The Plant Cell, Vol. 1, 249–258, February 1989, © 1989 American Society of Plant Physiologists

A Highly Conserved *Brassica* Gene with Homology to the S-Locus-Specific Glycoprotein Structural Gene

Beth A. Lalonde, Mikhail E. Nasrallah, Kathleen G. Dwyer, Che-Hong Chen, Bicka Barlow, and June B. Nasrallah[1]

Section of Plant Biology, Cornell University, Ithaca, New York 14853

The S-locus-specific glycoprotein of *Brassica* and the gene encoding it (the SLG gene) are thought to be involved in determining self-incompatibility phenotype in this genus. It has been shown that the *Brassica* genome contains multiple SLG-related sequences. We report here the cloning and characterization of a *Brassica oleracea* gene, SLR1, which corresponds to one of these SLG-related sequences. Like the SLG gene, SLR1 is developmentally regulated. It is maximally expressed in the papillar cells of the stigma at the same stage of flower development as the onset of the incompatibility response. Unlike SLG, the SLR1 genes isolated from different S-allele homozygotes are highly conserved, and this gene, which appears to be ubiquitous in crucifers, is expressed in self-compatible strains as well as self-incompatible strains. Most importantly, we show that the SLR1 gene is not linked to the S-locus and therefore cannot be a determinant of S-allele specificity in *Brassica*.

INTRODUCTION

Self-incompatibility systems in plants prevent self-pollination and are thought to promote outbreeding and increase genetic diversity within species (East, 1940; Lewis, 1979; Nasrallah and Nasrallah, 1986). In *Brassica*, the block to self-fertilization occurs soon after the initial contact between pollen and the stigma surface and is manifested as an inhibition of pollen germination and/or growth of the pollen tube into the surface cells of the stigma (Ockendon, 1972). Self-incompatibility in this genus is under the control of a single genetic locus, the S-locus, for which more than 50 natural alleles have been identified (Thompson, 1957; Ockendon, 1982), and occurs when the pollen and stigma are derived from plants carrying identical S-alleles.

Glycoproteins involved in self-incompatibility, S-locus-specific glycoproteins, have been identified in *Brassica* (Nasrallah and Nasrallah, 1984; Nasrallah et al., 1985b), and the genes encoding these glycoproteins, which we will refer to as SLG genes, have been isolated from several S-genotypes (Nasrallah et al., 1987). Expression of the SLG genes is under precise temporal and spatial regulation. *Brassica* stigmas are initially fully self-compatible and only become self-incompatible at about 1 day before flower opening. RNA blot analysis has demonstrated that the SLG gene is expressed in the stigma but not in style or leaf tissue and that maximal accumulation of SLG transcripts in the stigma occurs at the same developmental stage as the onset of the self-incompatible response (Nasrallah et al., 1985a). In situ hybridization experiments have demonstrated the localization of SLG transcripts in the papillar cells, the outermost layer of cells of the stigma (Nasrallah et al., 1988).

[1]To whom correspondence should be addressed.

Variability at the S-locus is reflected in extensive polymorphism among the S-locus-specific glycoproteins identified in different S-allele homozygotes (Nasrallah and Nasrallah, 1984; Nasrallah et al., 1985b) and among the SLG gene sequences isolated from these strains (Nasrallah et al., 1987). The cosegregation of these polymorphisms with S-alleles in F2 populations of plants has demonstrated the linkage of the SLG gene to the S-locus. Interestingly, DNA gel blot analysis of *Brassica* genomic DNA has revealed the existence of a family of sequences with varying degrees of homology to the SLG gene. Whether these sequences represent a family of genes that are functionally related to the SLG gene has not been determined. Some of the SLG-related sequences have been cloned and shown to contain out-of-frame deletions and insertions and so, clearly, are not functional genes.

In this paper, we report on the isolation and characterization of an expressed gene that corresponds to one of these SLG-related sequences. We describe the similar temporal and spatial patterns of expression of this gene and the SLG gene and the similarities between the predicted amino acid sequences of the two genes. We also report on substantial differences between these genes regarding linkage to the S-locus and involvement in the determination of S-allele specificity.

RESULTS

Isolation and Genomic Organization of SLR1

To determine whether any of the SLG-related genomic sequences are expressed, a cDNA library constructed

previously from mature *Brassica oleracea* stigmas (Nasrallah et al., 1985a) was subjected to hybridization with both stigma-derived cDNA probes and seedling cDNA probes. Several clones were isolated which hybridized to the stigma probes but not to seedling probes. All of these stigma-specific clones were shown to be homologous by cross-hybridization experiments and, upon further analysis, to fall into two classes. One class corresponded to the previously described SLG gene (Nasrallah et al., 1985a). The second class showed at least 10-fold less homology, as determined by hybridization analysis, to the SLG clones. cDNA and genomic sequences corresponding to this class will be designated as S-locus-related and abbreviated as SLR1.

The most striking difference between the SLG and SLR1 cDNAs is revealed by the very different genomic hybridization patterns exhibited by the two when used as probes. As shown in Figure 1A, when genomic DNA from a variety of *Brassica* species and S-genotypes is cut with restriction enzymes and subjected to DNA gel blot analysis, the SLG gene probe produces a complex pattern of bands with extensive size polymorphism observed between different S-allele homozygotes. It was shown previously that these restriction fragment polymorphisms segregate with the S-locus in F2 populations, indicating that the genomic sequences corresponding to these fragments are at the S-locus or very tightly linked to it (Nasrallah et al., 1985a). Furthermore, and as shown in Figure 1A, the intensity of hybridization differed among *Brassica* strains homozygous for different S-alleles. The SLG gene probe hybridized strongly to at least two bands and more weakly to several other bands in the majority of *B. oleracea* strains tested and in the *B. campestris* S8 homozygote. On the other hand, the *B. oleracea* S5 and S15 homozygotes showed very little, if any, hybridization to the SLG gene probe.

By contrast, a probe made from one of the SLR1 cDNA clones yielded a simpler genomic hybridization pattern. As shown in Figure 1B, a single strongly hybridizing band, exhibiting very little S-allele-associated size polymorphism, was identified in each of the *B. oleracea* and *B. campestris* strains analyzed, including the S5 and S15 strains. Furthermore, and as shown in Figure 1A, the genomic SLR1-homologous band corresponded, in every case, to one of the bands hybridizing weakly to the SLG probe, indicating that the SLR1 gene is one of the multiple SLG-related sequences described previously (Nasrallah et al., 1985a).

