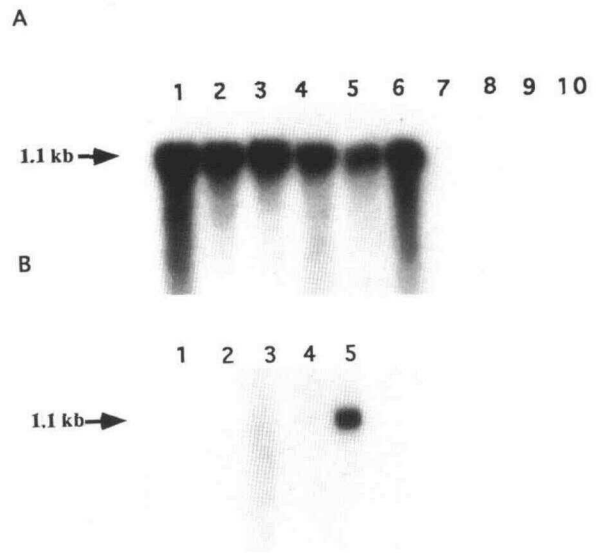


**Figure 2.** Spatial and Temporal Expression of the *S* Gene in *P. coerulescens*.

**(A)** Gel blot analysis of total RNA from different tissues. Lanes 1 to 6, pollen from plants of genotypes $S_{2.2}Z_{2.2}$ (lane 1), $S_{2.2}Z_{1.2}$ (lane 2), $S_{1.2}Z_{1.3}$ (lane 3), $S_{1.2}Z_{2.2}$ (lane 4), $S_{1.1}Z_{1.2}$ (lane 5), and $S_{F.F}Z_{1.1}$ (lane 6); lane 7, stigmas; lane 8, leaves; lane 9, stems; lane 10, roots.
**(B)** Gel blot analysis of total RNA from different stages of anther development. Lane 1, very young transparent anthers; lane 2, young green anthers; lane 3, anthers turning yellow; lane 4, yellow anthers; lane 5, anthers shedding pollen.

### Structure of the Gene and Variation among Alleles

Two genomic DNA libraries were constructed in the λGEM11 phage from plants with $S_{1.1}$ and $S_{2.2}$ genotypes. Screening of $2 \times 10^6$ recombinant phage in the $S_{1.1}$ library with the *Bm2* probe identified four positive plaques, whereas two clones were isolated from the $S_{2.2}$ library after screening $0.8 \times 10^6$ plaques. Two clones from each of the $S_{1.1}$ and $S_{2.2}$ libraries were further analyzed. Positive SacI fragments were subcloned into the pTZ18U vector and, together with cDNA clone *Bm2*, were sequenced using an automatic DNA sequencer. As suggested by the RNA gel blots, the sequencing confirmed that the 907-bp *Bm2* clone was not a full-length cDNA clone. The alignment of the genomic sequence with that of its cDNA showed that *Bm2* was transcribed from the putative $S_2$ allele. The nucleotide sequence of the coding region is interrupted by five introns. The first intron was predicted from mRNA length as determined by RNA gel blot analysis, from the intron–exon junction conserved sequence, from promoter signals, and by DNA gel blot hybridization analysis. Putative TATA and CAAT boxes were found at 57 and 93 bases from the potential translation initiation site, respectively. Another three putative TATA sequences were identified farther upstream (Figure 3).

The putative $S_1$ allele sequence contains 2942 bases from the ATG start codon to the TGA stop codon, three bases more than the putative $S_2$ sequence. The comparison of the two sequences showed that they indeed represent different alleles, although the two sequences are very similar (98% identity). Whereas differences are evenly distributed in the five introns, the majority of variations in the coding regions are concentrated in the second exon, where one base insertion, one base deletion, three consecutive base deletions, and three single substitutions distinguish $S_2$ from $S_1$. An important variation is the single-base insertion at position 415 that leads to a reading frame shift. This is then recovered by the one-base deletion after 57 bases. The genes are identical in exons 3 to 5. In exon 6, two substitutions occur, but these do not cause an amino acid change.

In the predicted amino acid sequences, a significant feature is that all variations are located at the N terminus and there

