The Plant Cell, Vol. 11, 2217–2231, November 1999, www.plantcell.org © 1999 American Society of Plant Physiologists

Structural and Transcriptional Comparative Analysis of the *S* Locus Regions in Two Self-Incompatible *Brassica napus* Lines

Yuhai Cui, Norbert Brugière, Lisa Jackman, Yong-Mei Bi, and Steven J. Rothstein[1,2]

Department of Molecular Biology and Genetics, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Self-incompatibility (SI) in Brassica is controlled by a single locus, termed the *S* locus. There is evidence that two of the *S* locus genes, *SLG*, which encodes a secreted glycoprotein, and *SRK*, which encodes a putative receptor kinase, are required for SI on the stigma side. The current model postulates that a pollen ligand recognizing the SLG/SRK receptors is encoded in the genomic region defined by the *SLG* and *SRK* genes. A fosmid contig of ~65 kb spanning the *SLG-910* and *SRK-910* genes was isolated from the *Brassica napus* W1 line. A new gene, *SLL3*, was identified using a novel approach combining cDNA subtraction and direct selection. This gene encodes a putative secreted small peptide and exists as multiple copies in the Brassica genome. Sequencing analysis of the 65-kb contig revealed seven additional genes and a transposon. None of these seven genes exhibited features expected of *S* genes on the pollen side. An ~88-kb contig of the *A14 S* region also was isolated from the *B. napus* T2 line and sequenced. Comparison of the two *S* regions revealed that (1) the gene organization downstream of *SLG* in both *S* haplotypes is highly colinear; (2) the distance between *SLG-A14* and *SRK-A14* genes is much larger than that between *SLG-910* and *SRK-910*, with the intervening region filled with retroelements and haplotype-specific genes; and (3) the gene organization downstream of *SRK* in the two haplotypes is divergent. These observations lead us to propose that the *SLG* downstream region might be one border of the *S* locus and that the accumulation of heteromorphic sequences, such as retroelements as well as haplotype-unique genes, may act as a mechanism to suppress recombination between *SLG* and *SRK*.

INTRODUCTION

Self-incompatibility (SI) is one of the mechanisms promoting outbreeding that have evolved in higher plants. In *Brassica* spp, SI is controlled sporophytically by a single Mendelian genetic locus, the *S* locus (Bateman, 1955). Up to 100 *S* alleles have been identified in the genus *Brassica* (Bateman, 1955; Ockendon, 1974; de Nettancourt, 1977). Two genes are known to segregate with the genetically defined *S* locus; therefore, the term *S* haplotype has been used to collectively describe the components of each *S* region. One gene encodes a secreted glycoprotein called the *S* locus glycoprotein (SLG; Nasrallah et al., 1987). The second gene encodes a protein predicted to consist of an extracellular domain linked to a cytoplasmic kinase domain by a transmembrane domain (Stein et al., 1991). This structure resembles the receptor kinases from animals (Ullrich and Schlessinger, 1990) and therefore is called the *S* receptor kinase (SRK).

Both *SLG* and *SRK* genes have been shown to be highly expressed in stigmas and at a lower level in anthers (Nasrallah et al., 1985; Stein et al., 1991; Goring and Rothstein, 1992). In the pistil, *SLG* is expressed predominantly in the surface papillar cells of the stigma (Sato et al., 1991), and its glycoprotein product accumulates to high levels in the walls of these cells (Kandasamy et al., 1989). SRK has been demonstrated to be a membrane-bound glycosylated protein in the stigmatic papillae (Delorme et al., 1995; J.C. Stein et al., 1996). These localizations are consistent with cytological observations that demonstrate the arrest of pollen or pollen tube development at the stigma surface. Although low levels of both *SLG* and *SRK* transcripts have been detected in anthers, the most sensitive detection methods available have failed to detect SLG and SRK proteins in this organ (J.C. Stein et al., 1996).

There is genetic evidence that the expression of both genes in papillar cells is required for the operation of SI, although some recent work has questioned the necessity of *SLG* for SI (Cabrillac et al., 1999). Several self-compatible mutant strains of Brassica have been identified in which self-compatibility is associated with spontaneous mutations at the *S* locus that disrupt the *SRK* gene (Goring et al., 1993; Nasrallah et al., 1994). In addition, spontaneous mutations at loci unlinked to the *S* locus that downregulate the *SLG* gene (Nasrallah et al., 1992) and transgene-induced mutations that downregulate the *SLG* (Toriyama et al., 1991) and

[1] To whom correspondence should be addressed. Email rothsteinsj @phibred.com; fax 515-270-4312.

[2] Current address: Department of Agronomic Traits, Pioneer Hi-Bred International, Inc., P.O. Box 1004, Johnston, IA 50131.

SRK (Conner et al., 1997; Stahl et al., 1998) genes are associated with the loss of the pistil's ability to inhibit self-pollen. From this evidence, it can be hypothesized that the *SLG* and *SRK* genes are necessary for SI, but it is not clear whether they are sufficient for the operation of the SI response in the pistil. Published attempts to modify SI specificity by transformation experiments have been unsuccessful (Conner et al., 1997).

Generally, within a haplotype, SLG sequences are very similar (∼90%) to extracellular domains of SRKs, suggesting that the two genes have coevolved and that the *SLG* gene very likely was generated by partial duplication of *SRK.* Alleles of the *SLG* and *SRK* genes from different haplotypes share ∼70 to 80% amino acid sequence identity, although exceptions have been found recently (Kusaba et al., 1997; Kusaba and Nishio, 1999). This allelic diversity is thought to determine the specificity of the SI reaction (Nasrallah and Nasrallah, 1993), and this implies that the *SLG* and *SRK* genes need to be tightly linked as one genetic unit. Indeed, recombination events between *SLG* and *SRK* have not been detected, even though the physical distance between the two genes may exceed 200 kb of DNA (Boyes and Nasrallah, 1993; Boyes et al., 1997).

A model that is analogous to the ligand-activated receptor tyrosine kinase described in animal systems (Ullrich and Schlessinger, 1990) for SRK activation and SI specificity has been proposed (Stein et al., 1991; Goring and Rothstein, 1992). The SRK protein kinase would be activated by contact between a papillar cell and self-pollen. By phosphorylating intracellular substrates, the SRK protein would couple the initial molecular recognition events at the papillar cell–pollen interface to the signal transduction pathway that ultimately leads to pollen rejection. Because SLG and SRK are both expressed in papillar cells in the absence of pollen, a pollen-borne component, possibly a ligand for SRK, has been postulated. Such an extracellular ligand would be highly polymorphic and encoded within the *S* locus complex. It would activate the receptor in a haplotype-specific manner, thus providing the specificity in self-recognition. SLG, which is freely diffusible in the cell wall, would be essential, either by acting as an extracellular regulator for accessing ligand to the signaling receptor or by being an integral part of a functional receptor complex.

Given the genetic analysis of SI, factors that contribute to the complementary stigma and pollen recognition system must be encoded by the *S* locus region. Thus, it is important to characterize the chromosomal region defined by the *SLG* and *SRK* genes in detail to determine whether there are any anther-specific genes encoded by this region and, if these are present, whether they are required for the expression and specificity of SI. Molecular analysis of regions flanking *SLG* and *SRK* genes in various *S* haplotypes has led to the identification of some new genes. These are vegetatively expressed (Boyes et al., 1997), reproductive tissue specific (Yu et al., 1996), or anther specific (Boyes and Nasrallah, 1995; Yu et al., 1996), but none of them has been shown to be SI related.

We have been working with two self-incompatible *B. napus* subsp *oleifera* lines called W1 and T2. The W1 line carries a functional *B. rapa S* haplotype in the self-compatible Westar background (Goring et al., 1992a); the T2 line carries a functional *B. napus* subsp *rapifera S* haplotype in the self-compatible Topas background (Goring et al., 1992b). The *SLG* and *SRK* genes (named the *910* allele for the W1 line and the *A14* allele for the T2 line) in the two lines have been characterized (Goring et al., 1992a, 1992b; Glavin et al., 1994). In this study, we report the isolation, transcriptional analysis, and comparison of the *S* locus regions in the *910* and *A14 S* haplotypes. Our goals are to continue the search for the pollen *S* gene and to gain insights into the structural basis for recombination suppression in the *S* locus.

