The S 15 Self-Incompatibility Haplotype in Brassica oleracea Includes Three S Gene Family Members Expressed in Stigmas

Didier Cabrillac, Valérie Delorme, Jérôme Garin, Véronique Ruffio-Châble, Jean-Loïc Giranton, Christian Dumas, Thierry Gaude, J. Mark Cock

The S_{15} Self-incompatibility Haplotype in *Brassica oleracea* Includes Three *S* Gene Family Members Expressed in Stigmas

Didier Cabrillac[a], Valérie Delorme[a,1], Jerome Garin[b], Véronique Ruffio-Châble[c], Jean-Loïc Giranton[a], Christian Dumas[a], Thierry Gaude[a], J. Mark Cock[a, 2]

[a] Reproduction et Développement des Plantes, UMR 9938 CNRS-INRA-ENSL, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France

[b] CEA/Grenoble, Laboratoire de Chimie des Protéines, 17 rue des Martyrs, 38054 Grenoble Cedex, France

[c] Amélioration des Plantes, Institut National de la Recherche Agronomique - Domaine de la Motte, BP29, 35650 Le Rheu Cedex, France

[1] Current address: Physiologie et Biologie Moléculaire des Plantes, Université de Perpignan, Faculte de Sciences, Bat. C, 52 avenue de Villeneuve, 66860 Perpignan Cedex, France.

RUNNING TITLE: A third stigma self-incompatibility gene

[2] To whom correspondence should be addressed. E-mail: Mark.Cock@ens-lyon.fr; fax 33-4-72-72-86-00.

ABSTRACT

Self-incompatibility in Brassica is controlled by a single, highly polymorphic locus which extends over several hundred kilobases and includes several expressed genes. Two stigma proteins, the *S* locus receptor kinase (SRK) and the *S* locus glycoprotein (SLG), are encoded by genes located at the *S* locus and are thought to be involved in the recognition of self-pollen by the stigma. We report here that two different *SLG* genes, *SLGA* and *SLGB*, are located at the *S* locus in the class II, pollen recessive S_{15} haplotype. Both genes are interrupted by a single intron but *SLGA* encodes both soluble and membrane-anchored forms of SLG whereas *SLGB* only encodes soluble SLG proteins. Thus, including *SRK*, the *S* locus in the S_{15} haplotype contains at least three members of the *S* gene family. The protein products of these three genes have been characterized and each SLG glycoform was assigned to an *SLG* gene. Evidence is presented that the S_2 and S_5 haplotypes carry only one or the other of the *SLG* genes indicating either that they are redundant or that they are not required for the self-incompatibility response.

INTRODUCTION

Brassica spp possess hermaphroditic flowers with functional male and female reproductive organs in close proximity. Auto-fertilization is, however, rare in the majority of the members of this genus because of the widespread occurrence of a genetically-controlled self-incompatibility (SI) system which allows recognition and rejection of self-pollen by the stigma (reviewed in Nasrallah et al., 1994b; McCormick, 1998). SI in Brassica is controlled

by a single, highly polymorphic locus, the *S* locus. The *S* locus has been shown to have a complex structure; a high level of polymorphism extends over several hundred kilobases and several different genes are located in this region of the genome.

Two of the genes that have been identified at the *S* locus, *SLG* (for *S* locus glycoprotein; Nasrallah et al., 1987) and *SRK* (for *S* locus receptor kinase; Stein et al., 1991), are thought to be involved in the first step of the SI response, the recognition of self-pollen. SRK is an integral membrane glycoprotein that closely resembles animal receptor protein kinases (Stein et al., 1991; Delorme et al., 1995b; Stein et al., 1996). It possess an extracellular receptor domain, a single membrane-spanning domain, and a cytosolic kinase domain. SLG is a secreted glycoprotein that closely resembles the extracellular domain of SRK and this similarity is often particularly marked within haplotype(s), indicating that there has been convergent evolution between linked *SLG* and *SRK* alleles (Stein et al., 1991; Goring et al., 1993; Kusaba et al., 1997). The *SLG* gene also shares similarity with a number of other sequences both in *Brassica* spp and in other species and these genes are referred to collectively as the *S* gene family. A member of the *S* gene family is defined as a gene which encodes a protein that possesses an *S* domain with sequence similarity to the *S* locus glycoprotein.

Both *SLG* and *SRK* are expressed specifically in stigmas in a developmentally regulated manner. The accumulation of *SLG* and *SRK* transcripts and proteins coincides with the acquisition of SI by the maturing stigma (Sato et al., 1991; Stein et al., 1996). The only other organ in which *SLG* and *SRK* transcripts have been detected is the anther, but transcripts of both genes are present at very low levels and the protein products of these genes have not been detected in this organ (Sato et al., 1991; Stein et al., 1991; Delorme et al., 1995b; Stein et al., 1996). Moreover, recent evidence indicates that *SRK* transcripts in anthers are copied from the noncoding strand of the gene (Cock et al., 1997). Both *SLG* and *SRK* are highly polymorphic and the alleles of both genes can be grouped into two classes based on the degree of sequence divergence. These two classes correlate closely with the two classes of *S*

haplotype that have been defined based on their associated SI phenotype. Class I *S* haplotypes possess highly polymorphic *SLG* and *SRK* genes, confer a strong SI phenotype and tend to be dominant in interactions with other haplotypes. *SLG* and *SRK* genes of class II haplotypes, on the other hand, are less polymorphic and these haplotypes confer a weaker SI response and tend to be recessive to the class I haplotypes.

There is therefore a great deal of circumstantial evidence implicating SLG and SRK in the SI response but, despite a significant effort in several different laboratories, direct evidence linking these genes to the SI response has been difficult to obtain. A reduction in *SLG* expression due to antisense inhibition or cosuppression has been correlated with the loss of the SI response (Toriyama et al., 1991; Shiba et al., 1995) but pleiotropic effects on *SRK* were not ruled out in these experiments. Whilst these observations suggested a role for SLG in the SI response, another study showed that the level of expression of SLG was not correlated with the strength of the SI response and indicated that, if SLG has a role in the SI response, it must be capable of carrying out its function even when expressed at a low level (Gaude et al., 1995).

Several studies have demonstrated a correlation between reduced *SRK* expression and an impaired SI response (Goring et al., 1993; Nasrallah et al., 1994a; Conner et al., 1997). In two instances, the presence of mutant *SRK* alleles has been correlated with a self-compatible phenotype (Goring et al., 1993; Nasrallah et al., 1994a), although in both cases it is possible that other genes were also affected. The recent observation that a mutant *SRK* gene induced partial self compatibility when transformed into a self-incompatible Brassica napus line strongly implicates SRK in the SI response, especially as the effect of the transgene was haplotype-specific (Stahl et al., 1998).

Based on the resemblance of SRK to animal receptor protein kinases, current models of self-pollen recognition in the SI response propose that SRK recognizes a pollen-borne ligand. Using a bioassay, Stephenson et al. (1997) have recently shown that the pollen coat of self pollen carries a factor which can initiate the SI response. The gene encoding the male

factor has not yet been identified. One possible candidate, the *S* locus anther (*SLA*) gene, an *S* locus linked gene expressed specifically in anthers (Boyes and Nasrallah, 1995), has recently been shown not to be required for the SI response (Pastuglia et al., 1997b).

In the female part of the flower, *S* locus genes encode a complex mixture of proteins in addition to SLG and SRK. Tantikanjana et al. (1993) showed that the class II SLG_2 allele possesses two exons and that alternative transcripts encode SLG and a membrane-anchored form of SLG, mSLG. mSLG has not been found in stigmas of class I haplotypes (class I *SLG* genes lack the second exon), and it has been suggested that this protein modulates the SI response and is responsible for the recessive phenotype associated with the class II haplotypes (Tantikanjana et al., 1993). Similarly, the class I SRK_3 allele has been shown to encode, in addition to an integral SRK protein, a soluble truncated protein, eSRK, which corresponds to the SRK extracellular domain (Giranton et al., 1995). eSRK is also thought to play a role in modulating the SI response.

In this study, we demonstrate that the class II S_{15} haplotype carried by the *B. oleracea* line P57Sc includes alleles of three different stigma-expressed genes; *SRK* and two different *SLG* genes, *SLGA* and *SLGB*. Evidence is also presented that the S_5 haplotype contains *SLGB* but not *SLGA* whereas the S_2 haplotype contains *SLGA* but not *SLGB*. These data raise questions both about the functional roles SLG and mSLG and about how sequences at the *S* locus have evolved.

RESULTS

Identification of a third *S* gene family member at the *S* locus in the *B. oleracea* line P57Sc (haplotype S_{15})

The self-compatible *B. oleracea* line P57Sc described by Gaude et al. (1993) was shown to

reject pollen from an S_{15} homozygous tester line indicating that it carries the S_{15} haplotype and that the SI system is functional on the female side (see Methods). A 116 kD protein, which is thought to correspond to SRK_{15}, has been identified in stigma extracts of the P57Sc line (Delorme et al., 1995b). To further characterize this protein, we attempted to identify the corresponding gene.

Alleles corresponding to two *S* gene family members have been shown to be linked to the *S* locus in the P57Sc line. The first, SLG_{Sc} is predicted to encode both a secreted and a membrane-anchored glycoprotein (Pastuglia et al., 1997b). SLG_{Sc} is unlikely to also encode an SRK protein (with a cytosolic kinase domain) because this gene is located directly upstream of another *S* locus gene, *SLA* (Pastuglia et al., 1997b; see Figure 1). For the second gene, a cDNA (designated CG15) was isolated which also encoded a secreted glycoprotein (Gaude et al., 1993). One possibility was that this cDNA corresponded to an alternative transcript of *SRK*. An analogous situation has been described for the S_3 haplotype in which SRK_3 has been shown to produce alternative transcripts that encode a truncated form of the SRK protein which resembles the SLG protein of the same haplotype (Giranton et al., 1995). To determine whether the gene which encoded CG15 also encodes an SRK protein we isolated the corresponding region of genomic DNA and sequenced a region of 2.8 kb downstream of the region corresponding to CG15 (Figure 1). No similarity with the transmembrane or kinase domains of other *SRK* alleles was found within the 2.8 kb region suggesting that the gene corresponding to CG15 does not encode an SRK protein. However, we could not rule out the possibility that the *S* domain and transmembrane domain were separated by a large intron which spanned the sequenced region.