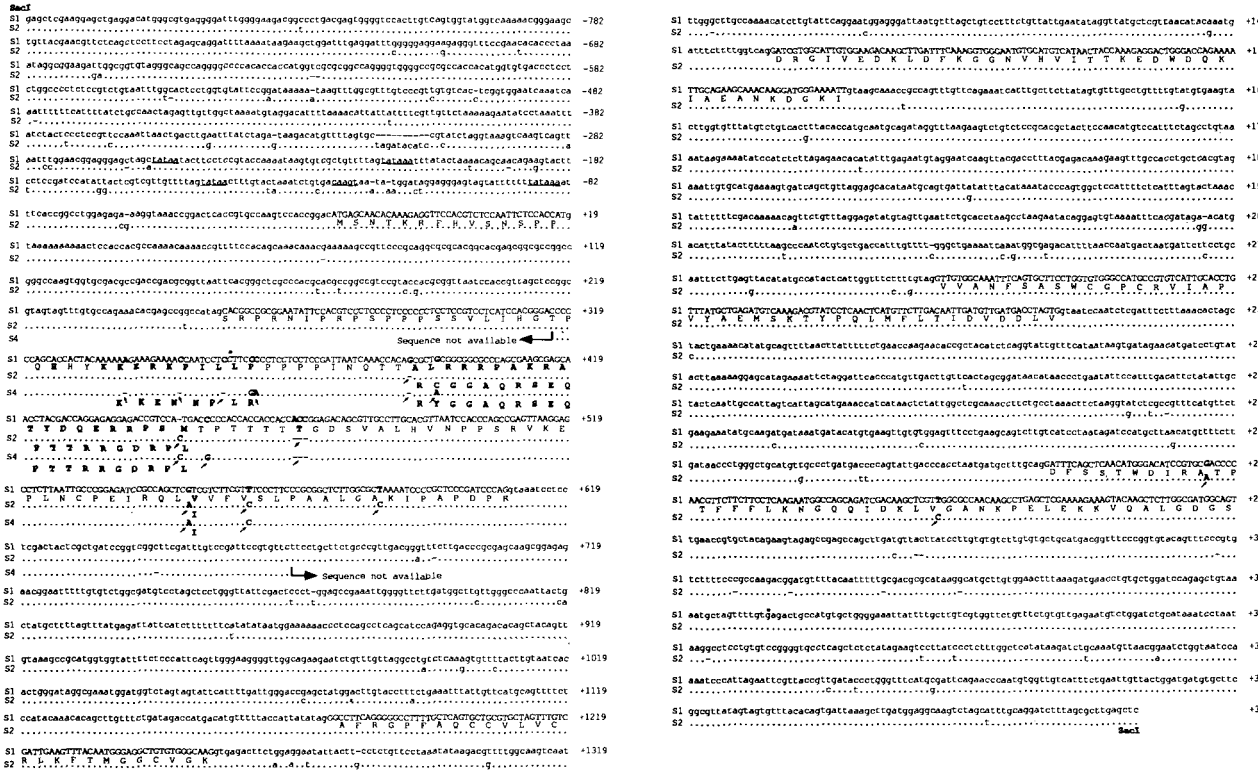


**Figure 3.** The Nucleotide and Deduced Amino Acid Sequences of the $S_1$, $S_2$, and Part of the $S_4$ Genes of *P. coerulescens*.

Coding regions are indicated by uppercase letters; the remaining regions are in lowercase letters. Nucleotides are numbered from the A in the ATG start codon. Putative TATA and CAAT boxes are underlined. The heavy dots above the sequence indicate the limits of the available cDNA. The complete $S_1$ sequence is given. For $S_2$ and $S_4$, only bases that differ are given; otherwise a dot indicates identity. Dashes indicate gaps that were introduced to maximize homology. Nucleotides and amino acids that vary in the coding regions are indicated in boldface, uppercase letters. Only divergent amino acids are shown in the $S_2$ and $S_4$ sequences. Arrows indicate the locations of the base changes and/or insertion/deletions. EMBL accession numbers are X81991 for $S_1$, X81992 for $S_2$, and X81993 for $S_4$.

is absolute sequence conservation at the C terminus. This feature raised the possibility that the gene has two distinct sections: an allele specificity region at the N terminus and a catalytic domain at the C terminus. These assumptions have been tested separately. Further evidence for the role of the N terminus in determining allele specificity was obtained from the putative $S_{4.4}$ sequence data. Primers flanking the second exon of the putative $S_1$ allele were synthesized and used to amplify the appropriate region from $S_{4.4}$ genomic DNA where most of the variations between the putative $S_1$ and $S_2$ sequences are located. The 344-bp fragment generated by PCR was cloned and sequenced (Figure 3). The putative $S_4$ sequence is more similar to $S_2$ than to $S_1$. In addition, the $S_4$ sequence shows a highly variable region of ~50 bases (three single deletions and two substitutions) at the beginning of the available sequence where $S_1$ and $S_2$ are identical. The concentrated nucleotide variations have changed more than 50% of the amino acids in this region, which is only nine amino acids upstream from the codon shift between the putative $S_1$ and $S_2$ sequences.

## Structural Features of the Predicted Amino Acid Sequences

The deduced $S_1$ protein sequence contains 282 amino acids with an $M_r$ of 31,215 (281 amino acids with an $M_r$ of 30,910 for the $S_2$ protein) and is rich in proline (11.3%), valine (7.8%), and lysine (7.5%). Hydropathy analysis of the predicted sequences using a 21–amino acid window showed that the two sequences have almost identical profiles (MacMolly software, Du Pont). After a long hydrophilic region at the N terminus, there are two hydrophobic sections with the length expected for a single membrane–spanning domain (Figure 4). The overall hydrophilic nature of the protein is apparent, suggesting that the protein encoded by the gene is cytoplasmic. However, the

presence of the two hydrophobic sections of approximately the correct size to cross a single membrane leaves the possibility open that the proteins are membrane associated. The prediction of protein conformation with SeqAid II software (Kansas State University, Manhattan, KS) revealed that the proteins are highly structured, with ~60% of their residues involved in secondary structural elements. Differences in the N-terminal sequences between the $S_1$ and $S_2$ proteins would alter protein conformation. The $S_1$ protein sequence is predicted to form a long α-helix around residues 60 to 75, which is absent in the $S_2$ protein. Similarly, the predicted $S_4$ protein has a hydrophilic feature in this region but a configuration different from that of $S_1$ and $S_2$. The α-helix around residues 36 to 41 is missing and replaced by a turn.