RESULTS

Cloning of the *SLG-910* and *SRK-910* Genomic Region

Previously, the *SLG-910* and *SRK-910* genomic region was cloned as two λ genomic clones. The *SLG* and *SRK* genomic clones did not overlap, and the gap between them was estimated to be ∼2 kb by pulsed-field gel electrophoresis (PFGE) analysis, which showed that both *SLG-910* and *SRK-910* cDNAs hybridized with an ∼25-kb EcoRI fragment (Yu et al., 1996). To verify the physical linkage of the two clones, we decided to clone the genomic region. A partial genomic cosmid library was therefore constructed as detailed in Methods and screened with the *SLG-910* and *SRK-910* kinase portion cDNAs as probes. Of ∼100,000 colonies, of which ∼15,000 contained inserts, one positive clone was isolated. DNA gel blot analysis confirmed that this cosmid clone contains *SLG-910*, *SRK-910*, *SLL1* (for S locus linked 1)-*910*, and *SLL2-910* genes. The region covered by the cosmid and the two λ clones was sequenced. The sequencing results show that the *SLG-910* and *SRK-910* genes transcribe toward opposite directions and that the distance between them (from start codon to start codon) is only ∼6 kb, which is the shortest among the *S* haplotypes characterized thus far (Boyes et al., 1997; Conner et al., 1998). These new data correct the previous work of Yu et al. (1996), in which the *SLG* and *SRK* genes were thought to be transcribed in the same direction and in which *SLL1* and *SLL2* were placed between *SLG* and *SRK.*

To further extend the cloned *S* locus region, we constructed a W1 fosmid (Kim et al., 1992) genomic library. Approximately 200,000 colonies were screened by using the *SLG-910* and *SRK-910* kinase cDNAs as probes. Four independent positive clones were isolated. The insert end sequences were obtained by sequencing the fosmid directly, with the T7 and SP6 primers flanking the cloning site on the vector. The end sequences then were used to locate and order the fosmid clones relative to the known sequences. As a result, the fosmid contig extends the cloned *910 S* region from ~30 to ~65 kb, with ~15 kb on the downstream side of the *SRK* and ~20 kb on the upstream side of the *SLG* gene.

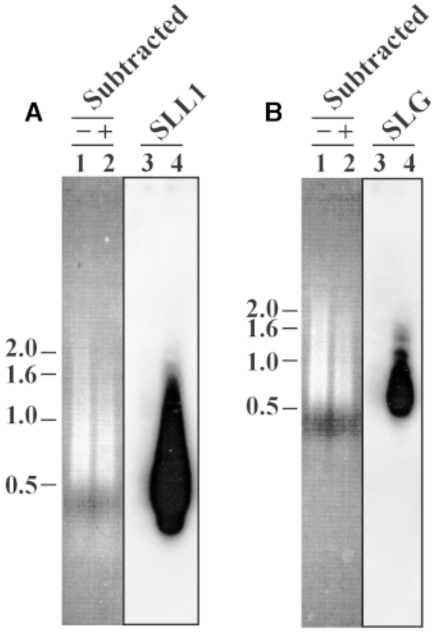

Figure 1. Enrichment of *S* Locus–Specific cDNAs by Subtraction.

(A) Enrichment of the *SLL1-910* cDNA. PCR-amplified W1 cDNA sequences that were not subtracted (−) and subtracted (+) (lanes 1 and 2, respectively) were probed with *SLL1* cDNA (lanes 3 and 4, correspondingly).
(B) Enrichment of the *SLG-910* cDNA. PCR-amplified W1 cDNA sequences that were not subtracted (−) and subtracted (+) (lanes 1 and 2, respectively) were probed with *SLG* cDNA (lanes 3 and 4, correspondingly).
Molecular length markers are indicated at left in kilobases.

Enrichment of *S* Locus–Specific cDNAs by Subtraction and Mapping of the cDNAs to the Cloned *S* Chromosomal Region

To search for other potential *S* genes encoded in the cloned genomic region, we performed a cDNA subtraction to enrich the *S* locus–specific cDNAs. The self-incompatible *B. napus* line W1 and self-compatible line Westar are nearly isogenic (Goring et al., 1992a), and the only difference between them is the *910 S* haplotype and possibly the flanking regions. Theoretically, the cDNAs encoded by the *910 S* haplotype should be enriched by subtraction, given the allele diversity shown by known *S* genes and expected for a new *S* gene(s). A technique called subtraction suppression hybridization (Diatchenko et al., 1996; Gurskaya et al., 1996) was used. The technique combines a high subtraction efficiency with an equalized representation of differentially expressed sequences, which is achieved by a specific form of polymerase chain reaction (PCR) called suppression PCR. This permits the exponential amplification of cDNAs that differ in abundance, whereas amplification of sequences of identical abundance in the two populations is suppressed.

Whole flower buds were used rather than anthers only, based on the following considerations: (1) the procedure is able to enrich rare messages, so the relative abundance of a specific mRNA is not a major concern for the success of the process; and (2) it potentially allows the enrichment of polymorphic/differentially expressed genes from the stigma as well. Because it is not known whether the pollen factor controlling SI specificity is expressed in diploid meiocytes premeiotically or in tapetal cells derived sporophytically, 1- to 2-mm buds were used. These stages correspond to the premeiotic diploid meiocyte or to the microspore release during microsporogenesis, respectively (Scott et al., 1991). It is known that the two well-characterized *S* genes, *SLG* and *SRK*, are expressed in 2- to 7-mm buds, with the peak in 4- to 6-mm buds. Therefore, we reasoned that their interaction partner(s) from pollen theoretically should exhibit a similar expression pattern or might be transcribed earlier in development, with the transcript or translated protein maintained in the pollen. Based on this reasoning, 3- to 4-mm buds also were included. Figures 1A and 1B show that the subtraction was successful: both *SLL1-910* and *SLG-910* were markedly enriched. *SLG* is a polymorphic gene, and *SLL1* has been shown to be expressed only in the W1 line (Yu et al., 1996). This result shows the suitability of this procedure for the enrichment of other potential *S* genes.

To map the subtracted W1 cDNAs to the cloned *S* genomic region, we then took an approach similar to direct cDNA selection (reviewed in Lovett, 1994). Cloned genomic DNA fragments were fractionated, blotted, and hybridized with a radioactively labeled cDNA mixture. Figures 2A to 2C show the mapping of the 14-kb *SRK* λ genomic clone. Based on the map of the clone, it was restricted with four enzymes individually, to separate the band containing *SRK* from the rest. One clear signal was obtained for the bands containing the 1.4-kb EcoRI fragment, which is 3 kb downstream of the *SRK* gene. The strength of the signal was weaker than that of *SLG* but comparable to that of *SRK.* The probe DNA was eluted and amplified by PCR. The PCR products were cloned and sequenced. When compared with the genomic sequence, four introns clearly could be identified. No matches were found in the public databases for this new gene, designated *SLL3.*

The entire 65-kb cloned region then was scanned using the same procedure (data not shown). For convenient identification of whether a signal was from a new gene or from a previously identified gene, genomic DNA was either amplified by PCR or restricted with appropriate restriction enzymes that could separate the fragment containing known genes from the rest. Two additional weaker signals were obtained. The probe DNAs were eluted and amplified by PCR.

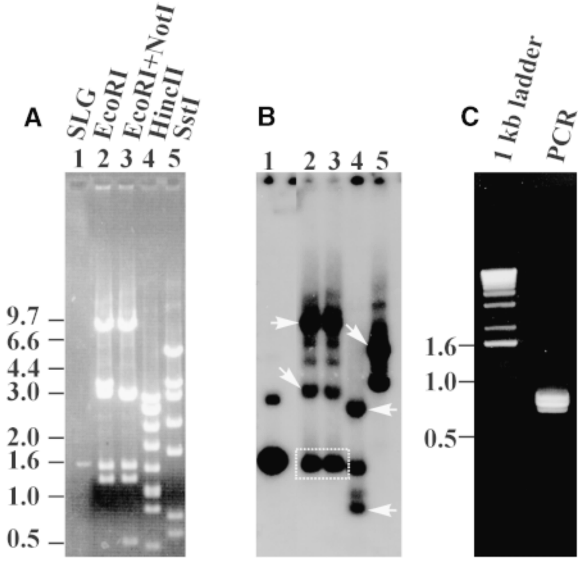

Figure 2. Mapping of Subtraction-Enriched cDNAs to the *S* Locus by Using the 14-kb *SRK* λ Clone as an Example.

(A) Gel separation of the 14-kb *SRK* λ genomic clone (recloned into pBluescript SK+) digested with four restriction enzymes as indicated above each lane (lanes 2 to 5). Lane 1 is *SLG-910* as control.
(B) DNA gel blot hybridization of the gel in **(A)** with radioactively labeled subtracted cDNA mixture. The arrows indicate signals from *SRK.* The lanes are as in **(A)**.
(C) PCR amplification of the probe DNAs eluted from the filter region corresponding to the bands in the rectangle in lanes 2 and 3 in **(B)**.
Molecular length markers are indicated in kilobases at left.

Cloning and sequencing of the PCR products revealed two more genes that encode homologs of methionyl-tRNA transformylase (*FMT*) and a *Caenorhabditis elegans* putative protein (*CePP*). These two genes also were revealed by homology search of the public databases (see Sequence Analysis of the N65-kb *910 S* Region).

Characterization of *SLL3*

Sequencing of the selected *SLL3* cDNAs revealed two kinds of sequences. The longer ones have RsaI sites at both ends, suggesting that it is a partial cDNA (because the cDNAs were digested with RsaI before subtraction). Therefore, 5′ and 3′ rapid amplification of cDNA end (RACE) reactions were performed to obtain the full-length cDNA. The 3′ RACE products were very short, with the longest only 130 bp from the RsaI site. A stop codon (TAG) is only 33 bp downstream of the RsaI site, followed immediately by a stretch of T residues, denoting that this is the likely 3′ end of the transcript. Sequences of eight randomly picked 5′ RACE product clones were determined. The sequences fell into two groups, each with four members. One group of the isolated cDNAs was very similar to the *910* genomic sequence, and two introns could be easily identified.