As an alternative approach to identify the *SRK* allele of the P57Sc line, oligonucleotides were designed to amplify *SRK* cDNA using the the polymerase chain reaction (PCR). The nucleotide sequences of the kinase domains of two class II *SRK* alleles, SRK_5 (see Methods) and SRK_2 (JMC, unpublished data), were aligned with the corresponding regions of the closely related genes *SLR3* (Cock et al., 1995), *SFRI* (accession number

Y14285), *SFR2* (Pastuglia et al., 1997a) and *SFR3* (accession number Y14286). Sequences conserved in the class II *SRK* alleles but diverged or absent in the other genes were selected for oligonucleotide synthesis. These oligonucleotides were used either in combination with PS3, an oligonucleotide that hybridizes near the ATG initiation codon of several class II *S* genes (Nishio et al., 1996), to amplify 5' sequences or with a oligo(dT) adapter in a rapid amplification of cDNA ends (RACE)-PCR protocol (Frohman et al., 1988) to amplify 3' sequences from stigma cDNA. These experiments allowed the reconstruction of a cDNA of 2883 bp. The cDNA is predicted to encode a polypeptide which consists of a signal peptide, an *S* domain, a membrane-spanning domain and a kinase domain. The predicted mass of the mature polypeptide is 94.0 kD. The *S* domain region of the amplified cDNA was highly similar to SLG_{Sc} and CG15 (89.5% and 87.3% respectively, Table 1) but was not identical, indicating that the cDNA was derived from a third gene.

Because P57Sc carries the S_{15} haplotype, the gene corresponding to the amplified cDNA was designated as SRK_{15}, SLG_{Sc} was renamed $SLGA_{15}$ and the allele corresponding to CG15 was designated as $SLGB_{15}$ (Figure 1). To confirm that the P57Sc line carried the S_{15} haplotype, part of the *S* domain of each of the three genes was amplified from genomic DNA of the S_{15} tester line. DNA sequence analysis of the amplified products showed that they were identical to the corresponding regions of the three genes carried by P57Sc (apart from a number of errors introduced during the PCR, each of which was only found in one of several clones analysed for each gene; data not shown).

Comparison of $SLGA_{15}$ and SRK_{15} revealed a region of 502 bp with 100% sequence identity (Figure 1). The presence of this region of identity indicates that there has been concerted evolution between the two genes probably involving gene conversion. A similar region was not found when SLG_2 and SRK_2 were compared indicating that gene conversion occurred after the divergence of these two *S* haplotypes.

Genetic mapping and expression patterns of SRK_{15}, $SLGA_{15}$ and $SLGB_{15}$

To demonstrate that *SRK*$_{15}$, *SLGA*$_{15}$ and *SLGB*$_{15}$ were linked to the *S* locus, we used the PCR to analyse their segregation in an F_2 population resulting from a cross between the P57Sc line (homozygous for the S_{15} haplotype) and a line homozygous for the S_{3a} haplotype. Figure 2A shows that *SRK*$_{15}$, *SLGA*$_{15}$ and *SLGB*$_{15}$ sequences were only amplified from genomic DNA of F_2 progeny carrying the S_{15} haplotype, indicating that all three genes are linked to the *S* locus.

The expression patterns of *SLGA*$_{15}$ and *SLGB*$_{15}$ were analysed by ribonuclease protection analysis of mRNA abundance in a range of different organs. Figure 2B shows that *SLGA*$_{15}$ and *SLGB*$_{15}$ transcripts were detected only in stigma RNA. No transcripts were detected in the other organs tested even after a long exposure. Analysis of autoradiographs by densitometry indicated that *SLGA*$_{15}$ transcripts were approximately four fold less abundant than *SLGB*$_{15}$ transcripts. Reverse transcriptase (RT)-PCR was used to analyse the organ-specific expression pattern of *SRK*$_{15}$. Figure 2C shows that *SRK*$_{15}$ transcripts were detected only in stigmas and, at a much lower abundance, in anthers at the trinucleate pollen stage.

Analysis of *S* locus-encoded glycoproteins in stigmas of Brassica plants homozygous for the S_{15} haplotype

Comparison of the sequences of SRK_{15}, $SLGA_{15}$ and $SLGB_{15}$ indicated that these three genes encoded mature protein products with identical amino-termini. The first 10 amino acid residues of each protein are predicted to be identical to the synthetic peptide that was used to raise the monoclonal antibody MAb 157-35-50 (Giranton et al., 1995) and this antibody should therefore recognize the protein products of all three genes. MAb 157-35-50 was used to detect proteins in stigma extracts from plants homozygous for five different *S* haplotypes, including S_{15}, following separation of the proteins by SDS-PAGE. Figure 3A shows that MAb 157-35-50 detected both a large protein of between 102 and 120 kD, depending on the *S* haplotype, and a number of smaller proteins of between 43 and 68 kD in stigmas of four of the five lines.

The molecular masses and polymorphic nature of the 102 to 120 kD proteins was consistent with their corresponding to the SRK proteins of each *S* haplotype. This was supported by the fact that, following deglycosylation, there was a shift in the mobilities of these proteins and a polypeptide of 94 kD was detected in extracts from lines homozygous for S_2, S_{15} and S_3 (Figure 3B). The method was not sensitive enough to detect deglycosylated SRK_5. The molecular mass of the deglycosylated proteins corresponded to the predicted molecular mass of the polypeptide backbone of SRK. The mobility of the protein detected in stigma extracts of the S_{15} homozygous line indicated a molecular mass of 105 kD, slightly lighter than the previous estimate of 116 kD (Delorme et al., 1995b). Further evidence that this protein corresponds to SRK_{15} was obtained by analysing 24 plants of the F_2 population segregating the S_{15} and S_{3a} haplotypes (Figure 3C). The 105 kD protein was expressed only in plants which possessed the S_{15} haplotype and is therefore encoded by a gene linked to the *S* locus in the S_{15} haplotype. This result was confirmed by analysing additional F_2 populations (data not shown).

In addition to the larger SRK proteins, MAb 157-35-50 detected smaller proteins of between 43 and 68 kD in the various *S* haplotypes (Figure 3A). Several of these proteins, particularly those from S_3 and S_2 have been characterized previously (Giranton et al., 1995; Tantikanjana et al., 1993). The 59.4 and 69.8 kD proteins detected in the stigma extract from the S_3 line are glycoforms of a soluble, truncated form of SRK_3 ($eSRK_3$) which have a polypeptide backbone of 46.8 kD (Giranton et al., 1995; Figure 3B). SLG_3 is not detected by MAb 157-35-50. On the other hand, the 55 to 68 kD proteins detected in the S_2 line have previously been shown to include both the soluble SLG_2 protein and a membrane anchored form of SLG_2 called $mSLG_2$ (Tantikanjana et al., 1993). This is consistent with the fact that proteins of 52 and 48 kD were detected in deglycosylated S_2 stigma extracts, presumably corresponding to the polypeptide backbones of $mSLG_2$ and SLG_2 respectively.

The pattern of bands detected by MAb 157-35-50 in stigma extracts of the S_{15} line was more similar to that observed with the S_2 line than with the S_3 line. The mobilities of the proteins of between 54 and 65 kD was altered following deglycosylation to give two proteins of 48 and 52 kD (Figure 3B). By analogy to S_2, we suggest that these correspond to soluble and membrane anchored forms of *SLG*, respectively. Interestingly, following deglycosylation of stigma proteins from the S_5 homozygous line a single major protein of 48 kD was detected with MAb 157-35-50 (the size predicted for soluble SLG protein) suggesting that stigmas of this line do not contain significant levels of mSLG (Figure 3B).

The above interpretation of the results presented in Figures 3A and 3B predicts that the 105 kD protein corresponds to SRK and is anchored in the membrane whereas the smaller 54 to 65 kD proteins include both soluble and membrane-anchored proteins. To investigate further the sub-cellular location of the *S* locus-encoded proteins of the S_{15} homozygous line, we separated stigma extracts into soluble and membrane fractions and then further fractionated the membrane preparation by aqueous two-phase partitioning to enrich for plasma membrane. An antibody against the plasma membrane-localised ATPase was used to verify that the fractions had been enriched in plasma membrane (data not shown). MAb 157-

35-50 was then used to detect *S* locus-encoded proteins in the different fractions. Figure 3D shows that the 105 kD protein was most abundant in the fraction enriched in plasma membrane. The group of proteins of between 54 and 65 kD could be separated into, on the one hand, a number of proteins of between 54 and 65 kD, which partitioned into the soluble fraction, and, on the other, a major protein of 65 kD, which was associated with the plasma membrane (Figure 3D). These results are consistent with the 105 kD protein corresponding to a plasma membrane-anchored SRK, the 65 kD protein corresponding to plasma membrane-anchored mSLG and the 54 to 65 kD proteins being different soluble glycoforms of SLG.