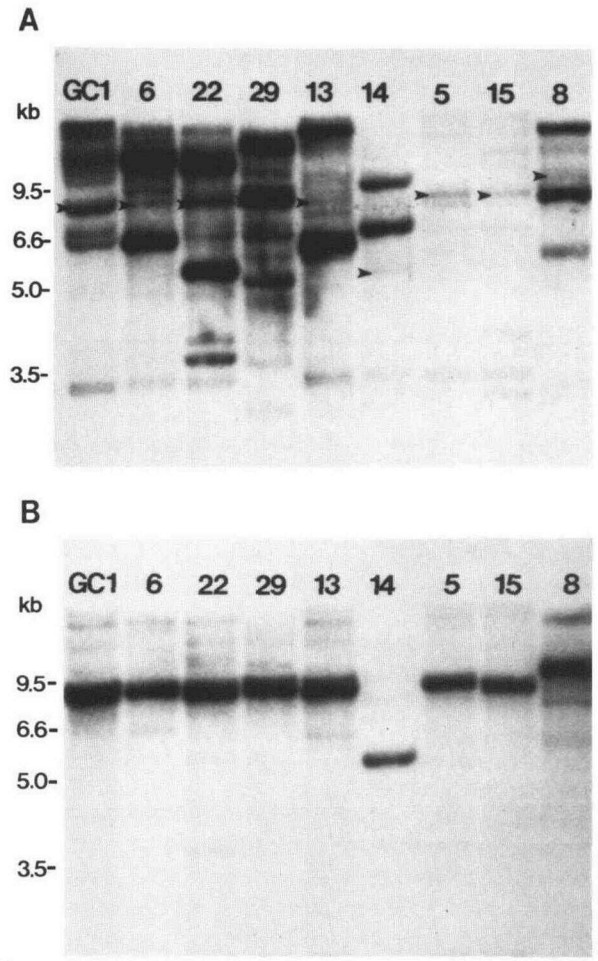

Figure 1. DNA Gel Blot Analysis of Sequences Homologous to the SLG Gene and the SLR1 Gene in the *Brassica* Genome.

Brassica genomic DNA was cut with EcoRI. Numbers above the lanes refer to the S-allele present in each homozygous plant. GC1 is a broccoli cultivar with an unidentified S-allele. Size markers are indicated to the left of the gel.

(A) Genomic sequences hybridizing to the SLG gene. The probe used was an S6 SLG cDNA. The arrowheads show the positions of the bands homologous to the SLR1 cDNA probe.

(B) Genomic sequences hybridizing to the SLR1 gene. The probe used was an EcoRI fragment bearing an SLR1 cDNA.

DNA Sequence of the SLR1 cDNA from the S22 Genotype of Kale

To understand more thoroughly the relationship of the SLR1 gene to the SLG gene, the DNA sequences of both genes isolated from a single S-homozygous genotype were examined. A stigma-derived cDNA library was constructed from kale plants homozygous for the S22 allele. Of 120,000 λgt10 recombinant plaques screened, 595 plaques were found to hybridize strongly to an SLR1 cDNA probe and 79 to an SLG gene cDNA probe. The SLR1 probe also hybridized weakly to the 79 SLG gene plaques. Thus, SLR1 appears to be a highly expressed gene, its transcript being approximately sevenfold more abundant than the SLG gene in this S22 stigma-cDNA library.

The two SLR1-homologous S22 cDNA clones with the longest inserts, 1.0 kb and 1.3 kb, were subcloned into a Bluescript plasmid (Stratagene) to generate the plasmids

```
 TG AGA GGT GTA ATA CCA AAC TAT CAT CAC TCT TAC ACC TTA CTC TTT TTC GTT ATA TTG
Met Arg Gly Val Ile Pro Asn Tyr His His Ser Tyr Thr Leu Leu Phe Phe Val Ile Leu
-28
60
GTT CTG TTT CCT CAT GTG TTC TCG ACC AAT ACT TTG TCA CCT AAC GAA GCT CTT ACA ATA
Val Leu Phe Pro His Val Phe Ser Thr Asn Thr Leu Ser Pro Asn Glu Ala Leu Thr Ile
                                1
120
TCA AGC AAC AAA ACC CTT GTG TCT CCC GGT GAT GTC TTC GAG CTT GGC TTC TTC AAA ACC
Ser Ser Asn Lys Thr Leu Val Ser Pro Gly Asp Val Phe Glu Leu Gly Phe Phe Lys Thr

180
ACC ACA AGA AAC TCT CCA GAT GGT ACT GAT CGT TGG TAT CTC GGT ATT TGG TAC AAG ACA
Thr Thr Arg Asn Ser Pro Asp Gly Thr Asp Arg Trp Tyr Leu Gly Ile Trp Tyr Lys Thr

240
ACC TCT GGT CAT AGA ACA TAT GTT TGG GTT GCC AAC AGA GAC AAC GCT CTT CAC AAC TCC
Thr Ser Gly His Arg Thr Tyr Val Trp Val Ala Asn Arg Asp Asn Ala Leu His Asn Ser

300
ATG GGG ACA CTC AAA ATC TCT CAC GCT AGC CTC GTC CTC CTT GAC CAT TCT AAT ACT CCT
Met Gly Thr Leu Lys Ile Ser His Ala Ser Leu Val Leu Leu Asp His Ser Asn Thr Pro

360
GTA TGG TCA ACG AAC TTT ACG GGA GTT GCG CAT TTA CCA GTT ACG GCA GAG CTT CTC GCT
Val Trp Ser Thr Asn Phe Thr Gly Val Ala His Leu Pro Val Thr Ala Glu Leu Leu Ala

420
AAC GGC AAC TTC GTG CTT AGA GAC TCC AAA ACC AAC GAC CTA GAC CGG TTC ATG TGG CAG
Asn Gly Asn Phe Val Leu Arg Asp Ser Lys Thr Asn Asp Leu Asp Arg Phe Met Trp Gln

480
AGC TTT GAT TAT CCG GTG GAT ACT TTG CTC CCG GAG ATG AAA CTT GGT CGG AAT CGC AAC
Ser Phe Asp Tyr Pro Val Asp Thr Leu Leu Pro Glu Met Lys Leu Gly Arg Asn Arg Asn

540
GGT TCA GGA AAC GAA AAA ATC CTC ACA TCT TGG AAA AGC CCT ACT GAT CCA TCA AGT GGA
Gly Ser Gly Asn Glu Lys Ile Leu Thr Ser Trp Lys Ser Pro Thr Asp Pro Ser Ser Gly

600
GAT TAT TCG TTC ATA CTC GAA ACC GAA GGG TTT TTA CAT GAG TTT TAT CTA CTG AAC AAT
Asp Tyr Ser Phe Ile Leu Glu Thr Glu Gly Phe Leu His Glu Phe Tyr Leu Leu Asn Asn

660
GAG TTC AAA GTG TAC CGA ACC GGT CCT TGG AAC GGA GTC CGG TTT AAC GGC ATA CCA AAA
Glu Phe Lys Val Tyr Arg Thr Gly Pro Trp Asn Gly Val Arg Phe Asn Gly Ile Pro Lys

720
ATG CAA AAC TGG AGC TAC ATT GAT AAC AGT TTC ATA GAT AAC AAC AAG GAA GTC GCG TAC
Met Gln Asn Trp Ser Tyr Ile Asp Asn Ser Phe Ile Asp Asn Asn Lys Glu Val Ala Tyr

780
AGT TTC CAA GTC AAC AAC AAC CAC AAC ATC CAC ACA AGA TTT AGA ATG AGT TCC ACA GGG
Ser Phe Gln Val Asn Asn Asn His Asn Ile His Thr Arg Phe Arg Met Ser Ser Thr Gly

840
TAC TTA CAA GTA ATC ACA TGG ACT AAG ACA GTA CCG CAA CGT AAC ATG TTT TGG TCG TTC
Tyr Leu Gln Val Ile Thr Trp Thr Lys Thr Val Pro Gln Arg Asn Met Phe Trp Ser Phe

900
CCG GAA GAT ACA TGC GAT CTG TAC AAA GTT TGT GGT CCT TAC GCT TAC TGT GAC ATG CAC
Pro Glu Asp Thr Cys Asp Leu Tyr Lys Val Cys Gly Pro Tyr Ala Tyr Cys Asp Met His

960
ACG TCG CCT ACG TGT AAC TGT ATC AAA GGC TTC GTT CCC AAG AAT GCT GGA AGA TGG GAT
Thr Ser Pro Thr Cys Asn Cys Ile Lys Gly Phe Val Pro Lys Asn Ala Gly Arg Trp Asp

1020
TTG AGA GAT ATG TCA GGT GGT TGT GTG AGG AGC TCG AAG CTA AGC TGT GGA GAG GGT GAT
Leu Arg Asp Met Ser Gly Gly Cys Val Arg Ser Ser Lys Leu Ser Cys Gly Glu Gly Asp

1080
GGG TTT CTG CGG ATG AGT CAG ATG AAG CTA CCG GAG ACA AGC GAA GCG GTT GTG GAC AAG
Gly Phe Leu Arg Met Ser Gln Met Lys Leu Pro Glu Thr Ser Glu Ala Val Val Asp Lys

1140
AGG ATC GGG TTG AAG GAA TGC AGG GAG AAG TGT GTT AGA GAT TGT AAC TGT ACC GGG TAT
Arg Ile Gly Leu Lys Glu Cys Arg Glu Lys Cys Val Arg Asp Cys Asn Cys Thr Gly Tyr

1200
GCG AAT ATG GAT ATC ATG AAT GGT GGG TCG GGA TGT GTG ATG TGG ACC GGA GAG CTC GAT
Ala Asn Met Asp Ile Met Asn Gly Gly Ser Gly Cys Val Met Trp Thr Gly Glu Leu Asp

1260
GAT ATG CGG AAG TAC AAT GCT GGA GGT CAA GAT CTT TAT GTC AAG GTA GCA GCT GCT AGT
Asp Met Arg Lys Tyr Asn Ala Gly Gly Gln Asp Leu Tyr Val Lys Val Ala Ala Ala Ser

1320
CTT GTC CCC TCG TAG ACATGTATCCAATAATATGCGAGATTTTTAAAAATATTTGTATCGATTGAAAGTATCTAA
Leu Val Pro Ser End

AAATAAATAATATGAAA
```