Attempts to clone the cDNA that matches perfectly with *SLL3-910* have been unsuccessful. Primers supposedly specific to the *910* allele were designed and used for PCR amplification with a W1 bud cDNA pool as the template. Cloning and sequencing of the PCR products revealed that they were similar to, but not identical to, the *910* genomic sequence, suggesting that *SLL3*-like genes constitute a multiple-copy gene family. This was confirmed by DNA gel blot analysis. As shown in Figures 3A and 3B, genomic blots prepared with DNAs from W1 and Westar exhibited multiple bands when probed with the *SLL3* coding region. Furthermore, the signals were much stronger than normal single-copy genomic blots, suggesting that each band actually might consist of multiple subbands. Fosmid clones were isolated by using the 5′ portion of the *SLL3* cDNA as a probe. When DNA blots of these were analyzed, they exhibited bands varying slightly at ∼1.4 kb, indicating that the ∼1.4-kb band in the genomic blot actually was a complex of multiple bands (Figure 3C).

This transcribed region is still of interest. The putative *SLL3-910* full-length cDNA sequence was obtained by eliminating these introns from the genomic sequence and assembling the exons and is shown in Figures 4A and 4B. Sequence analysis revealed two open reading frames (ORFs), encoding proteins of 65 (ORF1) and 291 (ORF2) amino acids, respectively. Neither of these has matches in the public databases. ORF1 encodes a putative signal peptide, with a potential cleavage site between positions 17 and 18, which conforms to the (−3, −1) rule (von Heijne, 1986; Bairoch, 1993). Furthermore, the small polypeptide is very hydrophobic. However, because of its multiple-copy nature, we were unable to determine the expression pattern of the *SLL3-910* by RNA gel blot and reverse transcription–PCR analyses.

Sequence Analysis of the ∼65-kb *910 S* Region

The ∼65-kb genomic DNA of the *910 S* region was completely sequenced, except for a gap of ∼400 bp, which could not be sequenced due to the presence of a long stem–loop structure. The sequence was used to search the databases. As a result, an *Enhancer*/*Suppressor-mutator* (*En*/*Spm*)–type transposon (*910Tn1*) and several homologs of known genes, such as those encoding methionyl-tRNA transformylase (*FMT*), the Drosophila seven-in-absentia protein (*SIAH1* and *SIAH2*), Clp protease (*ClpP*), and a protein kinase (*Bkin*), were detected. In addition, homologs of predicted putative proteins from the *C. elegans* (*CePP*) and Arabidopsis (*AtPP*) genome projects also were identified in the region. Next, the exon prediction program GenScan was used to search for new genes. Although a few more new

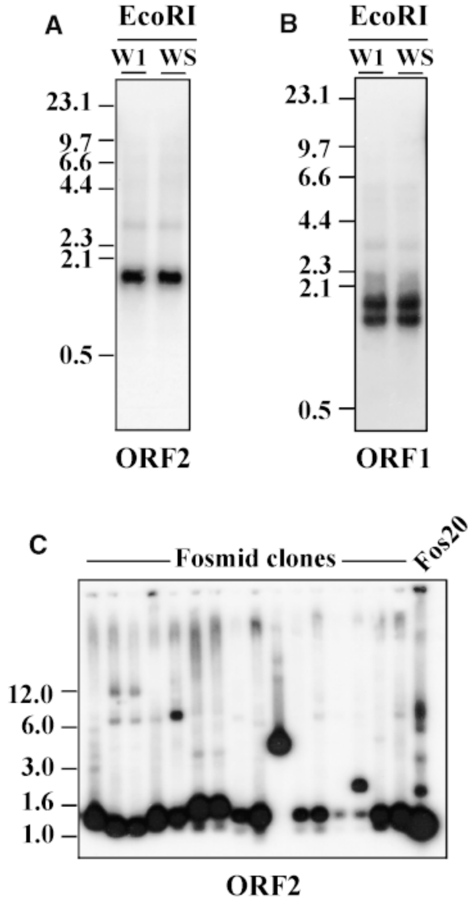

Figure 3. DNA Gel Blot Analysis of the Putative *SLL3* Gene.

(A) and **(B)** Genomic blot analysis of *SLL3.* Sources of genomic DNA and restriction enzyme used are indicated above the lanes. WS, Westar.

(C) Fosmid blot analysis of *SLL3.* Sources of plasmid DNA are indicated above the lanes. Fosmid clones were isolated with the 5′ portion of the *SLL3* cDNA as probe. Fos20 contains *SLL3-910.* The restriction enzyme used was EcoRI.

The probes used are indicated underneath each blot, with genomic DNA from either the ORF1 or the ORF2 regions used, as defined in Figure 4. Numbers at left indicate length markers in kilobases.

ORFs could be predicted, we failed to confirm their existence by reverse transcription–PCR analysis. The genes identified by similarity searches are summarized in Table 1. To confirm the presence of new transcription units in the *910 S* locus region, we designed specific primers and used them to PCR amplify cDNA from a cDNA pool prepared from W1 flower buds. Primers were chosen so that PCR-amplified cDNA could be easily distinguished from possible genomic DNA amplification products by the absence of at least one intron. Partial or complete cDNA sequences corresponding to the *CePP*, *FMT*, *SIAH1*, *AtPP*, and *Bkin* sequences were successfully amplified, confirming their existence as transcription units. Because the *SIAH2* sequence did not contain any intron in its coding region, we were unable to draw a similar conclusion for this putative gene.

Our sequence analysis revealed several features that placed *910Tn1* in the *En*/*Spm* family. First, as shown in Figure 5A, the only ORF found in *910Tn1* exhibits strong similarity to TNP2 of *Tam1* (Nacken et al., 1991) and the ORF of *Tdc1* (Ozeki et al., 1997). Both of these are characterized members of the *Spm* family. Both Tdc1–ORF and Tam1–TNP2 are the homologs of the *tnpD* gene product of the *En*/*Spm* element from maize (Pereira et al., 1985). Deletion derivatives of the *En*/*Spm* family that lack the homologous region of *tnpD* have either no or greatly reduced ability to transpose autonomously (Fedoroff, 1989). Second, as shown in Figure 5B, clustered direct repeats and inverted repeats were found no more than 500 bp upstream of *910Tn1–ORF.* This repetitive structure was shown to be essential for transposition and is another major structural feature of the *En*/*Spm* family (Gierl, 1996). There is another long stem–loop structure at the 3′ end side of the ORF, which is why we were unable to sequence it. We were unable to locate the terminal inverted repeats and direct repeats. The element is likely to be inactive because an in-frame stop codon was found.

Isolation of an ∼88-kb Fosmid Contig That Spans the *SLG-A14* and *SRK-A14* Genomic Region

Allelic sequence diversity is thought to be the critical feature for genes involved in SI recognition. The putative pollen *S* component is expected to be as polymorphic as that observed for *SLG* and *SRK*, and specific combinations of allelic forms of these genes are thought to define different SI specificities. Therefore, cloning and sequencing of another *S* haplotype are necessary, because sequence comparison of two haplotypes would allow direct identification of polymorphic sequences as candidate *S* genes. Sequence comparison of two haplotypes also would reveal the structural basis for the suppression of recombination in the *S* locus. A T2 fosmid genomic library therefore was constructed and screened with the *SLG-A14* and *SRK-A14* cDNAs. Four independent clones were isolated from ∼200,000 colonies. One clone (Fos163) hybridized with the *SLG* probe only, whereas the other three (Fos156, Fos158, and Fos162) hybridized with the *SRK* probe only. Thus, no clone appeared to contain both genes. To check whether the *SLG* clone overlaps with the *SRK* clones, we labeled Fos163 and used it to probe the blot prepared with all four fosmid DNAs digested with HindIII (data not shown). The result showed that Fos163 and Fos156 did contain three common fragments sized ∼3.5, ∼3, and ∼0.7 kb, respectively, indicating that the fosmid contig covers the entire *SLG-A14* and *SRK-A14* region. The contig was estimated to be ∼88 kb of DNA, based on restriction analysis.