Analysis of the DNA sequences of SRK_{15}, $SLGA_{15}$ and $SLGB_{15}$ indicated that the soluble glycoproteins detected by MAb 157-35-50 in stigma extracts of the S_{15} homozygous line could have been encoded by any of the three genes. These soluble glycoproteins can be resolved by isoelectric focusing (IEF) gel electrophoresis into four major glycoproteins which have been designated α,_β,_β' and β" (Gaude et al., 1993). Figure 4 shows that these glycoproteins vary in pI from 6.1 to 7.3 and in relative abundance from 16.5% to 41.5%. To characterize them further, they were each purified by selective, two dimensional gel electrophoresis (Gaude et al., 1991). The purified proteins were then digested with either endopepidase Lys-C or endopeptidase Gluc-C and HPLC-purified peptides were sequenced by Edman degradation. The protein sequence data indicated that α_and_β were encoded by $SLGB_{15}$ (Figure 4). No peptide sequence data was obtained for β" and the sequence obtained for_β' was conserved in SRK_{15}, $SLGA_{15}$ and $SLGB_{15}$. In order to confirm these results and to assign a gene to the β' and β" proteins, the IEF-resolved α,_β,_β' and β" proteins were digested 'in-gel' with trypsin and the tryptic peptides analysed by matrix-assisted laser desorption ionisation time-of-flight mass spectroscopy (MALDI-TOF-MS). This analysis demonstrated that α,_β and_β' were encoded by $SLGB_{15}$ and that β" was encoded by $SLGA_{15}$ (Table 2). None of the four proteins were encoded by SRK_{15}.

$SLGB_{15}$ encodes 79.5% of the SLG protein in the stigma, only 20.5% being encoded

by *SLGA*$_{15}$ (Figure 4). This is consistent with the fact that *SLGB*$_{15}$ transcripts are approximately 4 fold more abundant than those of *SLGA*$_{15}$ (Figure 2), bearing in mind that a proportion of the *SLGA*$_{15}$ transcripts presumably encode an mSLG protein. Expression of these two genes therefore appears to be regulated principally at the mRNA level.

Transcripts of the *S* locus genes of the S_{15} haplotype

Based on the above analysis of proteins detected in stigma extracts from the S_{15} homozygous line, SRK_{15} is predicted to encode a membrane-anchored, 105 kD protein (SRK), $SLGA_{15}$ to encode both a soluble protein (SLG β") and a membrane-anchored protein of 65 kD (mSLG) and $SLGB_{15}$ has been shown to encode several soluble proteins (SLGs α,_β and_β'). If this is the case then each gene would be expected to produce transcripts capable of encoding the corresponding protein products. This has been demonstrated for SRK_{15} and $SLGB_{15}$, which were both initially identified as cDNA clones which encode SRK and SLG proteins, respectively (Figure 5). However, $SLGA_{15}$ was identified as a genomic clone and $SLGA_{15}$ transcription has not yet been analysed. Therefore, we used RACE-PCR to characterize transcripts of this gene. Figure 5 shows that two different transcripts were identified, one which terminates within the single intron and therefore is predicted to encode a secreted SLG protein and a second which includes both exon 1 and exon 2 and hence is predicted to encode a membrane-anchored mSLG protein. Therefore, analysis of RNA transcripts of the SRK_{15}, $SLGA_{15}$ and $SLGB_{15}$ indicates that these three genes potentially encode all the proteins detected in stigma extracts with the MAb 157-35-50 antibody.

Interestingly, none of the soluble, *S*-locus-encoded proteins present in S_{15} stigma extracts corresponded to a truncated form of the SRK protein (Figure 4). This contrasts with the situation in the S_3 haplotype where significant levels of such a protein, eSRK, have been detected (Giranton et al., 1995). eSRK is encoded by transcripts which retain sequences corresponding to the first intron of the *SRK* gene with the result that a stop codon occurs at the end of the reading frame encoding the *S* domain (Giranton et al., 1995). This stop codon is found in a conserved position just downstream of the 5' splice site of intron 1 in all the *SRK* alleles that have been analysed to date. We isolated the first intron of SRK_{15} by PCR amplification from genomic DNA and sequence analysis showed that the stop codon is also present in SRK_{15} (data not shown). Furthermore, RACE-PCR analysis identified short SRK_{15}

transcripts which possessed sequences corresponding to the 5' end of intron 1 and which are therefore predicted to encode an eSRK protein (Figure 5). The absence of an eSRK protein amongst the soluble S_{15} stigma proteins suggests, however, that these transcripts are not translated or that they are only present in very low abundance.

Comparison of the sequences of the S_{15} alleles of *SLGA*, *SLGB* and *SRK* with those of other haplotypes.

The deduced amino acid sequences of the *S* domains of SRK_{15}, $SLGA_{15}$ and $SLGB_{15}$ were compared with SRK and SLG proteins encoded by other class I and class II *S* haplotypes and with two proteins, SLR2 and SFR2, which are also encoded by members of the *S* gene family in Brassica but whose genes are not linked to the *S* locus. Table 1 shows that $SLGA_{15}$ is most similar to SLG_2 whereas $SLGB_{15}$ is most similar to SLG_5. This confirms a previous observation that, whilst there is evidence for convergent evolution of *SLG* and *SRK* alleles within a particular *S* haplotype (Stein et al., 1991), alleles within a haplotype are not necessarily more similar to each other than to alleles from other haplotypes (Kusaba et al., 1997).

The similarities between $SLGA_{15}$ and SLG_2 and between $SLGB_{15}$ and SLG_5 are interesting because these two pairs of genes share similar structures. Both $SLGA_{15}$ and SLG_2 possess a single intron and are capable of encoding both soluble and membrane-anchored forms of SLG (Figure 5; Tantikanjana et al., 1993). Comparison of the sequences of the introns of these two alleles (Tantikanjana et al., 1993; Pastuglia et al., 1995b) revealed that they were 95.9% similar (data not shown). As for $SLGB_{15}$, it also possesses an intron but does not encode a membrane-anchored form of SLG (Figure 1). The predicted C-terminus of SLG_5 (-threonine-cysteine-threonine-glycine-COOH) is identical to the C-terminus of $SLGB_{15}$ which is encoded by the second exon. We showed, by PCR amplification from genomic DNA, that the last four residues of SLG_5 are also encoded by a second exon. Comparison of

the nucleotide sequence of the intron of SLG_5 (accession number Y18256) showed that it was 97.7 % similar to that of $SLGB_{15}$ (data not shown). Hatakeyama et al. (1998) have recently described an *SLG* allele with a similar structure to $SLGB_{15}$ and SLG_5 in the class II S_{29} haplotype of *B. rapa*. Although there are significant differences between the introns of $SLGB_{15}$ and *B. rapa* SLG_{29} (for instance, the intron of *B. rapa* SLG_{29} lacks the direct repeat found in the $SLGB_{15}$ intron, Figure 7), they share extensive regions of sequence similarity indicating that they are diverged alleles of the same genes. Taken together, these observations indicate that SLG_2 is an allele of the *SLGA* gene whereas SLG_5 and *B. rapa* SLG_{29} are alleles of *SLGB*. Thus alleles of *SLGA* and *SLGB* are present in other class II *S* haplotypes.

The deduced amino acid sequences of different alleles of *SLG* and *SRK* exhibit a high level of polymorphism but polymorphic residues are not distributed evenly along the gene sequence. In particular, two hypervariable regions have been identified (residues 217 to 224 and 273 to 292 in $SLGA_{15}$) and it has been suggested that these hypervariable regions may be involved in determining the identity of each allele (Nasrallah et al., 1987). In a survey of class I *SLG* alleles from *B. oleracea* and *B. campestris*, Kusaba et al. (1997) identified two pairs of alleles that shared identical hypervariable regions. One pair included an allele from each species, and may represent alleles with the same SI-specificity in the two species but the second pair were both from *B. oleracea* and were derived from two genetically distinct *S* haplotypes, S_8 and S_{46}. Comparison of $SLGA_{15}$ with SLG_2 showed that these two alleles are even more similar than the SLG_8/SLG_{46} pair, with only five amino acid differences between the mature proteins (compared with nine for SLG_8/SLG_{46}), none of which are located in the hypervariable regions (Figure 6). Moreover, none of the polymorphisms in the $SLGA_{15}/SLG_2$ pair are predicted to alter sites of N-linked glycosylation. These observations suggest that, if SLG plays a role in the recognition step of the SI response, the hypervariable domains are not always involved in determining allele-specificity.

Intron structure and evolution of *S* locus genes in the S_{15} haplotype

To better understand the evolutionary relationship between the three *S* locus genes, the DNA sequences of the first intron of SRK_{15} and the introns of $SLGA_{15}$ and $SLGB_{15}$ were compared. Figure 7A shows that the intron of each allele contained regions that were highly similar to sequences in introns of each the two other alleles. However, none of the three introns contained sequences which were significantly similar to both of the other introns. Figure 7B shows two possible models for the evolution of these intron sequences based on these similarities. In the first model, three copies of an ancestral gene would have been generated by triplication and each of the three copies would then have undergone a specific deletion of intron sequences to generate the intron sequences of SRK_{15}, $SLGA_{15}$ and $SLGB_{15}$. In the second model, multiple copies of an ancestral gene generated by gene duplication would have undergone a series of recombination events involving intron sequences to generate the patchwork of sequences shared between the three genes. In both cases, only major steps in the evolution of the sequences are shown, additional mutations would have lead to the divergence of the sequences from one another.