Figure 2. Nucleotide Sequence and Deduced Amino Acid Sequence of the S22 SLR1 cDNA.

Nucleotides are numbered above the DNA sequence. The first nucleotide of the initiation codon was not present in the cDNA. A stop codon is present at nucleotide 1332 and a presumptive polyadenylation signal at 1396. The predicted amino acid sequence is presented below the nucleotide sequence. Numbers below the sequence refer to amino acid residues. Residue number 1 is the predicted first amino acid of the mature protein.

pBL204 and pBL205, respectively. Restriction maps of these plasmids showed that they contained overlapping sequences. Appropriate restriction fragments were subcloned from each plasmid into the M13 sequencing vectors mp18 and mp19 to allow overlapping sequence determination of both DNA strands. Similarly, S22 SLG cDNA clones were characterized, subcloned into M13 vectors, and sequenced.

Figure 2 shows the DNA sequence and predicted amino acid sequence of the S22 SLR1 cDNA. Neither pBL204 nor pBL205 contains a full-length cDNA insert. Each contains a portion of the same single open reading frame. pBL204 extends from nucleotide 477 to nucleotide 1411 and includes a stop codon at nucleotide 1332 and a potential polyadenylation signal at nucleotide 1396 but does not contain a polyA tract. pBL205 is truncated at the 3′ end; it extends from nucleotide 1 to nucleotide 1307. Restriction enzyme analysis of an S22 SLR1 genomic clone is consistent with the cDNA sequence presented here (data not shown). In addition, and as discussed later in this paper, the DNA sequence of an S6 SLR1 genomic clone confirms the length and position of the open reading frame presented in Figure 2.

The S22 SLR1 cDNA sequence and the S22 SLG cDNA sequence contain approximately 68% (905/1334) identical nucleotides overall. As Figure 3 shows, the predicted amino acid sequences for these two genes bear a striking resemblance to one another. Polypeptides of 46.3 and 46.5 kilodaltons are predicted for the SLG- and SLR1-encoded proteins, respectively. They contain approximately 58% (256/444) identical residues overall with several stretches of 10 or 11 identical residues. A hydropathy plot of the predicted SLR1 protein sequence shows an extremely hydrophobic stretch at the 5′ amino terminus extending from residues −28 to −1. This hydrophobic sequence is 42% conserved with respect to a similar hydrophobic N-terminal stretch in the SLG-encoded protein. The sequence of the amino terminus of mature S-locus-specific glycoprotein has been determined by direct amino acid sequencing (Takayama et al., 1986) and shown to correspond to residues 1 to 29 of the SLG amino acid sequence. It was proposed, therefore, that the hydrophobic sequence extending from residues −1 to −32 of the SLG protein encodes a signal peptide (Nasrallah et al., 1987). Because residues 1 to 29 are 80% conserved between the SLG and SLR1 proteins, we propose that these residues correspond to the amino terminus of the SLR1 protein and that residues −1 to −28 encode a signal peptide.