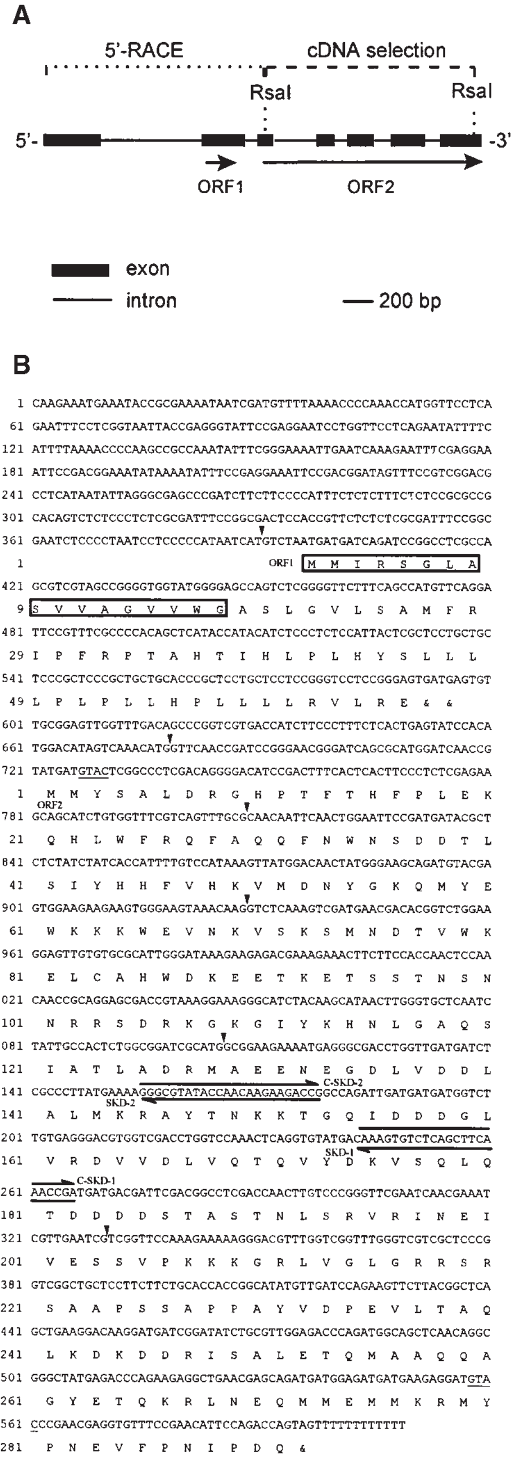
A
5'-RACE
cDNA selection
RsaI
RsaI
5'-
-3'
ORF1
ORF2
exon
intron
200 bp
B
1 CAAGAAATGAAATACCGCGAAAATAATCGATGTTTTAAAACCCCAAACCATGGTTCCTCA
61 GAATTTCCTCGGTAATTACCGAGGGTATTCCGAGGAATCCTGGTTCCTCAGAATATTTTC
121 ATTTTAAAACCCCAAGCCGCCAAATATTTCGGGAAAATTGAATCAAAGAATTTCGAGGAA
181 ATTCCGACGGAAATATAAAATATTTCCGAGGAAATTCCGACGGATAGTTTCCGTCGGACG
241 CCTCATAATATTAGGGCGAGCCCGATCTTCTTCCCCATTTCTCTCTTTCTCTCCGCGCCG
301 CACAGTCTCTCCCTCTCGCGATTTCCGGCGACTCCACCGTTCTCTCTCGCGATTTCCGGC
361 GAATCTCCCCTAATCCTCCCCCATAATCATGTCTAATGATGATCAGATCCGGCCTCGCCA
1 ORF1 M M I R S G L A
421 GCGTCGTAGCCGGGGTGGTATGGGGAGCCAGTCTCGGGGTTCTTTCAGCCATGTTCAGGA
9 S V V A G V V W G A S L G V L S A M F R
481 TTCCGTTTCGCCCCACAGCTCATACCATACATCTCCCTCTCCATTACTCGCTCCTGCTGC
29 I P F R P T A H T I H L P L H Y S L L L
541 TCCCGCTCCCGCTGCTGCACCCGCTCCTGCTCCTCCGGGTCCTCCGGGAGTGATGAGTGT
49 L P L P L L H P L L L L R V L R E & &
601 TGCGGAGTTGGTTTGACAGCCCGGTCGTGACCATCTTCCCTTTCTCACTGAGTATCCACA
661 TGGACATAGTCAAACATGGTTCAACCGATCCGGGAACGGGATCAGCGCATGGATCAACCG
721 TATGATGTACTCGGCCCTCGACAGGGGACATCCGACTTTCACTCACTTCCCTCTCGAGAA
1 M M Y S A L D R G H P T F T H F P L E K
ORF2
781 GCAGCATCTGTGGTTTCGTCAGTTTGCGCAACAATTCAACTGGAATTCCGATGATACGCT
21 Q H L W F R Q F A Q Q F N W N S D D T L
841 CTCTATCTATCACCATTTTGTCCATAAAGTTATGGACAACTATGGGAAGCAGATGTACGA
41 S I Y H H F V H K V M D N Y G K Q M Y E
901 GTGGAAGAAGAAGTGGGAAGTAAACAAGGTCTCAAAGTCGATGAACGACACGGTCTGGAA
61 W K K K W E V N K V S K S M N D T V W K
961 GGAGTTGTGTGCGCATTGGGATAAAGAAGAGACGAAAGAAACTTCTTCCACCAACTCCAA
81 E L C A H W D K E E T K E T S S T N S N
1021 CAACCGCAGGAGCGACCGTAAAGGAAAGGGCATCTACAAGCATAACTTGGGTGCTCAATC
101 N R R S D R K G K G I Y K H N L G A Q S
1081 TATTGCCACTCTGGCGGATCGCATGGCGGAAGAAAATGAGGGCGACCTGGTTGATGATCT
121 I A T L A D R M A E E N E G D L V D D L
C-SKD-2
1141 CGCCCTTATGAAAAGGGCGTATACCAACAAGAAGACCGGCCAGATTGATGATGATGGTCT
SKD-2
141 A L M K R A Y T N K K T G Q I D D D G L
1201 TGTGAGGGACGTGGTCGACCTGGTCCAAACTCAGGTGTATGACAAAGTGTCTCAGCTTCA
SKD-1
161 V R D V V D L V Q T Q V Y D K V S Q L Q
C-SKD-1
1261 AACCGATGATGACGATTCGACGGCCTCGACCAACTTGTCCCGGGTTCGAATCAACGAAAT
181 T D D D D S T A S T N L S R V R I N E I
1321 CGTTGAATCGTCGGTTCCAAAGAAAAAGGGACGTTTGGTCGGTTTGGGTCGTCGCTCCCG
201 V E S S V P K K K G R L V G L G R R S R
1381 GTCGGCTGCTCCTTCTTCTGCACCACCGGCATATGTTGATCCAGAAGTTCTTACGGCTCA
221 S A A P S S A P P A Y V D P E V L T A Q
1441 GCTGAAGGACAAGGATGATCGGATATCTGCGTTGGAGACCCAGATGGCAGCTCAACAGGC
241 L K D K D D R I S A L E T Q M A A Q Q A
1501 GGGCTATGAGACCCAGAAGAGGCTGAACGAGCAGATGATGGAGATGATGAAGAGGATGTA
261 G Y E T Q K R L N E Q M M E M M K R M Y
1561 CCCGAACGAGGTGTTTCCGAACATTCCAGACCAGTAGTTTTTTTTTTTT
281 P N E V F P N I P D Q &

Figure 4. Genomic Organization of the Putative *SLL3-910* Gene.

(A) Schematic representation of *SLL3* genomic organization. The regions obtained by cDNA selection and RACE experiments are indicated. Arrows indicate the transcriptional orientation of the ORFs.

(B) Nucleotide and deduced amino acid sequences of the putative *SLL3-910* gene-coding region. The arrowheads correspond to the sites of introns. The ampersands indicate the stop codons. RsaI sites (GTAC) are underlined. Boxed amino acids indicate the predicted signal peptide. Primers C-SKD-2 and C-SKD-1 are nested gene-specific primers for 3′ RACE, and SKD-1 and SKD-2 are for 5′ RACE.

Sequencing Analysis of the ~88-kb *A14 S* Region and Its Comparison with the *910 S* Region

The ~88-kb fosmid contig was subcloned and sequenced. Database searches revealed 12 more putative genes in addition to the *SLG* and *SRK* genes as well as two retroelements. The putative *SLL3-A14* allele was not found in this region. The genes identified by similarity searches are summarized in Table 1. *Athila* is a newly identified retroelement family in Arabidopsis (Pelissier et al., 1995, 1996; Thompson et al., 1996; Wright and Voytas, 1998). Because the similarity extends only over a relatively short stretch and we could not locate other structural features, such as long terminal repeats (LTRs) and direct repeats, the data are not shown here. *SLL2* was isolated previously from the *910* haplotype and was predicted to be closely linked to the *SLG* gene in the *A14* haplotype by PFGE analysis (Yu et al., 1996). Our sequence analysis confirms the previous report and shows further that *SLL2* is highly conserved between the two haplotypes. *SLL1* was isolated previously from the *910* haplotype as a candidate pollen ligand gene, because it is closely linked to *SLG* and only expressed in anthers. However, it only encoded a very short polypeptide, and when other *SLL1* cDNAs were PCR amplified and sequenced, no allelic sequence diversity was observed (Yu et al., 1996). Here, by genomic sequence analysis, we identified the *A14 SLL1* allele. When the genomic sequences of both alleles were aligned, significant similarity was found only in the 5′ and 3′ noncoding regions. The *A14 SLL1* cDNA was PCR amplified using specific primers from this region and sequenced. As was seen for the *910* allele, sequence analysis revealed no significant ORF.