A striking observation came from the sequence data base search. The conserved C termini of the proteins share 41% identity with the thioredoxin H proteins (Marty and Meyer, 1991; Rivera-Madrid et al., 1993). Significantly, the active site sequence, WCGPC, and many other amino acids important for thioredoxin structure, such as Phe-27, Ala-29, Trp-31, Pro-40, Asp-61, Pro-76, and Gly-92, are also conserved in the S sequences at the appropriate position relative to the active site (Figure 5). A hydrophobic region around the active site is also present. This has been proposed as a major site of interaction with other proteins (Eklund et al., 1991). Structure analysis showed that all thioredoxins have a similar three-dimensional structure, although the amino acid sequences of thioredoxins from different species show only 27 to 69% identity with *Escherichia coli* thioredoxin (Eklund et al., 1991). Therefore, it appears that the C terminus of the S protein has a tertiary structure similar to that of the thioredoxins.

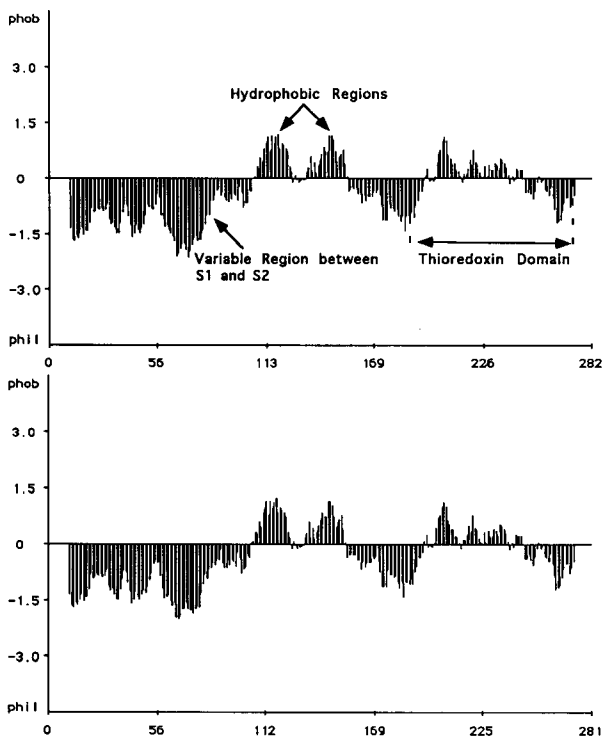


**Figure 4.** Hydropathy Plots of the Deduced Amino Acid Sequences of the $S_1$ and $S_2$ Proteins.

The deduced amino acid sequences were analyzed for hydropathy using MacMolly software with a window of 21 consecutive residues. The upper plot shows the $S_1$ profile, and the lower plot shows the $S_2$ profile. phob, hydrophobic; phil, hydrophilic.

### The *S* Complete Mutant Is Altered at the Conserved Thioredoxin Region

The proposed role of the conserved thioredoxin region in determining gene function was tested through the isolation and sequencing of the *S* complete mutant, in which both pollen and stigma have lost the self-incompatibility response at the *S* locus. The design of the experiment from which this mutant was recovered required that the mutant be derived from either an $S_1$ or $S_2$ allele (Hayman and Richter, 1992). Sequence comparison revealed that the sequence of the mutated allele is the same as the $S_2$ sequence at the variable 5′ end but differs from the $S_2$ and $S_1$ in the thioredoxin region (Figure 6). All changes were located in the sixth exon, where two transitions and four transversions were found when compared with the $S_1$ sequence. These include the two conserved substitutions that distinguish $S_1$ from $S_2$ at this region. These point mutations changed three amino acids: serine to arginine at residue 240, glutamine to glutamic acid at residue 259, and leucine to valine at residue 263. With our knowledge to date of the thioredoxin sequences, it is not possible to relate these mutations to function. However, the serine-to-arginine change may be significant in terms of post-translational modification and could be responsible for loss of the function of the protein.

The *Bm2* gene was also isolated and sequenced from an *S* pollen-only mutant. We could not identify a single base change in the coding region when compared with the $S_2$ sequence (data not shown).

## DISCUSSION

The *Bm2* probe, derived from a cDNA library of mature pollen, has identified an RFLP in BglII-digested *Phalaris* DNA; it cosegregates with the genotype of the *S* locus, which is one of the loci controlling self-incompatibility in *P. coerulescens*. The *Bm2* gene is expressed in mature pollen but not in any other tissue tested. The identification of characteristic N terminus variations among alleles and conserved C terminus variations in the mutant strongly suggests that the cDNA clone *Bm2* represents the *S* gene of *P. coerulescens*.