The high level of sequence similarity between the introns of $SLGA_{15}$ and SLG_2 and between the those of $SLGB_{15}$ and SLG_5 suggests that the major structural rearrangements of the introns indicated in Figure 7B would have occurred before the divergence of these class II *S* alleles. This suggests a relatively ancient origin for the intron sequences and we were interested in determining whether the introns of SRK_{15}, $SLGA_{15}$ and $SLGB_{15}$ resembled those of other members of the *S* gene family in Brassica. The *S* domain of *SFR2* shares 55.5% similarity with that of $SLGB_{15}$ at the amino acid level (Table 1). We isolated the first intron from *SFR2* (accession number Y18257) and compared its sequence with those of SRK_{15}, $SLGA_{15}$ and $SLGB_{15}$. No significant similarity was detected (data not shown). The *S* domain of *SLR2* shares greater similarity with those of the *S* locus genes and is most similar to *SLGB* (85.6% amino acid similarity; Table 1). *SLR2* is also similar to *SLGB* at the structural level, both genes contain an intron and the C-terminal end of SLR2 (glycine-threonine-glycine-

COOH) is encoded by a second exon (Tantikanjana et al., 1996). These similarities suggest that *SLR2* may have diverged from *SLGB* after the divergence of the three *S* locus genes and it will be interesting to compare the introns of these two genes to test this hypothesis.

Do other *S* haplotypes contain three *S* gene family members?

The above data indicated that the S_{15} haplotype includes three members of the *S* gene family. We were interested in determining whether this was also the case for other *S* haplotypes. A pair of oligonucleotides corresponding to conserved sequences within the *S* domains of the three S_{15} alleles were therefore used to amplify sequences from cDNA and genomic DNA of *B. oleracea* lines homozygous for two other class II *S* haplotypes, S_5 and S_2. No novel SLG sequences were detected using this approach (data not shown). To investigate this question further, fragments from the introns of $SLGA_{15}$ and $SLGB_{15}$ were used as gene-specific probes to detect alleles of these two genes in blots of genomic DNA. Figure 8 shows that a probe from intron 1 of $SLGA_{15}$ hybridized to a single fragment in BamHI and EarI restriction digested DNA from S_{15} and S_2 homozygous lines. A similar result was obtained with a number of different restriction enzymes and the fragments were identical size in each case, confirming the observation (see above) that $SLGB_{15}$ and SLG_2 share a high degree of similarity of both exon and intron sequences. No corresponding fragment was detected in digests of S_5 DNA despite the fact that the S_5 alleles of the two other *S* locus genes *SLGA* (i.e. SLG_5) and *SRK* are also highly similar to those of the S_{15} haplotype. This result suggests that the *SLGA* gene has been lost from the S_5 haplotype. When stigma proteins from the S_5 homozygous line were deglycosylated and *S* locus proteins identified with MAb 157-35-50, a single major band of 48 kD, the size expected for soluble SLG, was detected (Figure 3B). In contrast, two major proteins of 48 and 52 kD, which probably correspond to SLG and mSLG respectively, were detected in extracts from the S_2 and S_{15} homozygous lines (Figure 3B). Again, these results are consistent with the S_5 haplotype lacking a functional *SLGA* allele and,

therefore, being unable to encode mSLG.

A similar experiment was carried out with a probe from the *SLGB*$_{15}$ intron. In this case a single, conserved fragment hybridized to the probe in DNA from S_{15} and S_5 DNA but no corresponding fragment was detected in S_2 DNA (Figure 8). The probe did, however, hybridize weakly to a fragment of 2.9 kb in an BamHI digest of S_5 DNA and it was not clear whether this represented hybridization to *SLGB* or to another member of the *S* gene family.We favour the latter interpretation because a weakly hybridizing fragment was also detected in DNA of plants carrying the class I S_3 haplotype. The *SLGB* gene, therefore, is probably absent from (or present but significantly rearranged in) the S_2 haplotype. Gaude et al. (1995) have shown that the S_2 homozygous line expresses a very low level of SLG whereas SLG levels are significantly higher, and comparable with those of S_{15}, in the S_5 homozygous line. Figure 4 shows that, in the S_{15} haplotype, *SLGA* is strongly expressed and accounts for 79.5% of stigma SLG whilst *SLGB* only encodes 20.5%. The hypothesis that the S_5 and S_2 haplotypes have lost *SLGA* and *SLGB* respectively is therefore attractive because it may explain the significant difference in levels of SLG in stigmas of the two haplotypes.

DISCUSSION

Three members of the *S* gene family are linked to the *S* locus in the S_{15} haplotype

The results presented here demonstrate that three, closely-related members of the *S* gene family, *SRK*$_{15}$, *SLGA*$_{15}$ and *SLGB*$_{15}$, are linked to the *S* locus in the S_{15} haplotype. In a similar study, Suzuki et al. (1997) showed that three members of the S gene family are closely linked to the *S* locus in the *B. campestris* S_9 haplotype. However, the genes identified in the S_9 haplotype share less than 70% similarity with *SLG*$_9$ and *SRK*$_9$ at the nucleotide level and are expressed in both vegetative and floral tissues suggesting that they form a phylogenetic group

distinct from *SLG*/*SRK* and that they have (a) different funtional role(s). Furthermore, two of the genes identified by Suzuki et al. (1997) probably encode non-functional proteins. In contrast, the three genes identified in the *B. oleracea* S_{15} haplotype share several characteristics which have been associated with a function in self-incompatibility.

SRK_{15}, $SLGA_{15}$ and $SLGB_{15}$ encode four soluble glycoproteins (glycoforms of SLG) and two membrane-anchored glycoproteins (SRK and mSLG). These proteins accumulate specifically in stigmas. Both SRK and mSLG were shown to be localised to the plasma membrane fraction in Brassica stigmas. These data are consistent with the observation that SRK is targeted to the plasma membrane when expressed ectopically in transgenic tobacco (Stein et al., 1995). Peptide sequence and MALDI-TOF-MS analysis was used to assign each of the four soluble glycoproteins to an *S* locus gene. MALDI-TOF-MS was particularly useful for discriminating between the different products of the three genes and this technique is likely to have wide applications in distinguishing between gene products of other gene families and, coupled with genome information in model species, in correlating genes and gene products in general.

Functional roles of SLG, mSLG and eSRK

Considerable heterogeneity was observed when the *S*-locus-encoded proteins present in stigmas of plants homozygous for the S_{15} haplotype were compared with those of other class II haplotypes. The abundance of soluble SLG proteins varied between haplotypes and, whilst putative mSLG proteins were detected in both S_{15} and S_2 homozygous plants, mSLG was not detected in an S_5 homozygous plant. We present evidence that these differences reflect differences in the number of *S* gene family members at the *S* locus. DNA blot analysis indicated that the S_2 haplotype contains *SRK* and *SLGA* but not *SLGB*, whereas the S_5 haplotype contains *SRK* and *SLGB* but not *SLGA*. If these two lines each lack one of the two *SLG* genes it follows that (1) neither gene is required for the SI response or (2) the two genes

are functionally redundant with respect to the SI response. Moreover, $SLGA_{15}$ and $SLGB_{15}$ are more similar to proteins encoded by closely related *S* haplotypes (for example, there are only five amino acid differences between SLG_2 and $SLGA_{15}$) than they are to each other (45 amino acid differences, Figure 6). This argues against functional redundance of *SLGA* and *SLGB* and, therefore, suggests that SLG is unlikely to be required for the recognition step of the SI response.

The absence of an *SLGA* gene in the S_5 haplotype also raises questions about the functional role of mSLG. Tantikanjana et al. (1993) showed that mSLG was present in stigmas of the class II S_2 haplotype but absent from stigmas of the class I S_6 haplotype. They suggested that mSLG may interfere with SRK function and that this may account for the recessive nature of the S_2 haplotype in heterozygotes. The absence of mSLG from stigmas of the S_5 line, which is also a recessive class II haplotype, suggests that either mSLG is not responsible for this phenotype or that other factors can also have an influence. Comparative analysis of S_5 and S_2 homozygous lines may help to elucidate the role of the mSLG protein.

The four soluble glycoproteins recognised by MAb 157-35-50 in extracts of stigmas of the S_{15} line were all attributed to one or other of the *SLG* genes. Therefore, unlike S_3 (Giranton et al., 1995), this haplotype does not express a detectable level of eSRK. This suggests that either eSRK is not required for the SI response or that it is able to function at very low abundance (below the detection level of this antibody). It is interesting that eSRK was detected in plants carrying the class I S_3 haplotype but not in plants with the class II S_{15} haplotype despite the fact that both *SRK* alleles possess a stop codon at the 5' end of the intron 1. One possibility, which merits further investigation, is that the phenotypic differences between class I and class II haplotypes may be due to the presence or absence of eSRK.

Another important observation is that $SLGA_{15}$ and SLG_2 differ at only five amino acid positions, none of which are in the hypervariable domains which have been implicated in allele-specificity. Hence, if SLG does play a role in recognition of self-pollen, the hypervariable domains do not determine allele-specificity in these haplotypes. As pointed out

by Nasrallah (1997), the high level of polymorphism of the hypervariable domains may indicate a role in recognition or may merely be due to these being regions of minor functional importance that are more free to vary than other parts of the gene.

Recombination and gene conversion between *S* locus alleles

Despite the large physical size of the *S* locus in *Brassica* spp, no recombination event between the different genes at the *S* locus has been reported. Recombination is thought to be suppressed at the *S* locus because of extensive sequence rearrangements and divergence which have resulted in structural heteromorphisms between *S* haplotypes (Boyes et al., 1997). However, there is accumulating evidence that different sequences within the *S* locus have evolved independently, at least to some extent, suggesting that recombination or gene conversion events have occurred between haplotypes during evolution (Kusaba et al., 1997; Charlesworth and Awadalla, 1998). The fact that the similarity between *SLG* and *SRK* alleles within a haplotype is often greater than between many alleles of each of these two genes (Stein et al., 1991; Goring et al., 1993; Delorme et al., 1995b; Kusaba et al., 1997) indicates that sequence information may also have been exchanged between genes within the same *S* haplotype. This hypothesis is supported by two features which were identified when *S* locus genes of the S_{15} haplotype were compared.