As shown in Figure 6A, two nested retroelements were inserted in the *CaBP*-related sequence. Both elements have LTRs sharing 95% identity. The internal sequence between the longer LTRs is only 304 bp and contains no ORF, suggesting that this is a deletion derivative, designated *A14RT1A.* Deletion derivatives of retrotransposons have been found in plants (Bennetzen, 1996). The internal sequence within the shorter set of LTRs encodes a long ORF, which contains all of the conserved protein domains of LTR retrotransposons. The element is termed *A14RT1B* and is a member of the *Ty1*/*copia* family. All the protein domains share significant homologies with the *Melmoth* element, a *Ty1*/*copia*–type retroelement identified from the *Ssc* haplotype of *B. oleracea* (Pastuglia et al., 1997). Although *A14RT1B* contains almost all the essential components of an active retrotransposon, it is likely to

Table 1. Summary of the Genes Identified by Homology Searches

	Sequence Similarity to Known Genes			
Gene[a]	Homolog	Similarity (Accession Number)[b]	References	Description
CePP	*C. elegans* putative protein	24% (Q09305)		
FMT	Methionyl-tRNA transformylase	30 to 40%, bacterial isologs (P43523; AAC68132; CAB13446	Meinnel and Blanquet, 1994; Kunst et al., 1997; Stephens et al., 1998	First plant homolog; domain structure conserved with those of bacteria
SlAH1	Drosophila seven-in-absentia	23% Drosophila (P21461) 25% Arabidopsis (U90439)	Carthew and Rubin, 1990	Both C3HC4 and cysteine-rich domains conserved
SlAH2		21% (nt) Drosophila (P21461) 19% (nt) Arabidopsis (U90439)		Partial, likely a pseudogene; only the the second cysteine-rich domain identified
AtPP	Arabidopsis putative protein	29% (AL022223) 27% (Z99708)		
ClpP	Clp protease	96% *B. rapa* (L41144) 85% Arabidopsis (AF032103)	Letham and Nasrallah, 1998	Catalytic triad identified
SLL2-A14	*SLL2-910*	98% (66193)	Yu et al., 1996	
SLL1-A14	*SLL1-910*	42% (nt) (U66192)	Yu et al., 1996	
BKin	Arabidopsis SNF1-related protein kinase AKin10	80% (M93023)	Le Guen et al., 1992	Ser/Thr protein kinase consensus revealed; disrupted by multiple in-frame stop codons and frameshifts
CaBP	Arabidopsis putative calcium binding protein	84% (Z97343)	Bevan et al., 1998	Disrupted by the retroelement *A14RT1*
AtPP1	Arabidopsis putative protein	29% (AL022223) 27% (Z99708)		
AtPP2	Arabidopsis putative protein	24% (AL022223) 22% (Z99708)		
AtPP3	Arabidopsis putative protein	ND[c] (X97827)		
AtPP4	Arabidopsis putative protein	ND (AAB95232)		
DNA ligase	DNA ligase I	53% Arabidopsis (Q42572)	Tomkinson et al., 1991	Ligase active site identified: KYDGERA

[a] Gene designations are based on their homologs for known genes; putative genes predicted from the Arabidopsis/*C. elegans* genome sequencing projects are named *AtPP*/*CePP* for Arabidopsis/*C. elegans* putative protein, respectively. *BKin*, Brassica kinase; *CaBP*, calcium binding protein.
[b] The amino acid (or nucleotide if specified by nt) homology between the *S* locus genes and their database homologs as well as their GenBank accession numbers are indicated. Because *FMT-A14* is at the end of the *A14* contig, we have cloned only the 3′ portion. For *CaBP*, we could not identify the 5′ portion by homology to the Arabidopsis gene. Therefore, for these two genes, the homologies are based on the identified regions. For the DNA ligase, the homology is based on ~200 amino acids at the C terminus.
[c] ND, homology not determined because it is only over a relatively short stretch.

be inactive because an in-frame stop codon is found in the endonuclease domain (Figure 6B).

As represented in Figure 7, comparison of the gene organizations of the *910* and *A14 S* haplotypes revealed a few striking structural features. First, in the region downstream of the *SLG* gene, the gene organization is highly colinear, except that *AtPP-910*, *AtPP1*, and *2-A14* apparently are rearranged. In addition, the similarities between the gene pairs are extremely high, with all of them being >95% similar. Second, the distance between the *SLG-A14* and *SRK-A14* genes is ~32 kb of DNA, and this region is filled with retroelements and sequences not found in the corresponding region of the *910* haplotype. In comparison, in the *910* haplotype, the distance is only 6 kb, and no gene/ORF was found in this sequence. Third, the gene organization in the region downstream of *SRK* is very divergent between the two haplotypes.

DISCUSSION

Search for the Pollen *S* Gene

In this study, we isolated the chromosomal regions defined by the *SLG* and *SRK* genes from two *S* haplotypes. We then took two approaches to search the genes encoded in this

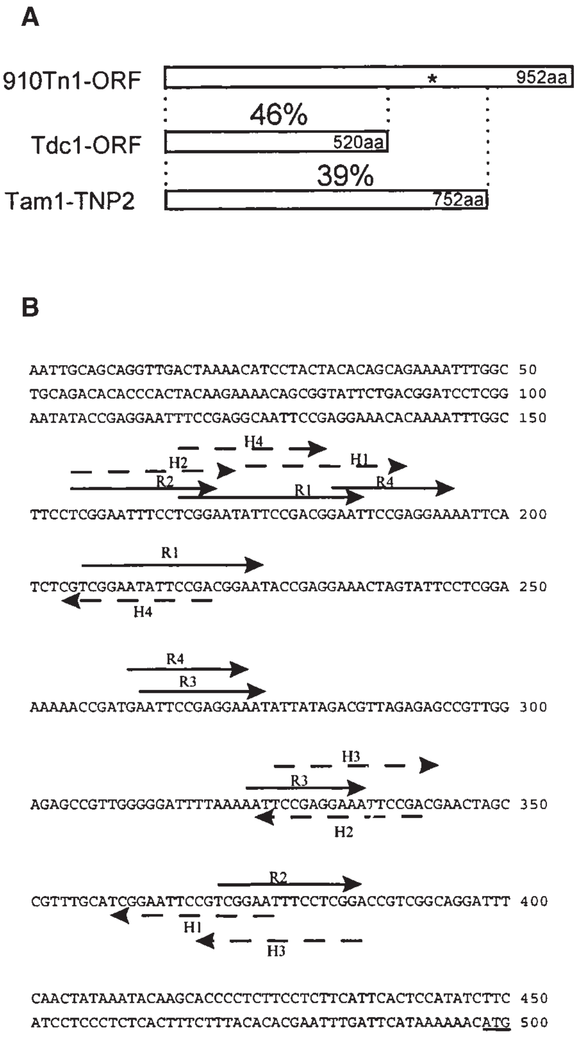
A
910Tn1-ORF
*
952aa
46%
Tdc1-ORF
520aa
39%
Tam1-TNP2
752aa
B
AATTGCAGCAGGTTGACTAAAACATCCTACTACACAGCAGAAAATTTGGC 50
TGCAGACACACCCACTACAAGAAAACAGCGGTATTCTGACGGATCCTCGG 100
AATATACCGAGGAATTTCCGAGGCAATTCCGAGGAAACACAAAATTTGGC 150
H4
H2
H1
R2
R1
R4
TTCCTCGGAATTTCCTCGGAATATTCCGACGGAATTCCGAGGAAAATTCA 200
R1
TCTCGTCGGAATATTCCGACGGAATACCGAGGAAACTAGTATTCCTCGGA 250
H4
R4
R3
AAAAACCGATGAATTCCGAGGAAATATTATAGACGTTAGAGAGCCGTTGG 300
H3
R3
AGAGCCGTTGGGGGATTTTAAAAATTCCGAGGAAATTCCGACGAACTAGC 350
H2
R2
CGTTTGCATCGGAATTCCGTCGGAATTTCCTCGGACCGTCGGCAGGATTT 400
H1
H3
CAACTATAAATACAAGCACCCCTCTTCCTCTTCATTCACTCCATATCTTC 450
ATCCTCCCTCTCACTTTCTTTACACACGAATTTGATTCATAAAAAACATG 500

Figure 5. Evidence Showing That *910Tn1* Is an *En*/*Spm*–Type Transposable Element.

(A) Schematic representation showing the homologies between the ORF of *910Tn1* (910Tn1–ORF) and the ORF/peptide encoded by other *En*/*Spm* family elements Tdc1 (Ozeki et al., 1997) and Tam1 (Nacken et al., 1991). aa, amino acids; asterisk, an in-frame stop codon within the 910Tn1–ORF.

(B) Arrangement of repetitive sequences in the 500-bp upstream region of *910Tn1–ORF.* The length of the line represents the length of the repetitive elements; the arrows indicate the orientation. Elements that are in the same direction as the sequence are labeled above the sequence, and those that are complementary are marked underneath the sequence. The same numbers are used for corresponding elements. H, inverted repeats (dashed arrows), which could form hairpin structures; R, direct repeats (solid arrows).

region. First, the *S* locus–specific genes were enriched by cDNA subtraction, and the enriched cDNAs were mapped to the cloned *S* region by direct cDNA selection. Second, the cloned *S* regions from two haplotypes were completely sequenced, and the coding regions were analyzed by database searches and exon prediction programs. The first approach, which combines the advantages of both suppression subtraction and direct cDNA selection, worked well for the identification of differentially expressed (such as *SLL1*) or polymorphic (such as *SLG*) genes, as evidenced by strong hybridization signals from *SLL1*, *SLG*, and *SRK.* In other words, if there were other *S* genes encoded in the cloned *S* region and they were as polymorphic/differentially expressed as *SLL1* and *SLG*, they should have been detected by using this procedure. The results of our novel cDNA mapping approach are consistent with those of the sequence analysis.