Comparison of the predicted amino acid sequences of the SLG genes isolated from the S6, S13, and S14 alleles of *B. oleracea* had demonstrated the presence of alternating conserved and diverged regions: Region A, containing amino acid residues 1 to 181, with 80% identical residues; Region B, from residue 182 to 274, with 42% identical residues; Region C, from residue 275 to 377, with 81% identical residues including 12 invariant cysteines; and

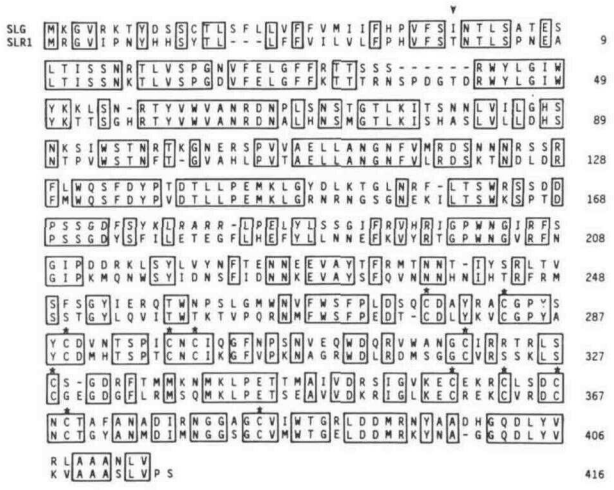

Figure 3. Comparison of the Predicted Amino Acid Sequences of the *B. oleracea* S22 SLR1 and SLG Genes.

The amino acid sequences were predicted from the nucleotide sequences determined for SLG and SLR1 cDNA clones isolated from an S22 homozygote. Identical residues are boxed. The arrowhead indicates the predicted, first amino acid of the mature protein. Dashes represent gaps introduced to maximize similarity. Cysteine residues are indicated by asterisks above the sequences.

Region D, from residue 378 to 405, with 46% identical residues (Nasrallah et al., 1987). When the S22 SLR1 amino acid sequence was compared with the S22 SLG gene protein sequence, a similar arrangement of conserved and diverged domains was apparent. A region of the SLR1 gene corresponding roughly to Region A and containing residues 1 to 148 shares 70% conserved residues with the SLG gene, and a region corresponding to Region B, from residue 149 to 266, contains only 42% conserved residues. A third region, from residue 267 to the end of the SLR1 gene, is more diverged from the corresponding Regions C and D of the SLG gene, containing 62% conserved residues. Interestingly, this region contains 12 cysteine residues which are, with the exception of a single amino acid insertion between cysteines 7 and 8, spaced exactly as in the SLG genes.

Despite the similarities between the predicted SLG and SLR1 amino acid sequences, several lines of evidence argue against the SLR1 gene product being an S-locus-specific glycoprotein. Of 10 potential N-glycosylation sites in the S22 SLG-encoded protein (residues 15, 90, 97, 124, 222, 238, 260, 292, 304, and 368), only three are conserved in the S22 SLR1-encoded protein (residues 15, 97, and 368). Also, the calculated pI value of 8.7 for the SLR1-encoded protein is quite different from the pI values of >10 calculated for the SLG-encoded proteins from a number of S genotypes. Finally, monoclonal antibodies raised against authentic S-locus-specific glycoproteins do not recognize a second protein on protein gel blots of S22 stigma proteins, nor do they recognize bacterial β-galactosidase-SLR1 fusion proteins. Thus, we conclude that the SLR1 gene codes for a protein related to but distinct from the SLG gene product.

Regulation of SLR1

It has been shown that expression of the SLG gene in the *Brassica* stigma occurs exclusively in the papillar cells (Nasrallah et al., 1988) and that this expression is temporally regulated during stigma development (Nasrallah et al., 1985a, 1985b). To determine whether the SLR1 gene shows a similar pattern of expression, RNA purified from staged stigmas and from leaf and style tissue was subjected to RNA gel blot analysis. Because the SLR1 transcript is approximately sevenfold more abundant than the SLG transcript in *Brassica* stigmas and the SLR1 and SLG gene sequences are only 68% conserved, it is likely that the signal observed on RNA gel blots is due primarily to hybridization to SLR1 transcripts. Figure 4 shows that, when the SLR1 cDNA was used as a hybridization probe, a message approximately 1.6 kb in length was detected in stigma tissue but not in style tissue. Furthermore, the levels of this RNA varied during stigma development.

Flowers and flower buds along a *Brassica* inflorescence represent a precise developmental progression, the stages

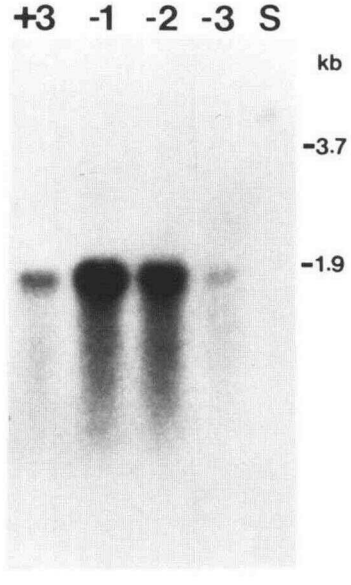

Figure 4. Regulation of SLR1-Homologous Transcripts in S6 Homozygous Plants.

Lane S contains total cellular RNA isolated from style tissue. Lanes −3, −2, −1, and +3 contain total cellular RNA isolated from stigmas 3, 2, and 1 days before and 3 days after flower opening, respectively. The positions of the rRNA bands are indicated at the right. The probe used was an EcoRI fragment containing the entire S22 SLR1 cDNA insert of pBL205.

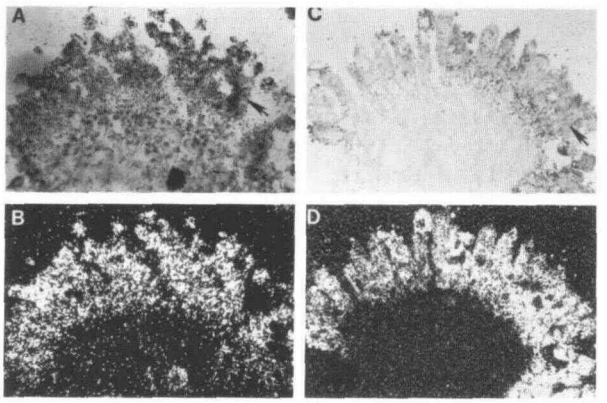

Figure 5. In Situ Localization of SLR1 Transcripts in the Developing *Brassica* Stigma.

(A) Bright-field view of a longitudinal section through a stigma at approximately 3 days prior to flower opening. The arrow indicates the layer of developing papillar cells.

(B) Dark-field view of the section in **A.** The formation of specific DNA:RNA hybrids is indicated by the concentration of white grains over the surface cell layer of the stigma.

(C) Bright-field view of a longitudinal section through one of the two lobes of a stigma at approximately 1 day prior to flower opening. The arrow indicates the layer of some 20 fully enlarged papillar cells measuring approximately 30 × 125 μm.

(D) Dark-field view of the section in **C.** Note the concentration of white grains over the layer of papillar cells indicating specific DNA:RNA hybrid formation in this region.