The cloning strategy assumed that the *S* and *Z* genes would

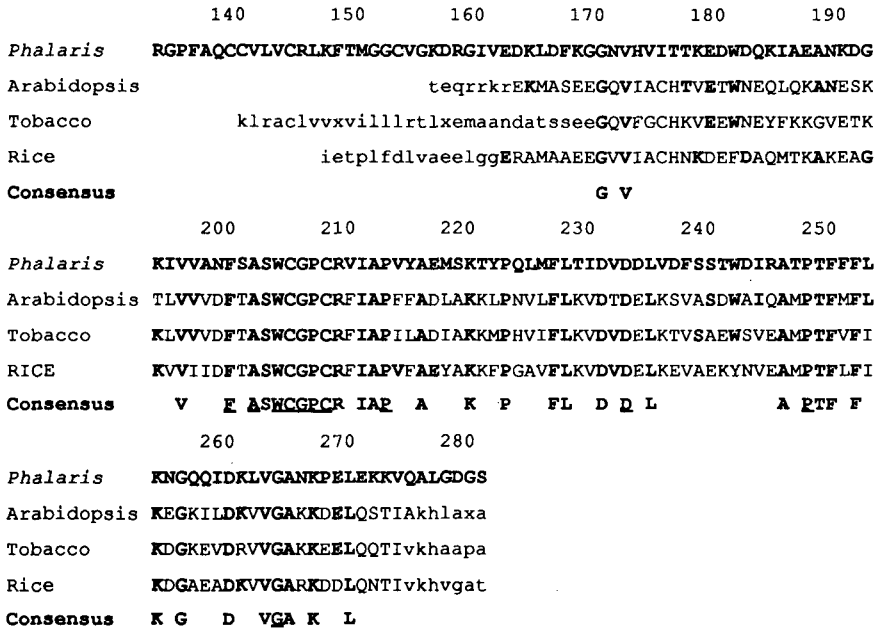


**Figure 5.** Comparison of the N-Terminal Sequences of the Predicted *Phalaris* S Protein and Thioredoxin H Proteins.

Numbers refer to the *Phalaris* sequence. Regions being compared with the *Phalaris* sequence are indicated by uppercase letters, and identity with the *Phalaris* sequence is shown by boldface letters. Conserved functional amino acids are underlined. Lowercase letters indicate that the regions are not being compared. Sequences being compared are Arabidopsis thioredoxin H (ATTHIOARA; Rivera-Madrid et al., 1993), tobacco thioredoxin H (NTTRNA; Marty and Meyer, 1991), and rice thioredoxin H.

be exclusively expressed in mature pollen and that alleles at each locus would be sufficiently divergent in sequence that they would not cross-hybridize. This second assumption was based on the results from the Solanaceae and Brassicaceae (Anderson et al., 1989). We were fortunate to retain *Bm2* in the analysis because more careful study has shown that the *S* alleles are very similar in sequence and readily cross-hybridize. In the predicted amino acid sequences, the $S_2$ protein shares 92% identity with the genetically distinct $S_1$ protein and 98% identity with the product of the *S* complete mutant. This figure is very close to the highest similarity between existing functional SI alleles so far reported in *Solanum* (Saba-El-Leil et al., 1994). The 98% identity between the products of the $S_2$ and the *S* complete mutant is comparable to that reported by Royo et al. (1994), who demonstrated that a mutant *S* allele in *Lycopersicon peruvianum* lost both incompatibility and RNase activity (Kowyama et al., 1994) through a single amino acid change (the loss of a histidine residue at the active site of *S* locus ribonuclease).

It is clear that the origin of new alleles in *Phalaris* is due to the progressive accumulation of point mutations rather than a major reconstruction of existing alleles. These point mutations accumulated only in certain regions. Variation is clustered at the N terminus between the $S_1$ and $S_2$ alleles and at the C terminus in the *S* complete mutant. The experiment to detect self-compatible mutants would have also detected mutations to new functional alleles (Hayman and Richter, 1992). No such mutations were detected in an estimated $10^8$ pollen grains. The sequence data from this study suggested that any such new alleles require changes at a number of sites in the second exon. Theoretically, a mutation to a new allele would enjoy a selective advantage, because it is compatible with all genotypes in the population.

The low sequence divergence among alleles in this species may be related to the complexity of the SI system. Preliminary experiments suggest that the S and Z proteins may form a heterodimer in the pollen (data not shown). This implies that the S pollen protein interacts with the Z or T pollen protein (*T* is a new SI locus identified in studies of mutants; see Hayman and Richter, 1992) before recognition with the stigma component can occur. Obviously, the S pollen protein must contain additional recognition sites in comparison with the S protein of the single locus systems. It is perhaps logical to assume that the *S* gene of the grasses is more delicate than the single locus *S* genes and less tolerant of mutations.

The distribution of variation among the *S* gene sequences highlights two distinct sections, a variable N terminus probably determining allele specificity and a conserved C terminus with the catalytic function. The hydropathy index showed that $S_1$, $S_2$, and part of $S_4$ are hydrophilic at the variable N terminus. This region is expected to be exposed at the surface of the protein accessible for interactions with other SI components, most likely, the S stigma protein. The subsequent recognition will be specified by different protein conformations

in this region, as demonstrated in the prediction of protein structure. Similar hydrophilic features but different secondary structures at the N terminus would be consistent with the requirements for this region to determine allele specificity.