Firstly, comparison of intron sequences of SRK_{15}, $SLGA_{15}$ and $SLGB_{15}$ showed that they included a patchwork of conserved sequence motifs each of which was found in the introns of only two of the three genes. The structure of the introns suggests that the three genes have evolved by a series of gene duplications followed by either deletion or recombination events (Figure 7). The hypothesis that recombination may have occurred between intron sequences of *S* locus genes is interesting in the light of data presented here which indicates that the number of *S* gene family members at the *S* locus may vary between haplotypes. Recombination events, occurring for example as a result of unequal sister

chromatid exchange or intrachromasomal recombination, could lead to the loss or gain of alleles. In addition, recombination within intron sequences would provide a mechanism for the acquisition or loss of exons encoding transmembrane and kinase domains by or from genes resembling *SLG* and *SRK* respectively (Tantikanjana et al., 1993). Moreover, when kinase domains of SRKs from different haplotypes are compared they exhibit fewer nonsynonymous differences, fewer non-conservative amino acid differences and a generally lower level of variability than SRK *S* domains (Charlesworth and Awadalla, 1998). This suggests that some recombination has occurred between the *S* and kinase domains of *SRK* genes from different haplotypes. Hence sequence information may have been exchanged by recombination both between alleles of different haplotypes and between genes of the same haplotype at the *S* locus.

The second sequence of interest is a 502 nucleotide region of 100% sequence identity which is shared by SRK_{15} and $SLGA_{15}$ (Figure 1). The presence of this shared sequence indicates that the two genes have been involved in a gene conversion event. Similar regions of 100% nucleotide sequence identity are also shared by the *SLG* and *SRK* alleles of both the *B. napus* S_{A10} and the *B. rapa* S_9 haplotypes (Goring et al., 1993; Suzuki et al., 1997). The positions of the identical regions are different in each *SLG*/*SRK* pair indicating that the events occurred independently after divergence of the *S* alleles. Hence, there is evidence that gene conversion has occurred in three out of the eleven *SLG*/*SRK* haplotype pairs for which sequence data is currently available. Similar regions of sequence identity are significantly less common between *SLG* alleles (JMC, unpublished results) suggesting that gene conversion events between genes of the same haplotype have occurred more frequently than between alleles of the same *S* locus gene. In humans, homogenisation of tandemly repeated genes occurs primarily by intrachromosomal events and exchanges of sequences between chromosomes are less frequent (Liao et al., 1997). Moreover, in yeast, intrachromasomal homogenisation of tandem repeats occurs primarily as a result of gene conversion (Gangloff et al., 1996). Evolution of tandemly repeated disease resistance genes at the *Cf-4*/*9* locus in

tomato is thought to have occurred by a similar mechanism involving shuffling of sequences between genes within the locus (Parniske et al., 1997). The above evidence that there has been a higher frequency of sequence exchange between alleles of genes within the same haplotype than between haplotypes suggests that more dispersed gene clusters such as the *S* gene family members at the Brassica *S* locus may evolve in a similar manner.

The concerted evolution of *SLG* and *SRK* alleles within a particular haplotype has been put forward as evidence of a shared function for these two genes (Stein et al., 1991). Although this may be the case, it is also possible that their similarity may be merely the result of general homogenisation mechanisms operating on linked genes with sequence homology. Indeed, theoretical considerations indicate that, under certain conditions, it may be difficult for duplicated genes to avoid the homogenising effects of gene conversion (Walsh, 1987).

METHODS

Plant Material, Genetic Crosses, and Determination of Incompatibility Phenotype

The *Brassica oleracea* var *acephala* line P57Sc was shown to be self-compatible by Gaude et al. (1993) and the *S* haplotype carried by this line was therefore arbitrarily designated S_{Sc}. Genetic crosses with closely related *B. oleracea* var *botrytis* lines have indicated that the self-compatible phenotype of P57Sc is due to a defect in the pollen (Pastuglia et al., 1997b). We were therefore able to identify the *S* haplotype carried by P57Sc by measuring seed set after application of pollen from tester lines from the collection at Horticulture Research International, Wellesbourne, UK. Pollen from an S_{15} homozygous line was incompatible on P57Sc stigmas indicating that P57Sc carries the S_{15} haplotype (data not shown). This was subsequently confirmed by comparing the sequences of P57Sc *S* locus genes with partial sequences of the corresponding genes from the S_{15} tester line (see results). The S_3 homozygous line has been described (Delorme et al., 1995b), the S_2 and S_5 homozygous lines were a gift from D. J. Ockendon (Horticulture Research International, Wellesbourne, UK). An F_2 population obtained by self-fertilisation of an S_{3a}/S_{15} F_1 hybrid (the parental lines are described in Delorme et al., 1995a, and Gaude et al., 1993, respectively where they are referred to as S_{3a} and S_{Sc}) was used to demonstrate linkage of SRK_{15}, $SLGA_{15}$ and $SLGB_{15}$ to the *S* locus. The incompatibility phenotypes of the F_2 progeny were determined by self-pollination and by crosses to tester plants using previously described procedures (Delorme et al., 1995b) and aniline blue staining of pollinated pistils was carried out as previously described (Ruffio-Châble et al., 1997).

Cloning of *S* locus Gene Sequences and DNA Sequence Analysis

The isolation of a genomic clone corresponding to $SLGA_{15}$ (previously called SLG_{Sc}; Pastuglia

et al., 1997b) and a cDNA corresponding to *SLGB*$_{15}$ (CG15; Gaude et al., 1993) has been described previously. A fragment of genomic DNA carrying the *SLGB*$_{15}$ gene was isolated from a genomic library (Pastuglia et al., 1997b) constructed with DNA of the P57Sc line. Restriction endonuclease digested fragments of isolated λ clones were subcloned into pBluescript II SK$^+$ (Stratagene, La Jolla, CA) for sequence analysis.

An *SRK*$_5$ cDNA was reconstructed by RT-PCR (reverse transcriptase-PCR) amplification of overlapping fragments from cDNA prepared from stigmas of a plant homozygous for the *S*$_5$ haplotype. Oligonucleotide pairs used were 5'-GCTATTGCGGATGTTCG-3' (SG57) and a degenerate oligonucleotide 5'-AAIATICKIGCCATICCRAARTC-3' (RK2), 5'-AATCTCTGGCAACAATC-3' (SG46) and 5'-AGCTTTGCAATTAGCCTGACTTCGT-3' (SK26), PS3 (Nishio et al., 1996) and 5'-CTCCTCACTGTTCTCCG-3' (SG48). The 3' end of the *SRK*$_5$ cDNA was isolated by RACE-PCR (Frohman et al., 1988) using 5'-AATCTCTGGCAACAATC-3' (SK27) followed by 5'-AGCTTTGCAATTAGCCTGACTTCGT-3' (SK29) in a nested PCR with the adapter oligonucleotide 5'-GACTCGAGTCGACATCG-3' (RA2).

An *SRK*$_{15}$ cDNA was reconstructed in a similar manner from PCR products amplified with the following oligonucleotide pairs: PS3 and 5'-TTCTCGCCCTCATAAACACAACAG-3' (SK30), 5'-ATTAAGCATTACTACTCGAGTCGCAGAAGC-3' (SK31) and 5'-ATGCAACACCTATTGTGGGAAATCAAGTTC-3' (SK32), and a nested RACE-PCR reaction using SK27 and SK29 in combination with RA2. The 3' ends of transcripts which terminate within the first intron of *SRK*$_{15}$ were amplified by a nested RACE-PCR reaction using oligonucleotides SG46 and 5'-AACAGGTTCCTTACATCATGGAGATCCTAT-3' (SG62) with RA2. Similarly, RACE-PCR and oligonucleotides SG46 and 5'-GGATTGCCAGAGTTTATTCTTAATCAAGGAC-3' (SG64) with RA2 were used to amplify the 3' ends of *SLGA*$_{15}$ transcripts. The first intron of *SRK*$_{15}$ was amplified from genomic DNA of the P57Sc line by the PCR using oligonucleotides SG62 and 5'-

CTCCTCCAAAAGCAGAACACGATAACACTC-3' (SG66). The first intron of *SFR2* was amplified from the cloned *SFR2* gene (Pastuglia et al., 1997a) by the PCR using oligonucleotides 5'-AATGTGAAGAGAAGTGC-3' (fK4) and 5'-CATCAGTCCCTTGTACG-3' (fK28). Part of the *S* domains of each of the three *S* genes (1020 bp from *SLGA*, 1019 bp from *SLGB* and 997 bp from *SRK*) were amplified from the S_{15} homozygous tester line using two oligonucleotides, 5'-TGGAACCCTCAAAATCT-3' (SG22) and 5' GGCCTGCAGCAGCATTCAATCTGAC-3' (SG2), which correspond to sequences conserved in the three genes.

The method used for extraction of genomic DNA has been described previously (Pastuglia et al., 1997b). For RT and RACE PCR, first strand cDNA was prepared from 1 μg of total stigma RNA using the Superscript Kit (Gibco BRL) and oligonucleotide RA1 (Frohman et al., 1988). PCR conditions were 2 min at 94°C followed by 35 cycles of 30 sec at 94°C, 30 sec at 55°C, and 2 min at 72°, with a final extension for 10 min at 72°C. The same conditions were used for nested PCR reactions except that 25 cycles were used for each reaction. RT-PCR and RACE-PCR products were cloned into pGEM-T easy (Promega). DNA sequencing was carried out using the dideoxynucleotide chain termination method (Sanger et al., 1977) either on an automatic sequencer (Applied Biosystems, Foster City, CA) or using a T7 DNA polymerase sequencing kit (Pharmacia Biotech, Uppsala, Sweden). Sequence data was analysed using Lasergene sequence analysis software (DNASTAR, London, UK).