(**A** and **B**, magnification ×60; **C** and **D**, magnification ×30)

of which are most simply designated in terms of days from flower opening or anthesis (stage 0). Stigmas are first self-compatible in the younger buds (stages −4,−3,−2) and become self-incompatible in buds at 1 day before anthesis (stage −1). SLR1 transcript levels reached a maximum 1 day before anthesis, whereas very low levels of message were detectable prior to 2 days before anthesis (stage −3) and in mature flowers (stage +3). No signal was detected in leaf tissue (data not shown) or style tissue.

This pattern of expression is identical to the pattern seen for SLG gene expression (Nasrallah et al., 1985a). In addition, as shown in Figure 5 and as already demonstrated for the SLG gene transcripts (Nasrallah et al., 1988), the SLR1 transcripts have been localized by in situ hybridization experiments to the papillar cells of the stigma surface. These experiments show that SLR1 transcripts can be detected, not only in fully differentiated papillar cells (Figure 5, C and D), but also very early in stigma development, in the immature papillar cells (Figure 5, A and B).

SLR1 Is not Linked to the S-Locus

The similar patterns of expression of the SLR1 and SLG genes suggest that the SLR1 gene might play a role in the incompatibility response of *Brassica*. Although genetic evidence argues that self-incompatibility is controlled by a single locus, the S-locus (Thompson, 1957), the possibility exists of a second tightly linked gene or cluster of genes. The linkage of the SLR1 gene to the S-locus was therefore examined. As shown in Figure 1B, DNA gel blot analysis of EcoRI-digested DNA from the S6 kale homozygote and from the S14 cabbage homozygote revealed a useful DNA polymorphism at the SLR1 locus. Due to the long life cycle of the kale and cabbage cultivars, an F2 population was not generated specifically for this linkage analysis. However, five vegetatively maintained F2 progeny plants from a cross between the S6 and S14 homozygous inbreds were available, and were subjected to DNA gel blot analysis.

The self-incompatibility genotype of these plants was determined by pollination analysis and associated S-locus-specific glycoprotein polymorphism, and is shown above the lanes in Figure 6, A and B. Hybridization of the filter with an SLG cDNA probe revealed that SLG DNA polymorphisms cosegregate with the S-genotype (Figure 6A). A very different result is obtained, however, when the same filter is probed with the SLR1 cDNA. As shown in the fourth and fifth lanes of Figure 6B, two of the F2 plants determined to be homozygous for the S6 allele exhibit both the restriction fragment characteristic of the kale parental line and that characteristic of the cabbage parental line, and are therefore heterozygous at the SLR1 locus.

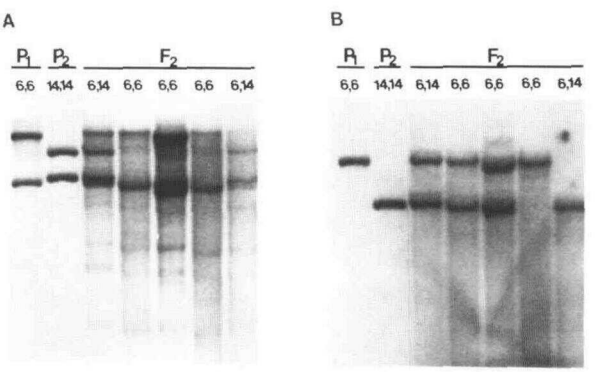

Figure 6. DNA Gel Blot Analysis of Plants Segregating for the S6 and S14 S-Alleles Demonstrates that SLR1 Is Unlinked to the S-Locus.

DNA was isolated from homozygous S6 and S14 plants and from the F2 progeny of a selfed F1 plant derived from a cross between the S6 and S14 homozygotes. DNA was digested with EcoRI. Lane P1, S6-homozygous parent; lane P2, S14-homozygous parent; lanes F2, F2 plants. The numbers above the lanes indicate the S-genotype of each plant as determined by pollination assays and by the presence of characteristic S-locus-specific glycoprotein polymorphisms.

(A) Hybridization with the SLG gene probe. The probe used was an EcoRI fragment bearing the entire S6 SLG cDNA.

(B) Hybridization with the SLR1 gene probe. The probe used was an EcoRI fragment bearing the pBL205 S22 SLR1 cDNA.

Similarly, one F2 plant determined to be heterozygous for the S6 and S14 alleles (right lane in Figure 6B) exhibits only the restriction fragment characteristic of the cabbage parental line and is therefore homozygous at the SLR1 locus. Thus, the SLR1 polymorphism segregates independently of S-genotype in these F2 plants, demonstrating that the SLR1 gene is not linked genetically to the S-locus.

SLR1 Is a Highly Conserved Gene

The presence of a single, generally nonpolymorphic, SLR1-homologous band on genomic DNA gel blots of a variety of *Brassica* strains suggested that SLR1 is a single, highly conserved gene. To examine the extent of sequence conservation among SLR1 genes, the genomic copies of SLR1 isolated from S6 and S13 *B. oleracea* strains were sequenced. Sequence extending from −362 to 1698 (using the same numbering scheme as in Figure 2) for the S6 gene and from −4 to 1370 for the S13 gene was determined. The S6 SLR1 sequence confirms the presence of a single open reading frame beginning at the methionine residue at position −28 of the S22 SLR1 amino acid sequence (Figure 2). The S22 and S13 SLR1 genes were identical, whereas the S22 and S6 SLR1 gene sequences differed only at nucleotide 1076, at which a thymine was replaced by an adenine. The deduced amino acid sequences are identical nevertheless, since this substitution does not result in an amino acid substitution. By contrast, the SLG amino acid sequences determined for the same S22, S13, and S6 strains are only about 80% conserved (Nasrallah et al., 1987).

Widespread sequence conservation among SLR1 genes is also indicated by the detection of SLR1-homologous sequences on genomic DNA gel blots of a wide variety of crucifer species, including incompatible strains of *B. nigra* and *Raphanus sativus*, and several self-compatible species such as *B. carinata*, *B. napus*, *B. juncea*, *Arabidopsis thaliana*, and *yellow sarson*, a self-compatible strain of B. campestris.

SLR1 Is Expressed in Self-compatible Strains

The extreme sequence conservation of SLR1 genes, even among strains that have diverged widely for other genes, is consistent with the SLR1 gene product playing a fundamental role in pollination. It was reasoned that if this is indeed the case, the SLR1 gene should be expressed equally in both self-compatible and self-incompatible strains. To address this question, a comparative analysis of the self-compatible *B. campestris* variety *yellow sarson* (YS), and a self-incompatible *B. campestris* strain homozygous for the S8 allele was undertaken. Self-compatibility in the *yellow sarson* line has been demonstrated to be due to the action of a recessive modifier locus unlinked to the S-locus (Hinata et al., 1983).