It is not clear which part of the gene determines the pollen specificity. Sequencing the *S* allele from an $S_2$ pollen-only mutant failed to identify amino acid changes when compared with the $S_2$ predicted sequence. One possibility is that transcriptional or translational modifications are involved in the pollen specificity, such as alternative splicing sites, methylation, glycosylation, and phosphorylation (Wehling et al., 1994). Such modifications cannot be identified by genomic DNA sequencing. It is also possible that the lesion in the pollen-only mutants is in a gene closely linked to the *S* locus and also involved in the incompatibility reaction. The interaction between the stigma and pollen proteins probably leads to a series of subsequent reactions that ultimately result in the cessation of pollen growth. A gene encoding one of the proteins involved in these downstream reactions may have been mutated in the pollen-only mutants. Genetic data have shown only that the mutated locus is linked to the *S* gene.

It is interesting that the conserved C terminus shows extensive homology with the thioredoxin H proteins. Furthermore, active site and functionally important residues in the thioredoxins are conserved. It is plausible to assume that the C terminus of the S protein has a three-dimensional structure similar to that of thioredoxins. It will be important to demonstrate the catalytic properties of this region. Thioredoxin is a ubiquitous, small, heat-stable protein that acts in many important biological reactions, including as a hydrogen donor for ribonucleotide reductase, as a substrate for reductive enzymes, as a protein disulfide oxide reductase, as a regulatory factor for enzymes or receptors, and as a subunit of a viral DNA polymerase (Holmgren, 1985). Thus, the presence of the conserved thioredoxin region in the *S* gene opens up several possible mechanisms for the self-incompatibility reaction in grasses. However, the multifunctional properties of the thioredoxin proteins have made it difficult to speculate on their involvement in this multilocus SI system at this stage. Further experiments are needed to clarify this crucial issue.

The localization of the lesion in the *S* complete mutant to the thioredoxin-like region reinforces the proposition that the conserved C terminus is the functional domain of this gene. The replacement of serine by arginine at residue 240 in the *S* complete mutant is probably the most significant mutation of the three changes. Serine contains an aliphatic hydroxyl group and is one of the three most important amino acids involved in protein phosphorylation (phosphorylation results from addition of a phosphate group to the hydroxyl group of serine, tyrosine, or threonine and triggers conformation changes that alter the properties of the proteins). Protein phosphorylation is known to play a vital role in signal transduction in higher plants (Cohen, 1992). Evidence is also accumulating to show that it is actively involved in the SI response. In *Brassica*, the *SRK* gene is a member of the *S* gene family and encodes a functional serine/threonine kinase (Goring and Rothstein, 1992; Stein and Nasrallah, 1993). The requirement of the *SRK* gene in the SI response has been genetically confirmed (Nasrallah et al., 1994). Recently, Wehling et al. (1994) have shown that the phosphorylation of pollen protein plays a key role in SI in

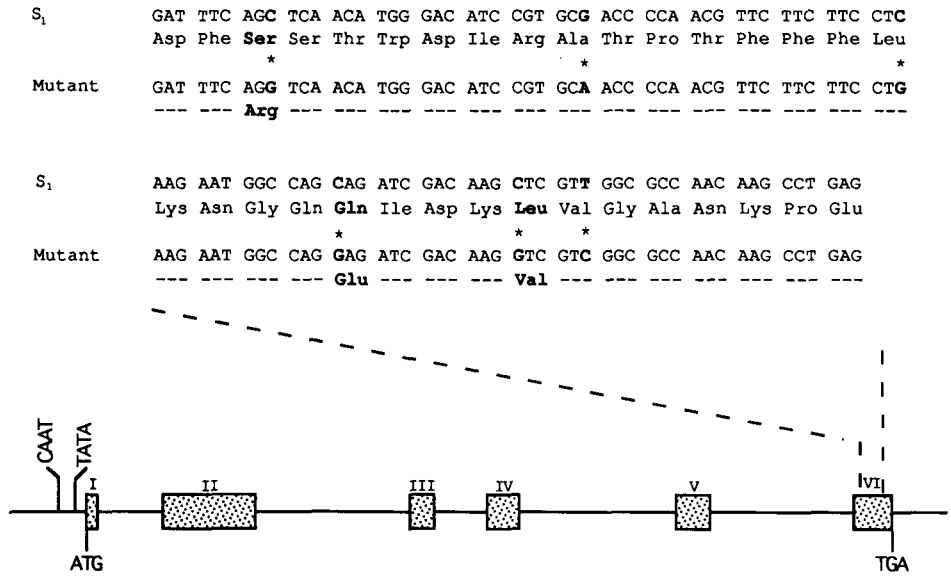


**Figure 6.** Comparison of Part of the Sixth Exon of the *S* Complete Mutant and $S_1$.

Stars indicate different nucleotides. For the mutant, only the amino acids that differ from the $S_1$ sequence are shown. Roman numerals above the stippled boxes indicate the number of exons.

rye. Rye has a gametophytic two-locus SI system similar to that of *P. coerulescens*. It is possible that the serine at residue 240 is involved in phosphorylation, an essential step in the SI pathway of this species. Replacement of the serine by arginine in the *S* complete mutant may abolish this vital post-translational modification and consequently inactivate the SI system. The identification of serine/arginine mutation in the *S* complete mutant further strengthens the suggestion that a protein kinase is involved in SI in grass species (Wehling et al., 1994). If so, it is possible that the product of the *T* locus in this species is a kinase (the *T* gene shows no allelic specificity).