Genetic Mapping of *S* Locus Genes

Segregation of $SLGB_{15}$ in an F_2 population was followed by PCR amplification of a 510 kb sequence from genomic DNA with oligonucleotides 5'-CTACTCCAGATTGACAATCAGTGAGTTG-3' (SG63) and SG2. Segregation of both SRK_{15} and $SLGA_{15}$ was followed using oligonucleotides SG66 and 5'-

TTCTGTTGGAACAACTAAATAAAAT-3' (SK34) for which give two different products of 424 bp and 526 bp respectively for the two genes.

Ribonuclease Protection and Reverse Transcriptase PCR analysis of Gene Transcription

Ribonuclease protection analysis of mRNA abundance was carried out as described previously (Cock et al., 1997). Radiolabeled probes were hybridized to 2 μg of total RNA. An $SLGA_{15}$ specific probe was synthesised by subcloning a 419-bp EcoRI-XhoI fragment into pBluescript II SK^+ (Stratagene, La Jolla, CA), linearising with SalI, and transcribing with T3 RNA polymerase. This probe included bases 553 to 360 of $SLGA_{15}$ (relative to the first base of the ATG initiation codon). For $SLGB_{15}$, the 3' end of the cDNA was sub-cloned as a BamHI fragment into pBluescript II SK^+, linearised with BglII and transcribed with T3 RNA polymerase. This probe included bases 1287 to 860 of $SLGB_{15}$ (relative to the first base of the ATG initiation codon). Reverse transcriptase PCR analysis of SRK_{15} expression was carried out on single-stranded cDNA, prepared as described above, using oligonucleotides SG62 and 5'-CATGAACTCATCGGTACCTTGAGC-3' (SK28). The region between these two oligonucleotide sequences in the SRK_{15} gene includes at least one intron. The expected size of the RT-PCR product is 1175 bp. PCR amplification conditions were as follows: 2 min at 94°C followed by 35 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 1 min and a final elongation step of 10 min at 72°C.

DNA blot analysis

Total genomic DNA was extracted from either young leafs or floral buds as described by Vallejos et al. (1992). DNA blot analysis was carried out using standard procedures (Sambrook et al., 1989). Intron probes were obtained by PCR amplification from plasmid

DNA using 5'-GTAATCTTCTAAAACTAAACACATCGG-3' (SK35) with 5'-TATTGTTCAATGCTAAAGTATGTGC-3' (SK36) to amplify a 415 bp fragment of the $SLGA_{15}$ intron and SK35 with 5'-ATATTAGCCGACCCGTTTCA3' (SK33) to amplify a 474 bp fragment of the $SLGB_{15}$ intron.

Protein Extraction, Deglycosylation, Fractionation of Membrane Proteins, Electrophoretic Analysis and Immunodetection of Proteins

Protein extraction, separation of proteins by isoelectric focusing or SDS-PAGE, electrotransfer onto nitrocellulose membranes and detection of antigen with antibodies were as described previously (Gaude et al., 1991, 1993). Deglycosylation of proteins was carried out as described previously (Delorme et al., 1995b). Stigma plasma membranes were prepared as described by Larsson et al. (1987). Enrichment for plasma membranes was followed using an antibody specific for the plasma membrane proton ATPase (Morsomme et al., 1996). Monoclonal antibody (MAb) 157-35-50, which was raised against a peptide corresponding to the amino-terminal end of a soluble stigma glycoprotein from the P57Sc line, has been described previously (Giranton et al., 1995).

Purification of Stigma Glycoproteins, Amino Acid Sequence Analysis and MALDI-TOF-MS Analysis of Purified Proteins

Soluble stigma glycoproteins were purified by two successive electrophoresis steps as described by Gaude et al. (1991). Purified proteins were electroblotted onto polyvinylidene difluoride membranes (Immobilon P, Millipore) and digested with endoproteinase Lys-C or endoproteinase Gluc-C (Promega, France) as described by Fernandez et al. (1992). Peptides were purified by reverse phase-HPLC using a 400 solvent delivery system (Applied Biosystem, Roissy, France) chromatograph apparatus equipped with a C18 column (Vidac,

Hesperia, CA). The gradient was from 0,095% (v/v) trifluoroacetic acid (TFA) in 4% (v/v) acetonitrile to 0,01% TFA / 48% acetonitrile for 60 min at a flow rate of 300µl/min. Collected peptides were sequenced using the Edman degradation method on an automatic sequencer (Procise 473A protein sequencer, Applied Biosystems).

For mass spectrometry analysis, purified proteins were directly cut from the Imidazole-Zinc stained SDS-PAGE gel (Matsui et al., 1991). The stained protein spots were digested in-gel essentially as desribed by Shevchenko et al. (1996). Briefly, spots were excised from the stained gel, and destained with citric acid. After washing with 50% acetonitrile, which was then removed, gel pieces were dried in a vacuum centrifuge. The dried gel fragments were re-swollen in 20 µl of 20 mM NH_4HCO_3, containing 0.5 µg of trypsin (Promega, Sequencing grade), and incubated for 3 hours at 37°C. A 0.4 µl volume of the digest solution was removed for mass spectrometric analysis. These samples were mixed on the sample probe with 0.4 µl of a saturated solution of α-cyano-4-hydroxy-*trans*-cinnamic acid prepared in 40% v/v acetonitrile, 0.1% v/v trifluoroacetic acid (TFA). Samples were rinsed by placing a 5 µl volume of 0.1% TFA on the matrix surface after the analyte solution had dried completely. After 2 min, the liquid was blown off with pressurized air. MALDI mass spectra of peptide mixtures were obtained using a Bruker Biflex mass spectrometer (Bruker-Franzen Analytik, Bremen, Germany) equipped with a SCOUT multiprobe inlet and a gridless delayed extraction ion source. The ion acceleration voltage was 19.5 kV and the reflectron voltage was 20.0 kV. For delayed ion-extraction, a 6.2 kV potential difference between the probe and the extraction lens was applied. Mass spectra were acquired as the sum of ion signals generated by irradiation of the target with 100 laser pulses. They were calibrated using ion signals from trypsin autodigestion peptides (MH^+ 842.50, MH^+ 1045.55, and MH^+ 2211.09 where MH^+ is the protonated molecular mass).

ACKNOWLEDGMENTS

We thank Marc Boutry (University of Louvain, Belgium) for the kind gift of an antibody against the plasma membrane proton ATPase and Monique Estienne, Anne-Marie Thierry, Richard Blanc, Hervé Leyral and Fabienne Deguerry for technical assistance. We would also like to thank Ioan Negrutiu, Peter Rogowsky and two anonomous reviewers for helpful comments on the manuscript. V.R.-C., and J.M.C. are members of the Institut National de la Recherche Agronomique, J.G. and T.G. are members of the Centre National de la Recherche Scientifique. This work was supported jointly by the Institut National de la Recherche Agronomique, the Centre National de la Recherche Scientifique and the Ecole Normale Supérieure de Lyon.

REFERENCES

Boyes, D.C., and Nasrallah, J.B. (1995). An anther-specific gene encoded by an *S* locus haplotype of *Brassica* produces complementary and differentially regulated transcripts. Plant Cell **7**, 1283-1294.

Boyes, D.C., Chen, C.-H., Tantikanjana, T., Esch, J.J. and Nasrallah, J.B. (1991). Isolation of a Second *S*-Locus-Related cDNA from *Brassica oleracea*: Genetic Relationships Between the *S* Locus and Two Related Loci. Genetics, **127**, 221-228.

Boyes, D.C., Nasrallah, M.E., Vrebalov, J. and Nasrallah, J.B. (1997). The self-incompatibility (*S*) haplotypes of Brassica contain divergent and rearranged sequences of ancient origin. Plant Cell **9**, 237-247.

Charlesworth, D. and Awadalla, P. (1998). Flowering plant self-incompatibility: the molecular genetics of *Brassica* S-loci. Heredity, **81**, 1-9.

Chen, C.-H. and Nasrallah, J.B. (1990). A new class of *S* sequences defined by a pollen recessive self-incompatibility allele of *Brassica oleracea*. Mol. Gen. Genet., **222**, 241-248.

Cock, J.M., Swarup, R. and Dumas, C. (1997). Natural antisense transcripts of the *S* locus receptor kinase gene and related sequences in *Brassica oleracea*. Mol. Gen. Genet., **255**, 514-524.

Conner, J.A., Tantikanjana, T., Stein, J.C., Kandasamy, M.K., Nasrallah, J.B. and Nasrallah, M.E. (1997). Transgene-induced silencing of *S*-locus genes and related genes in *Brassica*. Plant J. **11**, 809-823.

Delorme, V., Gaude, T. Heizmann, P., and Dumas, C. (1995a). Use of immunochemical and SSCP analyses to test homozygosity at the *S* locus of *Brassica oleracea* genotypes. Mol. Breeding **1**, 237-244.

Delorme, V., Giranton, J.L., Hatzfeld, Y., Friry, A., Heizmann, P., Ariza, M.J., Dumas, C., Gaude, T., and Cock, J.M. (1995b). Characterisation of the *S* locus genes, *SLG* and *SRK*, of the *Brassica* S_3 haplotype: identification of a membrane-localised protein

encoded by the *S* locus receptor kinase gene. Plant J. **7**, 429-440.

Fernandez J., DeMott M., Atherton D., and Mische S.M.(1992). Internal protein sequence analysis: Enzymatic digestion for less than 10μg of protein bound to polyvinylidene difluoride or nitrocellulose membrane. Anal. Biochem. **201**, 255-264.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci USA **85**, 8998-9002.

Gangloff, S., Zou, H. and Rothstein, R. (1996). Gene conversion plays the major role in controlling the stability of tandem repeats in yeast. EMBO J., **15**, 1715-1725.