An F2 population of plants was generated from a cross between these strains and shown to segregate for self-compatibility by microscopic monitoring of pollen tube growth following self-pollination. As Figure 7 shows, when RNA isolated from stigmas of the parental and F2 plants was subjected to RNA gel blot analysis, the SLR1 probe detected a strong signal in the incompatible, S8 homozygous parent, the compatible YS parent, and all of the F2 plants, whether self-compatible or incompatible. In addition, an SLR1 cDNA clone was isolated from the self-compatible YS strain and shown to be approximately 97% conserved with respect to the S22 SLR1 cDNA sequence; 581 of 602 residues sequenced are conserved (data not shown).

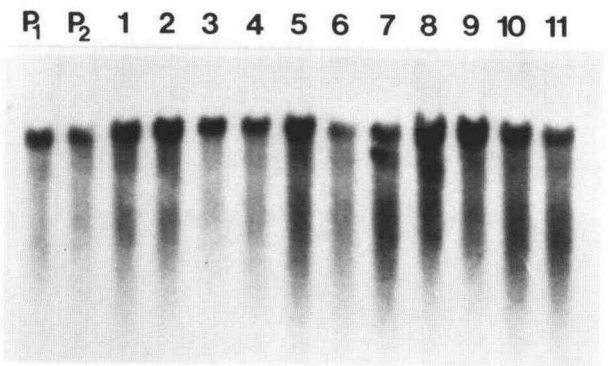

Figure 7. RNA Gel Blot Analysis of Compatible and Incompatible *B. campestris* Strains.

Lanes labeled P1 and P2 contain total stigma RNA from the self-incompatible S8 and self-compatible YS parent strains, respectively. Lanes 1 to 11 contain stigma RNA from the F2 progeny of F1 plants derived from a cross between the compatible and incompatible parents. Among the F2 population, self-incompatible plants are represented in lanes 1, 3, 6, 7, and 10, and self-compatible plants in lanes 2, 4, 5, 8, 9, and 11. The probe used was an S22 SLR1 cDNA gene.

Thus, the SLR1 gene is highly expressed in strains exhibiting either a self-incompatible or a self-compatible phenotype. This pattern of expression is consistent with a role for the SLR1 gene in a process that is required for pollination in plants of both types.

DISCUSSION

In this paper, we demonstrate that at least one of the multiple S-locus-related sequences present in the *Brassica* genome corresponds to an expressed gene, the SLR1 gene. Thus, it appears that the S gene family encodes at least two proteins, the S-locus-specific glycoprotein and the related SLR1-encoded protein.

The SLR1 gene was isolated as a stigma-specific cDNA clone with sequence homology to the SLG gene. Like SLG (Nasrallah et al., 1985a, 1988), it is developmentally regulated, being expressed only in the stigma and exhibiting maximal levels of transcript accumulation coincident with the onset of the incompatibility response in the developing flowers. Although very similar in broad outline, the patterns

of expression of the two genes differ in detail. When compared with SLG, SLR1 is expressed to a higher degree, as judged by the relative abundance of the corresponding sequences in the same cDNA library. In addition, and perhaps as a consequence of their greater abundance, SLR1 transcripts can be detected by in situ hybridization experiments 1 day earlier than SLG transcripts. The relatedness of the two genes is also evident from a comparison of their predicted protein products. The protein sequences share approximately 58% amino acid residues overall, and have a number of features in common. Among these are extensive regions of identity, a putative signal sequence at the amino terminus, and a region containing 12 cysteine residues at the carboxy terminus that are highly conserved between these two genes and among SLG genes isolated from different S-allele homozygotes.

The similarities in the spacial and temporal patterns of expression of SLR1 and SLG and in their predicted protein products are suggestive of a role for the SLR1 gene in determining incompatibility phenotype. The data presented in this paper demonstrate, however, that SLR1 differs from SLG in a number of significant aspects and is not likely to contribute directly to the specificity of S-alleles.

The conclusion that a gene and its protein product are involved in allelic specificity is predicated largely on the demonstration of S-allele-associated variability in the DNA and protein sequences encoded by this gene, and on the cosegregation of these polymorphisms with the S-alleles. The SLG gene fulfills these criteria. On genomic DNA gel blots, the SLG probe reveals extensive restriction fragment length polymorphisms and variations in hybridization intensities between different S-allele homozygotes, reflecting the high degree of allelic polymorphism at the S-locus (Nasrallah and Nasrallah, 1984; Nasrallah et al., 1985a). This variability is reflected also in the sequences of SLG genes isolated from different S-allele homozygotes. Although these genes are more than 90% homologous at the nucleotide level, they encode proteins that are 80% homologous in their most conserved regions and only 40% in their variable region (Nasrallah et al., 1987). Finally, segregation analysis of F2 populations shows that DNA and protein polymorphisms encoded by the SLG gene cosegregate with the S-alleles (Nasrallah et al., 1985a).

The SLR1 sequence, on the other hand, shows limited polymorphism on genomic DNA gel blots and hybridizes with equal intensity to all S-genotypes analyzed. Furthermore, comparison of SLR1 gene sequences isolated from three different S-allele homozygotes has demonstrated that this gene is highly conserved and encodes identical proteins in *Brassica* strains that differ in their S alleles. Most importantly, we have shown through segregation analysis of F2 plants that, unlike the SLG gene, this gene is not linked to the S-locus, and cannot, therefore, be involved in determining the self-incompatibility specificity of *Brassica* plants.

It is tempting to use a comparison of the SLR1 and SLG protein sequences to speculate on where the SLR1 and SLG gene functional domains lie. Because the interactions of incompatibility take place at the pollen-papillar cell interface, it has been suggested that the S-locus-specific glycoproteins, which are synthesized in the papillar cells, are transported to the papillar cell surface (Nasrallah et al., 1987). We propose that the highly related signal sequences of the SLG and SLR1 genes serve as transport signals to direct the SLG and SLR1 proteins to the surface of these cells. The amino terminal one-third, Region A, of the SLR1- and SLG-encoded proteins are highly conserved (73% homology), making it unlikely that specificity for either gene function lies in this region, but leaving it a candidate for a region of shared function. Conversely, the high variability in Region B, among SLG genes isolated from different S-allele homozygotes, makes it a candidate for a region controlling S-allele specificity.

A particularly interesting feature of the SLG and SLR1 genes is the extreme conservation of spacing between the 12 cysteine residues in Region C of these genes, despite the lack of overall sequence conservation (only 60% homology) in this region. This pattern of cysteines is consistent with a role for Region C in producing conserved tertiary or quaternary configurations that might be necessary for a function shared by both the SLR1 and SLG proteins. Other systems in which cysteine clusters have been reported include the barley thionine multigene family whose members code for cell wall-localized polypeptides with antifungal activity (Bohlmann et al., 1988) and a large collection of metal binding proteins, many of which have been implicated in DNA binding and gene regulation activities (Evans and Hollenberg, 1988). Whether the SLR1 gene product binds metals or has regulatory properties remains to be seen, although the high levels of expression of this gene are more in keeping with a structural function than a regulatory one.