Although we are not able to propose an experimentally supported working model for SI in *Phalaris*, the sequence of the *S* genes has provided some useful information about the action of the self-incompatibility gene products in this species. The S proteins are probably not transferred across the plasma membrane during pollen germination. This would imply that the incompatibility reaction occurs within the germinating pollen. Because the reaction can set in very early, sometimes before the pollen tube penetrates the stigma, it would appear that the product of the stigmatic genes is secreted onto the stigma surface, as is the case with the *S* gene products in the single-locus systems, and is taken up by the pollen upon germination.

The distinction between the genes expressed in the stigma and the pollen is not unique to *Phalaris*. Separate but closely linked genes are probably required, one for pollen and one for stigma. The mutation studies in *Phalaris* have identified several pollen-only mutants at both the *S* and *Z* loci in addition to the complete mutant described here. The lack of expression of the *S* gene in the stigma further supports the two-gene hypothesis (Lewis, 1963).

Because the *Phalaris* system involves at least two loci, it is probable that the *S* and *Z* gene products interact closely, possibly to form a dimer, within the germinating pollen. The stigmatic gene products may be taken up by the pollen as separate proteins or as a dimer. The reaction between the stigma and pollen-derived "dimers" would determine whether or not pollen tube growth will continue. Most probably, lack of recognition allows growth, whereas recognition disrupts cell function sufficiently to stop growth and, ultimately, kills the pollen.

The isolation and sequencing of the different alleles have provided strong evidence that these clones do represent the *S* genes, but it is not clear how they evolved or how the complete mutant arose. However, it is obvious that the SI system in *Phalaris* is fundamentally different from those operating in the single-locus systems of the dicotyledons. Previously isolated *S* genes apart from the *SRK* gene encode an abundant glycoprotein in the pistils, whereas the *Phalaris S* gene encodes an apparently soluble protein with a thioredoxin domain in the pollen. The *S* genes in dicotyledons frequently show extensive allelic diversity, but the alleles isolated here differ in only a small region. Furthermore, *S* genes reported to date have a typical leader sequence, suggesting that the proteins are secreted. However, the predicted S protein of *Phalaris* has no such signal and an overall hydrophilic nature. We cannot rule out the possibility that these differences are only a reflection of the distinction between the SI proteins of the pollen versus the stigma, rather than between the monocots and dicots.

Most grasses with a chromosome number based on 7 have an *S–Z* incompatibility system (Hayman, 1992) and are likely to have conserved linkage groups involving at least the *S* locus (Leach and Hayman, 1987) and probably the *Z* locus as well. Therefore, the results reported in this study of *Phalaris* will also apply to SI systems in economically important plants such as *Secale* and *Hordeum*.

Isolation of the probable *S* gene is the first step in understanding the mechanism of action of the SI genes in grasses. A high priority for future work will lie in elucidating the interactions between the *S* and *Z* gene products in the pollen. The confirmation of dimer formation may lead to the successful isolation of the *Z* gene. The implication of the thioredoxin domain in the SI response is also crucial.

## METHODS

### Plant Material

Self-incompatible lines of *Phalaris coerulescens* homozygous for $S_1$, $S_2$, $S_4$, $Z_1$, and $Z_2$ and heterozygous for $S_{1,2}$, $S_{2,4}$, and $Z_{1,2}$ were derived from a collection in the Department of Genetics, University of Adelaide, Adelaide. Pollen-only mutants and the complete mutant at the *S* locus were also used in this study (Hayman and Richter, 1992). All plants were maintained under glasshouse conditions.

### Isolation of DNA and Gel Blot Analysis

DNA was prepared as described by Guidet et al. (1991). Restriction endonuclease–digested DNA was fractionated on 1% agarose gels and blotted to Hybond $N^+$ membranes (Amersham International). Blots were prehybridized and hybridized at 65°C in a hybridization oven (HYBAID, Middlesex, UK). The membranes were washed in 2 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate), 0.1% SDS at 65°C for 20 min. A final wash was in either 0.2 × SSC, 0.1% SDS or 0.1 × SSC, 0.1% SDS.

### Isolation of RNA and Gel Blot Analysis

Total RNA was extracted from mature pollen, mature stigma, leaves, stems, and roots with RNA extraction buffer, followed by centrifugation in CsCl solution according to the procedure of Sambrook et al. (1989). For RNA gel blots, 10 μg of total RNA isolated from these tissues was loaded and hybridized with the $^{32}$P-labeled *Bm2* probe. rRNA presented in each lane was used as an indication to ensure that an equal amount of RNA had been loaded. RNA was separated on 1.5% agarose denaturing gels and transferred to a Hybond $N^+$ membrane. RNA was fixed to the membrane by UV light. Prehybridization and hybridization were performed at 42°C in 50% formamide.

### Construction of cDNA and Genomic DNA Libraries

Total RNA (50 μg) isolated from mature pollen of plants with the $S_{1.2}Z_{1.1}$ genotype was used to synthesize cDNA using a cDNA synthesis kit (Pharmacia). Following the addition of an EcoRI linker, cDNA was cloned into λgt10 arms, packaged in vitro, and propagated on the *Escherichia coli* C600Hf1 strain.