Gaude, T., Friry, A., Heizmann, P., Mariac, C., Rougier, M., Fobis, I., and Dumas, C. (1993). Expression of a self-incompatibility gene in a self-compatible line of *Brassica oleracea*. Plant Cell, **5**, 75-86.

Gaude, T., Denoroy, L. and Dumas, C. (1991). Use of a fast protein electrophoretic purification procedure for N-terminal sequence analysis to identify S-locus related proteins in stigmas of *Brassica oleracea*. Electrophoresis, **12**, 646-653.

Gaude, T., Rougier, M., Heizmann, P., Ockendon, D.J. and Dumas, C. (1995). Expression level of the *SLG* gene is not correlated with the self-incompatibility phenotype in the class II *S* haplotypes of *Brassica oleracea*, Plant Mol. Biol., **27**, 1003-1014.

Giranton, J.-L., Ariza, M.J., Dumas, C., Cock, J.M., and Gaude, T. (1995). The *S* locus receptor kinase gene encodes a soluble glycoprotein corresponding to the SRK extracellular domain in *Brassica oleracea*. Plant J. **8**, 101-108.

Goring, D.R., Glavin, T.L., Schafer, U., and Rothstein, S.J. (1993). An *S* receptor kinase gene in self-compatible *Brassica napus* has a 1 bp deletion. Plant Cell **5**, 531-539.

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

Hatakeyama, K., Takasaki, T., Watanabe, M. and Hinata, K. (1998). Molecular Characterization of *S* Locus Genes, *SLG* and *SRK*, in a Pollen-Recessive Self-Incompatibility Haplotype of *Brassica rapa* L. Genetics **149**, 1587-1597.

Larsson, C., Widell, S. and Kjellbom, P. (1987). Preparation of high purity plasma membranes. Meth. Enzymol., **148**, 558-568.

Liao, D., Pavelitz, T., Kidd, J.R., Kidd, K.K. and Weiner, A.M. (1997). Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the *RNU2* locus) involves rapid intrachromasomal homogenization and rare interchromasomal gene conversion. EMBO J. **16**, 588-598.

McCormick, S. (1998). Self-incompatibility and other pollen-pistil interactions. Curr. Opin. Plant Biol. **1**, 18-25.

Matsui N.M., Smith D.M., Clauser K.R., Fishmann J., Andrews L.E., Sullivan C.M., Burlingame A.L. and Epstein L.B. (1997). Immobilized pH gradient two-dimentional gel electrophoresis and mass spectrometric identification of cytokine-regulated proteins in ME-180 cervical carcinoma cells. Electrophoresis **18**, 409-417.

Morsomme, P., de Kerchove d'Exaerde, A., De Meester, S., Thinès, D., Goffeau, A. and Boutry, M. (1996). Single point mutations in various domains of a plant plasma membrane H^+-ATPase expressed in Saccharomyces cerevisiae increase H^+-pumping and permit yeast growth at low pH. EMBO J. **15**, 5513-5526.

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 **326**, 617-619.

Nasrallah, J.B. (1997). Evolution of the *Brassica* self-incompatibility locus: A look into *S*-locus gene polymorphisms. Proc. Natl. Acad. Sci. USA **94**, 7673-7678.

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

Nasrallah, J.B., Stein, J.C., Kandasamy, M.K., and Nasrallah, M.E. (1994b). Signalling the arrest of pollen tube development in self-incompatible plants. Science **266**, 1505-1508.

Nishio, T., Kusaba, M., Watanabe, M. and Hinata, K. (1996). Registration of *S* alleles in *Brassica campestris* L by the restriction fragment sizes of *SLGs*. Theor. Appl. Genet. **92**, 388-394.

Parniske, M., Hammond-Kosack, K.E., Golstein, C., Thomas, C.M., Jones, D.A., Harrison, K., Wulff, B.B.H. and Jones, J.D.G. (1997). Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell **91**, 821-832.

Pastuglia, M., Roby, D., Dumas, C., and Cock, J.M. (1997a). Rapid induction by wounding and bacterial infection of an *S* gene family receptor-like kinase gene in *Brassica oleracea*. Plant Cell **9**, 49-60.

Pastuglia, M., Ruffio-Chable, V., Delorme, V., Gaude, T., Dumas, C. and Cock, J.M. (1997b). A functional *S* locus anther gene is not required for the self-incompatibility response in *Brassica oleracea*. Plant Cell, 9, 2065-2076.

Ruffio-Châble, V., Hervé, Y., Dumas, C., and Gaude, T. (1997). Distribution of *S*-haplotypes and its relationship with self-incompatibility in *Brassica oleracea*. Part 1. In inbred lines of cauliflower (*B. oleracea* var *'botrytis'*). Theor. Appl. Genet. **94**, 338-346.

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

Sato, T., Thorness, M.K., Kandasamy, M.K., Nishio, T., Hirai, M., Nasrallah, J.B., and Nasrallah, M.E. (1991). Activity of an *S* locus gene promoter in pistils and anthers of transgenic Brassica. Plant Cell **3**, 867-876.

Scutt, C.P. and Croy, R.R.D. (1992). An *S5* self-incompatibility allele-specific cDNA sequence from *Brassica oleracea* shows high homology to the *SLR2* gene. Mol. Gen. Genet. **232**, 240-246.

Shevchenko A, Wilm M, Vorm O, Mann M (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem. **68**, 850-858

Shiba, H., Hinata, K., Suzuki, A. and Isogai, A. (1995). Breakdown of self-incompatibility by the antisense RNA of the *SLG* gene. Proc. Japan Acad., **71**, 81-83.

Stahl, R.J., Arnoldo, M., Glavin, T.L., Goring, D.R. and Rothstein, R.J. (1998). The self-incompatibility phenotype in *Brassica* is altered by the transformation of a mutant *S* locus receptor kinase. Plant Cell **10**, 209-218.

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.

Stein, J.C., Dixit, R., Nasrallah, M.E., and Nasrallah, J.B. (1996). SRK, the stigma-specific *S* locus receptor kinase of Brassica, is targeted to the plasma membrane in transgenic tobacco. Plant Cell **8**, 429-445.

Stephenson, A.G., Doughty, J., Dixon, S., Elleman, C., Hiscock, S. and Dickinson, H.G. (1997). The male determinant of self-incompatibility in *Brassica oleracea* is located in the pollen coating. Plant J. **12**, 1352-1359.

Suzuki, G., Watanabe, M. and Hinata, K. (1997). Highly conserved 5'-flanking regions of two self-incompatibility genes SLG^9 and SRK^9. Gene 191, 123-126.

Suzuki, G., Watanabe, M., Kai, N., Matsuda, N., Toriyama, K., Takayama, S., Isogai, A. and Hinata, K. (1997). Three members of the *S* multigene family are linked to the *S* locus of *Brassica*. Mol. Gen. Genet. **256**, 257-264.

Tantikanjana, T., Nasrallah, M.E., Stein, J.C., Chen, C.-H. and Nasrallah, J.B. (1993). An alternative transcript of the *S*-locus glycoprotein gene in a class-II pollen-recessive self-incompatibility haplotype of *Brassica oleracea* encodes a membrane-anchored protein. Plant Cell **5**, 657-666.

Tantikanjana, T., Nasrallah, M.E. and Nasrallah, J.B. (1996). The *Brassica S* gene family: molecular characterisation of the *SLR2* gene. Sex. Plant Reprod. **9**, 107-116.

Toriyama, K., Stein, J.C., Nasrallah, M.E. and Nasrallah, J.B. (1991). Transformation of *Brassica oleracea* with an *S*-locus gene from *B. campestris* changes the self-incompatibility phenotype. Theor. Appl. Genet. **81,** 769-776.

Walsh, J.B. (1987). Sequence-Dependant Gene Conversion: Can Duplicated Genes Diverge Fast Enough to Escape Gene Conversion? Genetics **117,** 543-557.

Vallejos, C.E., Sakiyama, N.S. and Chase, C.D. (1992). A molecular marker-based linkage map of Phaseolus vulgaris L. Genetics **131,** 733-740.

FIGURE LEGENDS

Figure 1. Structure of three *S* gene family members, $SLGA_{15}$, $SLGB_{15}$, and SRK_{15}, from the *B. oleracea* S_{15} self-incompatibility haplotype.

Schematic representation of the sequenced region of each gene. Transcribed sequences are represented by narrow boxes and coding regions by thick boxes. The hatched box represents a region of 502 bp which is 100% identical at the nucleotide level between $SLGA_{15}$ and SRK_{15}. Membrane-spanning domains (TM) are represented by black boxes. The exon/intron structure of the SRK_{15} kinase domain has not been determined in detail and this region is therefore represented by dashed lines. The transcribed region of the *SLA* gene downstream of $SLGA_{15}$ is represented by dotted lines. $SLGA_{15}$ was previously called SLG_{Sc} (Pastuglia et al., 1997b); a cDNA corresponding to $SLGB_{15}$ (CG15) has been described (Gaude et al., 1993). Accession numbers for the $SLGB_{15}$ and SRK_{15} sequences are Y18261 and Y18260 respectively. *S* dom, *S* domain; end *S* dom, end of the *S* domain; kin, kinase domain.

Figure 2. Genetic linkage to the *S* locus and expression patterns of $SLGA_{15}$, $SLGB_{15}$, and SRK_{15}.