Although members of the same S-multigene family, the SLR1 and SLG genes differ in their sequence homology to each other and to other members of the family. SLR1 is highly conserved, with little sequence polymorphism among different S-genotypes and little homology to other members of the S-family of genes; it hybridizes strongly to only one restriction fragment of genomic DNA. The SLG sequence, on the other hand, shows substantial sequence polymorphism among S-genotypes and hybridizes to several different restriction fragments of *Brassica* genomic DNA. Thus, although the SLG gene has diverged among different S-genotypes, it has apparently maintained some degree of homology to the other related sequences. The SLG gene and other members of the family appear, however, to have diverged significantly from the SLR1 gene. Whatever the evolutionary relationships among these sequences, two genes have emerged: one, the SLR1 gene, perhaps supplying a product required for some fundamental aspect of pollination; the other, the SLG gene, performing similar functions, with similar temporal and spatial

patterns of expression, but adding a degree of specificity to the pollination process.

The SLR1 protein product has not been identified in *Brassica* stigmas as yet and a search of the GenBank data base has uncovered no significant homology of either SLR1 or SLG to any other protein sequences. The extreme degree of sequence conservation and the ubiquitous nature of the SLR1 gene in crucifers suggests, however, that its product performs an important function in *Brassica*, one that cannot tolerate even minor sequence alterations. The expression of the SLR1 gene in the surface papillar cells of the stigma is consistent with a role for this gene in the highly localized reactions of pollination. In addition, we have shown that the SLR1 gene is expressed at high levels in a self-compatible *B. campestris* strain, in keeping with a possible role for this gene in some global aspect of pollination. Should the SLR1-homologous sequences that have been detected in other self-compatible crucifer species such as *Arabidopsis* also be expressed, a role for this gene in pollination events common to all crucifers would be indicated.

The interaction between pollen and papillar cells during pollination is a complex one and consists of many steps involving pollen contact and adhesion, pollen hydration, germination, and pollen tube growth (Heslop-Harrison, 1975). At each of these steps, a set of responses must be elicited in both pollen and pistil tissue. We propose that the S gene family is a cohort of related sequences involved in the complex cellular interactions that occur at the papillar cell-pollen interface and that the SLR1 gene is one member of this family that plays a more general role in these pollination events than the SLG gene.

METHODS

Plant Material

The *Brassica oleracea* var. *acephala* inbred kale lines homozygous for the S5, S6, S13, S15, and S22 self-incompatibility alleles and the *B. oleracea* var. *alboglabra* line homozygous for the S29 self-incompatibility allele were derived from plant material initially obtained from the Gene Bank Facility at Wellesbourne, U.K., courtesy of D. J. Ockendon. The S14 homozygote is an inbred line developed at Cornell University and belongs to the variety *capitata* (cabbage). *B. oleracea* var.*italica* is a commercially available broccoli line. The *B. campestris* self-incompatible S8 homozygous line and the self-compatible *yellow sarson* variety (USDA strain C634) were kindly provided by K. Hinata. *B. carinata*, *B. napus*, *B. nigra*, *B. juncea*, and *Raphanus sativus* are rapid cycling strains and were obtained from the Crucifer Genetics Cooperative, courtesy of Paul Williams. *Arabidopsis thaliana* (L.) Heynh. is descended from the Columbia wild-type strain and was obtained from C. Somerville.

Pollination Assays

Stigmas were harvested from *Brassica* flowers and brushed with pollen granules from isolated anthers. The pollinated stigmas were stained with decolorized aniline blue, and pollen-tube development was monitored by fluorescence microscopy (Linskens and Esser 1957; Martin, 1959).

In Situ Hybridization

In situ localization of SLR1 transcripts was performed on frozen sections of *Brassica* stigmas. Frozen sections were prepared and hybridized with a ^{3}H-dCTP-labeled SLR1 cDNA insert as described previously (Nasrallah et al., 1988). The specificity of hybridization was verified in control experiments in which labeled pUC8 DNA or heterologous probes showing no homology to *Brassica* RNA resulted in background levels of grains equally distributed over and outside the sections.

DNA Gel Blot Analysis

DNA was extracted from nuclei isolated from *B. oleracea* leaves by the procedure of Bingham et al. (1981) and purified by CsCl/ethidium bromide centrifugation. The DNA was digested with restriction enzymes and 10 μg per sample was fractionated on 0.9% agarose gels and transferred (Southern, 1975) to GeneScreen Plus membranes (Du Pont-New England Nuclear). Filters were prehybridized and hybridized at 65°C in 10% dextran sulfate (w/v), 330 mM sodium phosphate (pH 7), 10 mM EDTA, 5% SDS, and calf thymus DNA at 150 μg/ml buffer with DNA probes labeled by the random primed labeling reaction of Feinberg and Vogelstein (1983) (Boehringer Mannheim). Filters were washed at 65°C in 0.3 M NaCl, 40 mM Tris (pH 7.8), 2 mM EDTA, and 0.5% SDS.

RNA Gel Blot Analysis

Total cellular RNA was isolated from *Brassica* plants using an SDS-proteinase K extraction procedure (Hall et al., 1978, Beachy et al., 1979). Approximately 1 μg of denatured RNA per sample was electrophoresed on 1% (w/v) agarose gels in the presence of formaldehyde (Maniatis et al., 1982) and transferred to nitrocellulose paper or GeneScreen (Thomas, 1980). Blots were prehybridized and hybridized as for DNA gel blots and washed at 65°C first in 0.5% SDS, 2 × SSPE and then in 0.5% SDS, 0.2 × SSPE.

Cloning of the S22 SLR1 and SLG cDNA Genes

Poly(A)$^{+}$ RNA isolated from stigmas of an S22-homozygous *B. oleracea* line was used as a template for a cDNA library constructed in λgt10 using EcoRI linkers (Gubler and Hoffman, 1983) and the Gigapack in vitro packaging system (Stratagene, San Diego, CA). Plaques from this library were hybridized (Maniatis et al., 1982) with random primer labeled (Feinberg and Vogelstein, 1983) SLR1 and SLG cDNA probes to identify SLR1 and SLG cDNA clones, respectively. DNA from recombinant SLR1- and SLG-homologous clones was analyzed by restriction enzyme analysis.