It has been postulated that the putative pollen ligand gene should possess the following properties: (1) it should be *S* locus linked; (2) it should exhibit haplotype-specific polymorphism (as do the *SLG* and *SRK* genes); (3) it is likely to be expressed in anthers in a manner consistent with the genetic and developmental regulation of SI; (4) it is likely to encode a secreted peptide; (5) it may be deleted or mutated in a nonfunctional self-fertile *S* haplotype; and (6) a functional equivalent of the gene might not be found in the homologous region of the Arabidopsis genome. This last hypothesis is based on a recent comparative genomic mapping analysis between the S_8 haplotype of *B. rapa* and its homolog in Arabidopsis (Conner et al., 1998). The major conclusion of that comparison is that the autogamy in Arabidopsis has evolved by deletion of the SI genes. According to the above criteria, SLL3–ORF1 is interesting, although it does not match these perfectly. First, its 5′ end encodes a putative signal peptide. Second, it is very hydrophobic and would therefore fit the lipid-rich environment of the Brassica pollen surface (Murphy and Ross, 1998). Third, no cross-hybridizing sequences were detected in the homologous region of the Arabidopsis genome (data not shown). However, because of its multiple-copy nature, we were unable to study its expression pattern.

SLL1 was a potential candidate ligand gene, although based on the cDNA clones that were isolated, it did not seem to exhibit allelic polymorphism or code for a polypeptide of significant length (Yu et al., 1996). Based on comparisons of the *910* and *A14* genomic sequences of *SLL1*, there is a significant sequence diversity between them (~42% at the nucleotide level), but no significant ORF was detected. A subset of other genes could be eliminated as ligand candidates, based on sequence similarity alone, such as DNA ligase, *FMT*, and *ClpP.* The *CaBP*, *Bkin*, and *SIAH* homologs all are likely to play some roles in signal transduction pathways as they do in other systems, although their products are unlikely to act as a ligand. *CaBP-A14* is interrupted by retroelements, so it is not functional. *Bkin* also is not functional because it is interrupted by in-frame stop codons and

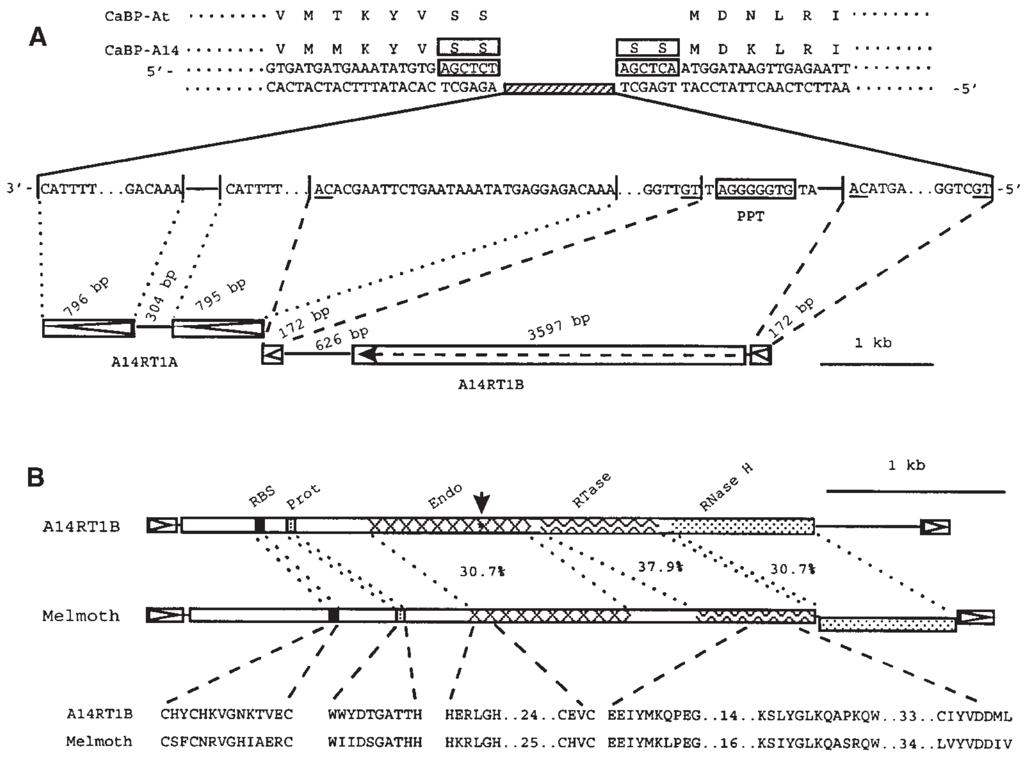
A
CaBP-At ········ V M T K Y V S S M D N L R I ········
CaBP-A14 ········ V M M K Y V S S S S M D K L R I ········
5′- ········GTGATGATGAAATATGTG AGCTCT AGCTCA ATGGATAAGTTGAGAATT········
········CACTACTACTTTATACAC TCGAGA TCGAGT TACCTATTCAACTCTTAA········ -5′
3′- CATTTT...GACAAA CATTTT... ACACGAATTCTGAATAAATATGAGGAGACAAA ...GGTTGT T AGGGGGTG TA ACATGA...GGTCGT -5′
PPT
796 bp
304 bp
795 bp
172 bp
626 bp
3597 bp
172 bp
1 kb
A14RT1A
A14RT1B
B
1 kb
RBS
Prot
Endo
RTase
RNase H
A14RT1B
30.7%
37.9%
30.7%
Melmoth
A14RT1B CHYCHKVGNKTVEC WWYDTGATTH HERLGH..24..CEVC EEIYMKQPEG..14..KSLYGLKQAPKQW..33..CIYVDDML
Melmoth CSFCNRVGHIAERC WIIDSGATHH HKRLGH..25..CHVC EEIYMKLPEG..16..KSIYGLKQASRQW..34..LVYVDDIV

Figure 6. *A14RT1B* Is a *Ty1*/*copia*–Type Retroelement.

(A) Schematic representation of *A14RT1* and its insertion site into the calcium binding protein in the *A14 S* haplotype (CaBP-A14). Target site duplication (in boxes) can be easily identified by comparing the *CaBP-A14* and its Arabidopsis homolog *CaBP-At.* The sequences at the borders of the different domains are shown. The sizes of domains are labeled in base pairs. The boxes with arrowheads represent the LTRs. Each box represents the ORF, and the arrow within each box indicates the transcriptional orientation. The border sequences and the boxes are connected by dotted and dashed lines for *A14RT1A* and *A14RT1B*, respectively. Lines represent noncoding regions. The inverted repeat sequences TG . . . CA at the ends of LTRs also are underlined. The left LTR of *A14RT1A* and the right LTR of *A14RT1B* overlap by 32 bp. The sequence similarities between the left and right LTRs of *A14RT1A* and those of *A14RT1B* are both 95%. PPT, putative polypurine tract, the priming site for initiation of reverse transcription.

(B) Comparison of the structure and amino acid sequence of *A14RT1B* and the *Melmoth* element (Pastuglia et al., 1997), a *Ty1*/copia–type retroelement in the *Ssc* haplotype of *B. oleracea.* The positions of conserved regions corresponding to the RNA binding site (RBS), the protease site (Prot), the endonuclease (Endo), the reverse transcriptase region (RTase), and the RNase H region are indicated by various boxes. The RNase H of the *Melmoth* element is encoded by a small separate ORF. The percentage of amino acid similarity between *A14RT1B* and the *Melmoth* for the different domains is indicated. Blocks of conserved amino acid sequence within each region are shown. Their position relative to the diagram is indicated with dashed and dotted lines. The arrow above the asterisk represents an in-frame stop codon mutation. The LTRs are represented by boxes with arrowheads inside. The coding and noncoding regions are represented by boxes and lines, respectively.

frameshift mutations. The *CePP* and *AtPP* homologs are potentially interesting, although the latter is homologous to sequences in Arabidopsis. However, it is certainly possible that some genes that are common between the two homologous regions might fulfill an SI-related function in Brassica while serving a different function in Arabidopsis.

Physical Boundaries of the *S* Locus

Classic genetic analysis defined the *S* locus as a single Mendelian locus. Indeed, recombination events between *SLG* and *SRK* never have been observed even when 500 plants in a single F_2 population were analyzed (Boyes et al., 1997). Thus, theoretically, the boundaries of the *S* locus could be defined genetically by mapping out the recombination break-points flanking the *SLG-SRK* region, with the breakpoints defining the *S* locus region. Thus, it is important to analyze the recombination frequency in the chromosomal region encompassing and flanking the *S* locus and to determine whether it differs significantly from recombination frequencies in other regions of the Brassica genome. A similar strategy has been used successfully in defining the boundaries of the mating-type locus of the green alga