(A) $SLGA_{15}$, $SLGB_{15}$ and SRK_{15} are genetically linked to the *S* locus. DNA was extracted from two parental plants, P57Sc (S_{15}) and a plant homozygous for the S_{3a} haplotype (S_{3a}), and from F_2 progeny (F_2) descended from the two parental plants. PCR amplification was carried out using oligonucleotides SG66 and SK34 for the top gel and SG63 and SG2 for the bottom gel (see Methods), and the PCR products were separated on an agarose gel and stained with ethidium bromide. A water control (H_2O) and three plasmid controls corresponding to SRK_{15} ($pSRK_{15}$), $SLGA_{15}$ (pCG514) and $SLGB_{15}$ (pCG15) were also included to verify that the

amplification was gene-specific. The numbers at right indicate the length of the PCR products in base pairs. The allele corresponding to each band is shown at left. Segregation of the two *S* haplotypes in the F_2 population was followed by immunoblot analysis of stigmatic SLG proteins separated by IEF (Delorme et al., 1995a), and the *S* haplotypes carried by the different progeny are indicated below the lanes.

(B) RNase protection analysis of $SLGA_{15}$ and $SLGB_{15}$ transcripts in a range of organs. Each lane represents 2 μg of total RNA assayed. Autoradiographs were exposed for 7 days (for $SLGA_{15}$) or 16 h (for $SLGB_{15}$). The numbers at right indicate the lengths of the probes and the major protected fragments in nucleotides. P: probe; t: tRNA control; U, B, and T, anthers at the unicellular, bicellular or tricellular stage of microspore development, respectively; S, stigma; R, root; L, leaf.

(C) RT-PCR detection of SRK_{15} transcripts in a range of organs. PCR amplification was carried out with oligonucleotides SG62 and SK28. The number at right indicates the length of the PCR product in base pairs. Abbreviations are as given in (B).

Figure 3. Analysis of molecular mass, glycosylation status and subcellular location of *S*-locus-encoded proteins in the S_{15} and other *B. oleracea S* haplotypes.

(A) Immunoblot analysis of proteins in stigma extracts from *B. oleracea* lines homozygous for the S_2, S_5, S_{15}, S_{25}, and S_3 haplotypes. Proteins were separated by SDS-PAGE and probed with MAb 157-35-50 as the primary antibody. An asterisk marks a 110-kD protein in the S_5 stigma extract that is only weakly recognized by the antibody. Ten micrograms of protein was loaded in each lane. The positions of molecular mass markers are shown at left in kilodaltons.

(B) Deglycosylation of proteins in stigma extracts from *B. oleracea* lines homozygous for the S_2, S_5, S_{15}, and S_3 haplotypes. Untreated (-) or deglycosylated (+) proteins were separated by SDS-PAGE and probed with MAb 157-35-50 as the primary antibody. Asterisks mark the 94

kD deglycosylated SRK_{15} protein and two smaller deglycosylated proteins of 48 and 52 kD. Five μg of protein was loaded in each lane. The positions of molecular mass markers are shown at left in kilodaltons.

(C) The 105-kD SRK_{15} protein is encoded by a gene linked to the *S* locus. Stigma proteins from two parental plants, P57Sc (S_{15}) and a plant homozygous for the S_{3a} haplotype (S_{3a}), and from 24 F_2 progeny (F_2; 11 representative progeny are shown) descended from the two parental plants were separated by SDS-PAGE, and SRK_{15} was detected using MAb 157-35-50 as the primary antibody. The positions of molecular mass markers and SRK_{15} (105-kD) are shown at left and right, respectively. Segregation of the two *S* haplotypes in the F_2 population was followed by immunoblot analysis of stigmatic SLG proteins separated by IEF (Delorme et al., 1995a), and the *S* haplotypes carried by the different F_2 progeny are indicated below the gel.

(D) SRK_{15} is located in the plasma membrane of stigma cells. Total stigma proteins (T) from an S_{15} homozygous plant were fractionated into soluble (S) and microsomal membrane fractions. The microsomal membranes were then further fractionated by aqueous two-phase partitioning to give fractions enriched in intracellular membranes (IM) and plasma membranes (PM). Samples from the different fractions were separated by SDS-PAGE and immunoblotted. *S*-locus-encoded proteins and the plasma membrane proton ATPase were detected with MAb 157-35-50 (top gel) or with an anti-ATPase antibody (bottom gel), respectively. The positions of molecular mass markers are shown at left in kilodaltons.

Figure 4. Biochemical analysis of soluble *S*-locus-encoded proteins in stigmas of the S_{15} haplotype.

Soluble stigma proteins were separated by IEF gel electrophoresis and immunoblotted with MAb 157-35-50. Data to the right show the pIs and relative abundances of the four proteins,

α,_β, β' and β", detected by MAb 157-35-50. The four proteins were purified by selective two-dimensional gel electrophoresis (Gaude et al., 1991) and analyzed by Edman degradation. Based on peptide sequence data, α and_β were assigned to *SLGB* (at right). Question marks indicate that the peptide sequence data did not permit the unequivocal assignment of a protein to a particular *S* locus gene.

Figure 5. RNA transcripts of *SLGA*$_{15}$, *SLGB*$_{15}$, and *SRK*$_{15}$.

Schematic representation of *SLGA*$_{15}$, *SLGB*$_{15}$ and *SRK*$_{15}$ showing transcripts which have been identified by cDNA cloning and RACE-PCR. Transcribed regions of the genes are represented by narrow boxes and coding regions by thick boxes. Membrane-spanning domains (TM) are represented by black boxes. The exon/intron structure of the *SRK*$_{15}$ kinase domain (kin) has not been determined in detail, therefore, this region is represented by dashed lines. The transcribed region of the *SLA* gene downstream of *SLGA*$_{15}$ is represented by dotted lines. Transcripts are shown as arrows with their heads at the site of polyadenylation. Expanded views show the sequence adjacent to each polyadenylation site. Two different polyadenylation sites were found near the 5' end of the first intron of *SRK*$_{15}$. Transcripts identified by the RACE-PCR technique are shown with a dashed line at the 5' end as only the 3' part of the transcript was cloned. The sizes of the introns shown are 833 bp, 1132 bp and 623 bp for *SLGA*$_{15}$, *SLGB*$_{15}$ and *SRK*$_{15}$ respectively. *S* dom, *S* domain; end *S* dom, end of the *S* domain; kin, kinase domain.

Figure 6. Alignment of the deduced amino acid sequences of the S domain of SRK_{15}, SLG_2, $SLGA_{15}$, $SLGB_{15}$, and SLG_5.

Dashes indicate identity with the SRK_{15} sequence. Gaps have been introduced into the SLG_2 and $SLGA_{15}$ sequences to optimize the alignment. The sequence of SLG_5 is incomplete and, therefore, the N-terminal end of this protein is represented by a dotted line. The five amino acid differences between SLG_2 and $SLGA_{15}$ are marked by black spots. The two hypervariable domains are underlined.

Figure 7. Conserved sequences within the introns of $SLGA_{15}$, $SLGB_{15}$ and SRK_{15} suggest a complex evolutionary history.

(A) Schematic representation of the introns of $SLGA_{15}$ and $SLGB_{15}$ and intron 1 of SRK_{15} (only intron sequences are represented). Sequences conserved between different introns are shown as black, hatched or stippled boxes. The percentage similarity between two regions that are conserved in different introns is indicated. A long direct repeat within the $SLGB_{15}$ intron is indicated by arrows. Accession numbers for the $SLGB_{15}$ and SRK_{15} intron sequences are Y18261 and Y18258 respectively.

(B) Two possible models for the evolution of *SLGA*, *SLGB* and *SRK* based on the structure of the introns of the alleles present in the S_{15} haplotype. Gene duplication events are not shown for the sake of clarity. In the deletion model, the ancestor of each *S* locus gene would have lost a different region of the precursor intron (indicated by the bracketed bars) by deletion. The deletion in the *SLGB* lineage would have also removed downstream sequences encoding the membrane-spanning region and hence created a new 3' splice site. In the recombination model, it is hypothesized that recombination would have allowed the membrane-spanning and kinase domains of a receptor-like kinase gene (RLK) to be positioned downstream of an *S* domain. See text for details. DELET., deletion of intron sequences; DUPLICATION, duplication of intron sequences; RECOMB., recombination between intron sequences.

Figure 8. Detection of *SLGA* and *SLGB* alleles in different class II haplotypes.

DNA gel blot of BamHI and EarI digested genomic DNA from *B. oleracea* lines homozygous for the S_2, S_5, S_{15} and S_3 haplotypes. Fragments of the introns of $SLGA_{15}$ and $SLGB_{15}$ were used as gene-specific probes, as indicated below each gel. The positions of molecular length markers are shown at left in kilobases.

TABLES

Table 1. Percent amino acid similarity[a] between the *S* domain sequences of $SLGA_{15}$, $SLGB_{15}$ and SRK_{15} and those of selected members of the *S* gene family in Brassica[b].

---

[a]The region compared is actually slightly shorter than the *S* domain because the sequence of the amino-terminal end of SLG_5 was not available. The region compared runs from the 21st residue of the mature *S* domain (for $SLGB_{15}$) to the end of the *S* domain. The results of the comparisons are expressed as percentage similarity which is calculated as follows: 100 times the number of matched amino acids divided by the sum of the length in amino acids of the aligned region, plus the number of gaps introduced to optimise the alignment.

[b]$SLGA_{15}$ was previously SLG_{Sc} (Pastuglia et al, 1997b), SLG_2 (Chen and Nasrallah, 1990), $SLGB_{15}$ corresponds to the cDNA CG15 (Gaude et al., 1993), SLG_5 (Scutt and Croy, 1992), SRK_{15} (this study, accession number Y18260), SRK_2 (Stein et al., 1991), SRK_5 (this study, accession number Y18259), SLR2-2 (Boyes et al., 1991), SLG_3 and SRK_3 (Delorme et al., 1995b), SFR2 (Pastuglia et al., 1997a).

Table 2. MALDI-TOF-MS analysis of soluble *S*-locus-encoded proteins in stigmas of the S_{15} haplotype.

---

[a] Numbers refer to the position of a peptide in the sequence of the corresponding protein.

[b] M, methionine.