For the genomic libraries, CsCl-purified DNAs prepared from fresh leaves of genotype $S_{1.1}$, $S_{2.2}$, *S* pollen-only mutant $S_{2.2}$, and *S* complete mutant $S_{F.F}$ were partially digested with Sau3A and fractionated on sucrose gradients. DNA fractions in the range of 9 to 20 kb were ligated into λGEM11 BamHI arms (Promega). The resulting DNA was packed in vitro with package extract (Promega) and plated with *E. coli* KW251.

### Analysis of Recombinant Clones and DNA Sequencing

The genomic DNA libraries were screened with the *Bm2* cDNA clone. Positive plaques were purified, and the DNA was isolated. After SacI digestion of the DNAs, the fragments that hybridized to *Bm2* were subcloned into pTZ18U. Nested deletion libraries from the SacI fragments were generated with the Erase-A-Base system (Promega) according to the manufacturer's instructions. A series of deletion clones and the *Bm2* cDNA clone were sequenced using a 373A DNA sequencer (Applied Biosystems, Foster City, CA). Alignments of the nucleotide sequences were performed with SeqED software (Applied Biosystems).

### Polymerase Chain Reaction Amplifications

Primers flanking the second exon were synthesized (upper primer, 5′-CCCCCAGCACCACTACAAAAA-3′; lower primer, 5′-GAACACGGAATCGGACAAATC-3′) and used to amplify a 344-bp fragment corresponding to the second exon of the $S_{4.4}$ genomic DNA. The standard polymerase chain reaction (PCR) was performed for 35 cycles in a final volume of 50 μL. An annealing temperature of 58°C was used. The amplification products were directly ligated into pTZ18U.

## ACKNOWLEDGMENTS

This work was supported by grants from the Australian Research Council (ARC) and the ARC Special Research Centre for Basic and Applied Plant Molecular Biology.

Received August 17, 1994; accepted October 17, 1994.

## REFERENCES

**Anderson, M.A., McFadden, G.I., Bernatzky, R., Atkinson, A., Orpin, T., Dedman, H., Tregear, G., Fernley, R., and Clarke, A.E.** (1989). Sequence variability of three alleles of the self-incompatibility gene of *Nicotiana alata*. Plant Cell **1,** 483–491.

**Charlesworth, D.** (1994). The key to specificity. Curr. Biol. **4,** 545–546.

**Clark, K.R., and Sims, T.L.** (1994). The *S*-ribonuclease gene of *Petunia hybrida* is expressed in nonstylar tissue, including immature anthers. Plant Physiol. **106,** 25–36.

**Clark, K.R., Okuley, J.J., Collins, P.D., and Sims, T.L.** (1990). Sequence variability and developmental expression of S-alleles in self-incompatible and pseudo-self-compatible petunia. Plant Cell **2,** 815–826.

**Cohen, P.** (1992). Signal integration at the level of protein kinases, protein phosphatases and their substrates. Trends Biochem. Sci. **17,** 408–413.

**Dodds, P.N., Bönig, I., Du, H., Rödin, J., Anderson, M.A., Newbigin, E., and Clarke, A.E.** (1993). *S*-RNase gene of *Nicotiana alata* is expressed in developing pollen. Plant Cell **5,** 1771–1782.

**Eklund, H., Gleason, F.K., and Holmgren, A.** (1991). Structure and functional relations among thioredoxins of different species. PROTEINS: Structure, Function, and Genetics **11,** 13–28.

**Foote, H.C.C., Ride, J.P., Franklin-Tong, V.E., Walker, E.A., Lawrence, M.J., and Franklin, F.C.H.** (1994). Cloning and expression of a distinctive class of self-incompatibility (*S*) gene from *Papaver rhoeas* L. Proc. Natl. Acad. Sci. USA **91,** 2265–2269.

**Franklin-Tong, V.E., and Franklin, F.C.H.** (1992). Gametophytic self-incompatibility in *Papaver rhoeas* L. Sex Plant Reprod. **5,** 1–7.

**Franklin-Tong, V.E., Atwal, K.K., Howell, E.C., Lawrence, M.J., and Franklin, F.C.H.** (1991). Self-incompatibility in *Papaver rhoeas* L.: There is no evidence for the involvement of stigmatic ribonuclease activity. Plant Cell Environ. **14,** 423–429.

**Goring, D.R., and Rothstein, S.J.** (1992). The *S*-locus receptor kinase gene in a self-incompatible *Brassica napus* line encodes a functional serine/threonine kinase. Plant Cell **4,** 1273–1281.

**Guidet, F., Rogowsky, P., Taylor, C., Song, W., and Langridge, P.** (1991). Cloning and characterisation of a new rye-specific repeated sequence. Genome **34,** 81–87.

**Guilluy, C.M., Trick, M., Heizmann, P., and Dumas, C.** (1991). PCR detection of transcripts homologous to the self-incompatibility gene in anthers of *Brassica*. Theor. Appl. Genet. **82,** 466–472.

**Hayman, D.L.** (1956). The genetic control of incompatibility in *Phalaris coerulescens* Desf. Aust. J. Biol. Sci. **9,** 321–331.

**Hayman, D.L.** (1992). The *S-Z* incompatibility system. In Grass Evolution and Domestication, G.P. Chapman, ed (Cambridge: Cambridge University Press), pp. 117–137.

**Hayman, D.L., and Richter, J.** (1992). Mutations affecting self-incompatibility in *Phalaris coerulescens* Desf. (Poaceae). Heredity **68,** 495–503.