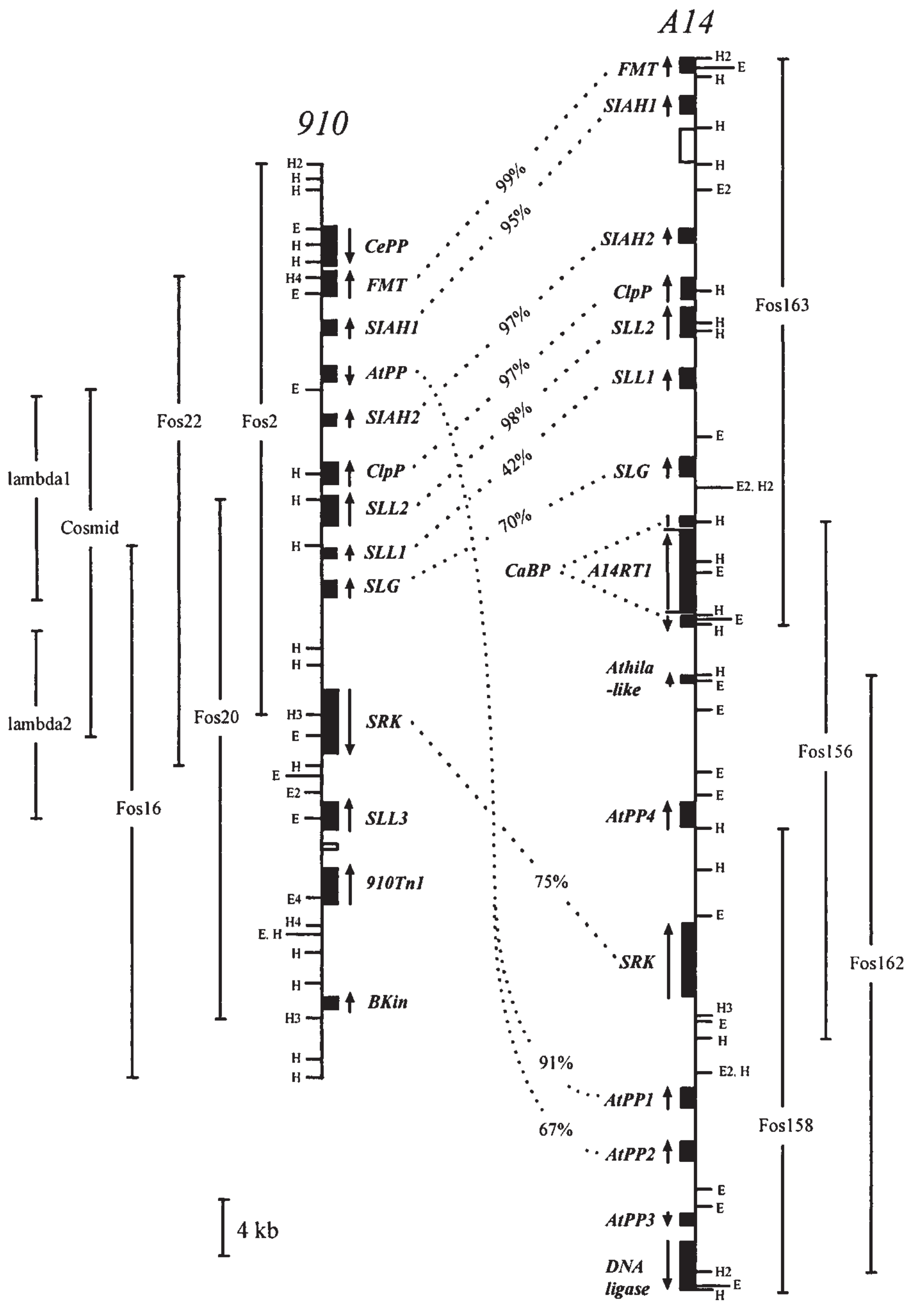

Figure 7. Comparative Map of the *910* and *A14 S* Regions That Have Been Cloned and Sequenced.

The 65-kb *910 S* region and its coverage by λ clones, cosmid clone, and fosmid clones (Fos) are shown at left, and the 88-kb *A14 S* region and its coverage by a fosmid contig are shown at right. For both maps, the recognition sites of EcoRI (E) and HindIII (H) are shown. The numbers after the letters indicate multiple sites. Genes are represented by boxes, and the arrows indicate the orientation of transcription. The white box in the *910* map indicates a region of ∼400 bp that could not be sequenced due to a long stem–loop structure. The white box in the *A14* map represents four nearly identical 673-bp HindIII fragments. The relative positions of these fragments have not been determined. Genes found in common between the two haplotypes are connected by dotted lines, and the amino acid similarities (nucleotide similarity for *SIAH2* and *SLL1*) between the corresponding gene pairs are shown. The *910* and *A14 S* region sequences have GenBank accession numbers AJ245479 and AJ245480, respectively.

Chlamydomonas (Ferris and Goodenough, 1994). However, largely due to the lack of dense molecular markers, currently no estimates exist for the actual frequency of recombination in the chromosomal region containing the *S* locus and in other regions of the Brassica genome, although Brassica linkage maps are available (Slocum et al., 1990; Song et al., 1991), and the *S* locus has been positioned on at least one such linkage map (Camargo et al., 1997). Furthermore, an F_2 segregating population, which is adequately large for measuring low-frequency recombination events, is needed.

Alternatively, the *S* locus boundaries also could be derived. The analysis of the Chlamydomonas mating-type locus provides a paradigm again, where the genetic and physical mapping data match perfectly in defining the locus boundaries (Ferris and Goodenough, 1994). The mating-type locus of Chlamydomonas exists as two apparent alleles (mt^+ and mt^-), and recombination suppression has long been associated with the mating-type locus. A chromosome walk through the *mt* locus was conducted until the frequent recombination breakpoints were reached. In total, 1.1 Mb of DNA from a wild-type mt^- strain and 0.85 Mb of the corresponding mt^+ DNA were ordered by chromosome walking. When restriction maps of the two regions are compared, three domains can be recognized: the central rearranged domain and two flanking domains. In the central domain, the mt^+ and mt^- maps are not colinear. The two flanking domains, although not necessarily identical, are clearly colinear for the full extent of the walk. Genetic mapping located the frequent recombination breakpoints to be 525 and 110 kb from the borders of the central domain. Within these intervals (from the borders of the central domain to the frequent recombination breakpoints), recombination has not been entirely eliminated. Therefore, recombination suppression extends out to either side of the central domain in a gradient. Ferris and Goodenough (1994) concluded that the central domain may in fact contain all the mating-type-specific genes.

As shown in Figure 7, the genes downstream of *SLG* in *910* and *A14 S* haplotypes are highly colinear, except that *AtPP-910*, *AtPP1*, and *2-A14* are apparently rearranged. Not only are they colinear, but the sequence similarities between the counterparts are extremely high. By analogy to the Chlamydomonas mating-type locus, we propose that the *SLL2-SLL1* region actually may represent one border of the *S* locus core region. However, more physical data from other *S* haplotypes and genetic mapping data are needed to support this hypothesis.

Implications for the Mechanisms of How the Recombination between *SLG* and *SRK* Is Suppressed

The *S* locus shares a related feature with mating-type loci and sex chromosomes: their control genes are transmitted as a unit and fail to recombine. Four mechanisms have been identified as explaining the observed suppression of recombination (as summarized in Ferris and Goodenough, 1994). First, the relevant genes at the locus may be highly dissimilar, precluding recombination. Second, genes may be present in one locus but absent in the other, with the extreme example being the male fertility genes on the Y chromosome. Third, in an apparently special case involving *Schizosaccharomyces pombe*, a novel chromatin configuration, designed to regulate mating-type switching, has the additional effect of suppressing recombination. Fourth, homology may persist, but recombination is suppressed because of chromosomal rearrangements. From a physical mapping comparison of three haplotypes, Boyes et al. (1997) found some rearrangements in the *S* locus. Based on this preliminary data, a hypothesis was proposed for the *S* locus evolution by analogy to the prevalent theory of sex chromosome evolution, which proposes that homomorphism gives rise to heteromorphism via the accumulation of chromosomal rearrangements (Bull, 1983).

From our comparative sequence analysis of the *910* and *A14 S* haplotypes, we find some rearrangements in the locus; for example, *SRK-910* and *SRK-A14* are divergently transcribed, suggesting a possible inversion event, and *AtPP-910* might have been translocated and duplicated in the *A14 S* haplotype. More importantly, we find that haplotype-specific genes and retroelements have contributed to the expansion and restructuring of the *SLG-SRK* genomic region. Transposon-mediated expansion of chromosomal regions and chromosomal restructuring have been widely found in the evolution of plant genomes (Wessler et al., 1995; SanMiguel et al., 1996). Two nested retrotransposons and one putative *Athila*-like retroelement have been identified in the *A14 S* region. The left and right LTRs of both nested retroelements show 95% identity, suggesting that the two were inserted at roughly the same time and that the insertion is an ancient event, because the two LTRs of a retrotransposon are usually identical at the time of its insertion into the host genome due to the nature of the transposition process (Lewin, 1997). Nucleotide substitution causes sequence divergence between the two LTRs, and the degree of divergence can be used as an estimate for the age of the insertion event (SanMiguel et al., 1998). In summary, chromosomal rearrangements, accumulation of transposons, and haplotype-unique genes all have contributed to the heteromorphism in the *S* locus, helping to suppress recombination and maintaining the gene pair as a genetic unit.

Based on our results, a few statements can be made regarding the structure and gene organization of the *S* locus: (1) the *S* locus is a gene-rich region, that is, the gene density is approximately one gene per 5 kb, which is similar to that of Arabidopsis (Bevan et al., 1998), which has a much more compact genome; (2) the *S* genes are embedded in non-SI-related genes; and (3) transposable elements, both transposons and retrotransposons, have contributed to the evolution of the *S* locus.