Cloning of the S6, S13, and S22 SLR1 Genomic Copies

DNA isolated from S6 and S13 homozygous *B. oleracea* lines was partially digested with Sau3A, and 10- to 20-kb fragments

were ligated into BamHI-digested EMBL4 DNA as described previously (Nasrallah et al., 1988). Plaques from these libraries were hybridized with an SLG cDNA probe essentially as described previously (Nasrallah et al., 1988). Among the 77 S6 and 22 S13 SLG-homologous clones identified, one S6 clone and one S13 clone were shown to hybridize strongly to the SLR1 cDNA probe.

DNA isolated from an S22-homozygous *B. oleracea* strain was digested to completion with EcoRI and fractionated on a 0.9% agarose gel. DNA from a duplicate lane was transferred to GeneScreen membrane and hybridized to a random primer labeled SLR1 cDNA probe to identify the region of the gel containing the SLR1 gene. DNA approximately 8 kb in length was eluted from the gel using Gene Clean glass beads and ligated into the EcoRI site of the vector λGem11 (Promega). Plaques from this library were hybridized with an SLR1 cDNA probe. One SLR1-homologous recombinant clone was characterized further. It was shown, by restriction enzyme mapping, to contain a region colinear with the S22 SLR1 cDNA gene.

Sequencing cDNA and Genomic SLR1 and SLG Clones

Gene inserts from recombinant phage were subcloned into Bluescript plasmids (Stratagene). Detailed restriction maps were created for these plasmids and overlapping restriction fragments were subcloned into the m13 sequencing vectors mp18 and mp19 (Yanisch-Perron et al., 1985). The DNA sequence of both strands for the entire length of these gene inserts was determined by the dideoxynucleotide chain termination method (Sanger et al., 1977) using the Kilobase (Bethesda Research Laboratories) or Sequenase (U.S. Biochemical, Cleveland, OH) sequencing systems.

ACKNOWLEDGMENTS

This work was supported by U.S. Department of Agriculture grant 85-CRCR-1-1600. B. Lalonde and K. Dwyer were recipients of McKnight Foundation postdoctoral fellowships, C-H. Chen of an American Cancer Society postdoctoral fellowship, and B. Barlow of a McKnight Foundation predoctoral fellowship.

Received December 2, 1988.

REFERENCES

Beachy, R.N., Barton, K.A., Madison, J.T., Thompson, J.F., and Jarvis, N. (1979). The mRNAs that code for soybean seed proteins. In Genome Organization and Expression in Higher Plants. C. Leaver, ed (New York: Plenum Publishing Corp.), pp. 273–281.

Bingham, P.M., Levis, R., and Rubin, G. M. (1981). Cloning of DNA sequences from the white locus of *D. melanogaster* by a novel and general method. Cell **25,** 693–704.

Bohlmann, H., Clausen, S., Behnke, S., Giese, H., Hiller, C., Reimann-Philipp, U., Schrader, G., Barkholt, V., and Apel, K. (1988). Leaf-specific thionins of barley—A novel class of cell wall proteins toxic to plant-pathogenic fungi and possibly involved in the defence mechanism of plants. EMBO J. **7,** 1559–1565.

East, E.M. (1940). The distribution of self-sterility in flowering plants. Proc. Am. Philos. Soc. **82,** 449–518.

Evans, R.M., and Hollenberg, S. M. (1988). Zinc fingers: Gilt by association. Cell **52,** 1–3.

Feinberg, A.P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132,** 6–13.

Gubler, U., and Hoffman, B.J. (1983). A simple and very efficient method for generating cDNA libraries. Gene (Amst.) **25,** 263–269.

Hall, T.C., Ma, Y., Buchbinder, B.U., Pyne, J.W., Sun, S.M., and Bliss, F.A. (1978). Messenger RNA for G1 protein of French bean seeds: Cell-free translation and product characterization. Proc. Natl. Acad. Sci. USA **75,** 3196–3200.

Heslop-Harrison, J.(1975). Incompatibility and the pollen-stigma interaction. Annu. Rev. Plant Physiol. **26,** 403–425.

Hinata, K., Okazaki, K., and Nishio, T. (1983). Analysis of self-compatibility in *Brassica compestris* var. *yellow sarson* (A case of recessive epistatic modifier). Proc. 6th Intern Rapeseed Conf. Paris. **1,** 354–359.

Lewis, D. (1979). Genetic versatility of incompatibility in plants. N. Z. J. Bot. **17,** 637–644.

Linskens, H.F., and Esser, K. (1957). Uber eine specifische Anfarbung den Pollenschlauche im Griffel und die Zahl der kallosepfropfen nach Selbstung und Fremdung. Naturwissenschaften **44,** 1–2.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).

Martin, F.W. (1959). Staining and observing pollen tubes in the style by means of fluorescence. Stain Technol. **34,** 125–128.

Nasrallah, J.B., and Nasrallah, M.E. (1984). Electrophoretic heterogeneity exhibited by the S-allele specific glycoproteins of *Brassica*. Experientia (Basel) **40,** 279–281.

Nasrallah, M.E., and Nasrallah, J.B. (1986). Molecular biology of self-incompatibility in plants. Trends Genet. **2,** 239–244.

Nasrallah, J.B., Kao, T.-H., Goldberg, M.L., and Nasrallah, M.E. (1985a). A cDNA clone encoding an S-locus specific glycoprotein from *Brassica oleracea*. Nature (Lond). **318,** 263–267.

Nasrallah, J.B., Doney, R.C., and Nasrallah, M.E. (1985b). Biosynthesis of glycoproteins involved in the pollen-stigma interaction of incompatibility in developing flowers of *Brassica oleracea* L. Planta (Berl.) **165,** 100–107.

Nasrallah, J.B., Kao, T.-H., Chen, C.-H., Goldberg, M.L., and Nasrallah, M.E. (1987). Amino-acid sequence of glycoproteins encoded by three alleles of the S-locus of *Brassica oleracea*. Nature (Lond.) **326,** 617–619.

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.

Ockendon, D.J. (1972). Pollen tube growth and the site of the incompatibility reaction in *Brassica oleracea*. New Phytol. **71,** 519–522.

Ockendon, D.J. (1982). An S-allele survey of cabbage (*Brassica oleracea* var. *capitata*). Euphytica **31,** 325–331.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74,** 5463–5467.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98,** 503–517.

Takayama, S., Isogai, A., Tsukamoto, C., Ueda, Y., Hinata, K., Okazaki, K., and Suzuki, A. (1986). Isolation and some characterization of S-locus-specific glycoproteins associated with self-incompatibility in *Brassica campestris*. Agric. Biol. Chem. **50,** 1365–1367.

Thomas, P.S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77,** 5201–5205.

Thompson, K.F. (1957). Self-incompatibility in marrow-stem kale, *Brassica oleracea* var. *acephala*. I. Demonstration of sporophytic system. J Genet. **55,** 45–60.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene (Amst.) **33,** 103–119.