**Hinata, K., Watanabe, M., Toriyama, K., and Isogai, A.** (1993). A review of recent studies on homomorphic self-incompatibility. Int. Rev. Cytol. **143,** 257–295.

**Holmgren, A.** (1985). Thioredoxin. Annu. Rev. Biochem. **54,** 237–271.

**Ioerger, T.R., Golke, G.R., Xu, B., and Kao, T.H.** (1991). Primary structural features of the self-incompatibility proteins in the *Solanaceae*. Sex Plant Reprod. **4,** 81–87.

**Kowyama, Y., Kunz, C., Lewis, I., Newbigin, E., Clarke, A.E., and Anderson, M.A.** (1994). Self-compatibility in a *Lycopersicon peruvianum* variant (LA2157) is associated with a lack of style *S*-RNase activity. Theor. Appl. Genet. **88,** 859–864.

**Leach, C.R., and Hayman, D.L.** (1987). The incompatibility loci as indicators of conserved linkage groups in the Poaceae. Heredity **58,** 303–305.

**Lewis, D.** (1963). A protein dimer hypothesis on incompatibility. Proceedings of the Eleventh International Congress on Genetics (Oxford: Pergamon Press), pp. 656–663.

**Marty, I., and Meyer, Y.** (1991). Nucleotide sequence of a cDNA encoding a tobacco thioredoxin. Plant Mol. Biol. **17,** 143–147.

**Matton, D.P., Nass, N., Clarke, A.E., and Newbigin, E.** (1994). Self-incompatibility: How plants avoid illegitimate offspring. Proc. Natl. Acad. Sci. USA **91,** 1992–1997.

**McClure, B.A., Haring, V., Ebert, P.R., Anderson, M.A., Simpson, R.J., Sakiyama, F., and Clarke, A.E.** (1989). Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. Nature **342,** 955–957.

**McClure, B.A., Gray, J.E., Anderson, M.A., and Clarke, A.E.** (1990). Self-incompatibility in *Nicotiana alata* involves degradation of pollen rRNA. Nature **347,** 757–760.

**Nasrallah, J.B., and Nasrallah, M.E.** (1993). Pollen–stigma signaling in the sporophytic self-incompatibility response. Plant Cell **5,** 1325–1335.

**Nasrallah, J.B., Yu, S.M., and Nasrallah, M.E.** (1988). Self-incompatibility genes of *Brassica oleracea*: Expression, isolation, and structure. Proc. Natl. Acad. Sci. USA **85,** 5551–5555.

**Nasrallah, J.B., Nishio, T., and Nasrallah, M.E.** (1991). The self-incompatibility genes of *Brassica*: Expression and use in genetic ablation of floral tissues. Annu. Rev. Plant Physiol. Plant Mol. Biol. **42,** 393–422.

**Nasrallah, J.B., Rundle, S.J., and Nasrallah, M.E.** (1994). Genetic evidence for the requirement of the *Brassica S*-locus receptor kinase gene in the self-incompatibility response. Plant J. **5,** 373–384.

**Newbigin, E., Anderson, M.A., and Clarke, A.E.** (1993). Gametophytic self-incompatibility systems. Plant Cell **5,** 1315–1324.

**Rivera-Madrid, R., Marinho, P., Chartier, Y., and Meyer, Y.** (1993). Nucleotide sequence of an *Arabidopsis thaliana* cDNA clone encoding a homolog to a suppressor of Wilms' tumor. Plant Physiol. **102,** 329–330.

**Royo, J., Kunz, C., Kowyama, Y., Anderson, M., Clarke, A.E., and Newbigin, E.** (1994). Loss of a histidine residue at the active site of *S*-locus ribonuclease is associated with self-compatibility in *Lycopersicon peruvianum.* Proc. Natl. Acad. Sci. USA **91,** 6511–6514.

**Saba-El-Leil, M.K., Rivard, S., Morse, D., and Cappadocia, M.** (1994). The S11 and S13 self incompatibility alleles in *Solanum chacoense* Bitt. are remarkably similar. Plant Mol. Biol. **24,** 571–583.

**Sambrook, J., Fritsch, E.F., and Maniatis, T.** (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

**Singh, A., Ai, Y., and Kao, T.H.** (1991). Characterization of ribonuclease activity of three S-allele-associated proteins of *Petunia inflata.* Plant Physiol. **96,** 61–68.

**Stein, J.C., and Nasrallah, J.B.** (1993). A plant receptor-like gene, the *S*-locus receptor kinase of *Brassica oleracea*, encodes a functional serine/threonine kinase. Plant Physiol. **101,** 1103–1106.

**Stein, J.C., Howlett, B., Boyes, D.C., Nasrallah, M.E., and Nasrallah, J.B.** (1991). Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea.* Proc. Natl. Acad. Sci. USA **88,** 8816–8820.

**Thompson, R.D., and Kirch, H.H.** (1992). The *S* locus of flowering plants: When self-rejection is self-interest. Trends Genet. **8,** 381–387.

**Wehling, P., Hackauf, B., and Wricke, G.** (1994). Phosphorylation of pollen proteins in relation to self-incompatibility in rye (*Secale cereale* L.). Sex Plant Reprod. **7,** 67–75.