METHODS

Plant Materials

The *910* haplotype in the W1 line originated from a self-incompatible *Brassica rapa* plant and was introgressed into *B. napus* subsp *oleifera* cv Westar through a series of backcrosses, as described by Goring et al. (1992a). The self-incompatible T2 line also was produced by a series of backcrosses involving a field-vernalized stock of the self-incompatible rutabaga (*B. napus* subsp *rapifera*) line Z as the donor of the *S* locus and a spring canola quality variety Topas as the recipient (Goring et al., 1992b).

Genomic DNA Gel Blots

Genomic DNA was extracted according to the method described by Goring et al. (1993). Approximately 5 to 10 μg of genomic DNA was digested with restriction enzymes and fractionated on a 0.8% agarose gel. Blotting, hybridization, and probe labeling were performed as described previously (Yu et al., 1996).

Construction and Screening of Cosmid and Fosmid Genomic Libraries

High molecular weight DNA was isolated from nuclei and embedded in agarose plugs as described by Zhang et al. (1995). To clone the 25-kb EcoRI fragment in the *910* haplotype, we made a partial cosmid W1 genomic library. Plugs were completely digested with EcoRI and then subjected to Gelase (Epicentre, Madison, WI) digestion. The digested DNA was precipitated as recommended by the manufacturer. Cosmid vector pLAFR1 (Friedman et al., 1982) was completely digested with EcoRI and dephosphorylated with HK phosphatase (Epicentre), as recommended by the manufacturer. Approximately 600 ng of digested genomic DNA was ligated overnight at 4°C to 500 ng of vector in a total volume of 10 μL with 400 units of T4 DNA ligase (New England Biolabs, Beverly, MA). Four microliters of this ligation mixture was packaged in vitro by using the Gigapack gold packaging system (Stratagene, La Jolla, CA). The cosmid particles were transfected to *Escherichia coli* XL1-blue MR (Stratagene), and the cells were spread onto Luria-Bertani plates containing 15 mg/L tetracycline.

The procedure for constructing fosmid libraries was developed based on previously described protocols (Kim et al., 1992; J.L. Stein et al., 1996). The plugs were partially digested with HindIII and fractionated by pulsed-field gel electrophoresis (PFGE). The target region containing the partially digested DNA ranging from ~40 to 50 kb was excised. The agarose was digested, and the DNA was dephosphorylated by using Gelase and HK phosphatase (Epicentre), respectively, as recommended by the manufacturer. Protein was removed by gentle phenol–chloroform extraction, and the DNA was pelleted. Vector arms were prepared from pFOS1 as described previously (Kim et al., 1992). The partially digested genomic DNA was ligated overnight at 4°C to the pFOS1 arms in a 15-μL ligation reaction mixture containing 1.5 μg each of vector and insert and 1 μL of T4 DNA ligase (400 units per μL; New England Biolabs). The ligated DNA in 4 μL of this reaction mixture was packaged in vitro by using the Gigapack gold packaging system (Stratagene); the fosmid particles were transfected into *E. coli* XL1-blue MR competent cells and spread onto Luria-Bertani plates containing 12.5 μg/mL chloramphenicol. Screening of genomic libraries was performed by standard methods (Sambrook et al., 1989). Prehybridization, hybridization, and washing conditions were the same as those described above for genomic gel blots.

cDNA Subtraction

The PCR-Select cDNA subtraction kit (Clontech, Palo Alto, CA) was used to enrich the W1-specific cDNA species. All the procedures were conducted according to the manufacturer's instructions. mRNA from 1- to 4-mm flower buds was used. Equal amounts of tissues not only in total but also at every size level from W1 and Westar plants were used simultaneously for RNA extraction, according to Jones et al. (1985), to minimize false positives. Polyadenylated RNA was purified with the mRNA purification kit from Pharmacia Biotechnology. The RNA concentration was estimated by comparison with RNA standard markers (Gibco BRL) run on the same gel. cDNA made from Westar mRNA was used as the driver and that from W1 as the tester. The Advantage cDNA PCR kit (Clontech) was used for all of the polymerase chain reactions (PCRs). The cycling was conducted with the Perkin-Elmer GeneAmp PCR System 9600; thus, the denaturing time used was 10 sec instead of 30 sec.

Mapping of the Subtracted cDNAs to the Cloned *S* Region

An approach similar to the direct cDNA selection (reviewed in Lovett, 1994) was used to map the subtraction-enriched cDNAs to the *S* locus. The procedure was intended to combine the advantages of both cDNA subtraction and direct cDNA selection. The subtracted cDNA mixture was labeled by random priming and used to probe the DNA blot prepared with *910* haplotype genomic DNA. After autoradiography, the region of the blot corresponding to the signal of interest was excised. The bound probe DNA was eluted as described previously (Ausubel et al., 1988). The membrane piece was rinsed in 1 mL of 0.1 × SSC (1 × SSC is 0.15 M NaCl and 0.015 M sodium citrate, pH 7.0), transferred to 300 μL of sterile water containing 5 μg of yeast tRNA, boiled for 2 min, frozen in a dry ice–ethanol bath immediately, and thawed at room temperature. Thirty microliters of sodium acetate, pH 5.2, and 900 μL of ethanol were mixed well with the solution, which was incubated in a dry ice–ethanol bath for 15 min. The DNA was pelleted by centrifugation for 15 min at 13,000 rpm. The pellet was washed with 70% ethanol, dried, and dissolved in 10 μL Tris-EDTA. Two microliters of this DNA was amplified with the Advantage cDNA PCR kit and the nested primers 1 and 2 that had been used in the subtraction. The amplified DNAs were cloned into pGEM-T/pGEM-T Easy vectors (Promega) and sequenced with the universal sequencing primers on the vectors.

Rapid Amplification of cDNA Ends

A double-stranded cDNA pool was synthesized from poly(A)$^+$ RNA isolated from 1- to 4-mm W1 flower buds by using the cDNA synthesis reagents provided in the PCR-Select cDNA subtraction kit. For 3′ rapid amplification of cDNA ends (RACE), two nested gene-specific primers, C-SKD-1 and 2 as shown in Figure 3B, were synthesized. The first-round PCR was performed on the cDNA pool, using the 3′ distal primer and the dT_{17} adapter primer (for the sequences of the adapter and dT_{17} adapter primer, see Goring et al., 1992b). The products then were subjected to a second-round PCR with the 3′ proximal primer and the adapter primer.

For 5′ RACE, the cDNA pool was ligated to the adapters 1 and 2 provided in the PCR-Select cDNA subtraction kit, and as was done during 3′ RACE, two nested gene-specific primers also were synthesized as indicated in Figure 3B. The first-round PCR was conducted with the cDNA pool by using the 5′ distal primer and the nested PCR primers 1 and 2, which are portions of the adapters 1 and 2, respectively. The products then were subjected to a second round of PCR by using the 5′ proximal primer and nested PCR primers 1 and 2. The products from the second-round PCR were run on an agarose gel, and the major bands were purified and cloned into pGEM-T/pGEM-T Easy and sequenced.

Sequence Analysis of the Cloned *S* Regions

For the *910 S* haplotype, two λ genomic clones containing *SLG* and *SRK*, respectively, were identified previously (Yu et al., 1996). The two λ clones were subcloned and sequenced. However, there was still a gap between the two of ~2.5 kb as determined by PFGE blot analysis (Yu et al., 1996). To fill in this gap sequence, we cloned the 25-kb EcoRI fragment, which covers both *SLG* and *SRK* (Yu et al., 1996), into the cosmid vector pLAFR1. The cosmid clone was sequenced directly, starting with primers located at the ends of the two λ clones, and extended by primer walking until the sequences from both sides overlapped. Several fosmid clones were identified during this work with the W1 fosmid genomic library. Their end sequences were obtained by direct sequencing of the clones with T7 and Sp6 primers present on the vector and were used to order the clones in the *S* region relative to the known sequence. Compared with the previously cloned region, the fosmid contigs extended the cloned *S* region by ~35 kb. The EcoRI restriction fragments were subcloned into pBluescript KS+ (Stratagene) and sequenced by primer walking. Sequences from different subclones were assembled via the sequences from direct sequencing of the fosmid clones, with complementary primers located within the subclones. Some other regions were not subcloned and were sequenced directly with the fosmid clones by primer walking.

For the *A14 S* haplotype, four independent fosmid clones were identified from the T2 fosmid genomic library. The fosmids were digested to completion with HindIII and cloned into pBluescript KS+. The HindIII restriction fragments ranged in size from ~0.5 to ~15 kb. Seventeen subclones that contained inserts ranging from ~0.5 to ~5 kb were sequenced directly. Five subclones, which contained inserts ranging from ~8 to ~15 kb, were further subcloned. Several enzymes were tested for each of them, and EcoRI, PstI, SacI, and XhoI then were chosen for the five, respectively. Twenty subclones, which contained inserts of ⩽6 kb, were obtained and sequenced. Sequencing of all of the subclones was initiated with the universal primers on the vector and subsequently extended by primer walking. To assemble the sequences from the subclones, we synthesized complementary primers located within the subclones, usually 100 to 200 bp away from the ends, and used them to sequence the fosmid directly. These fosmid sequences were not only used for joining the ends of subclones but also were used in at least three cases to close the small gaps missed during subcloning due to their small sizes.

Sequencing was performed mostly on one strand. For regions in which the sequences were not good, both strands were sequenced. Computer analysis was performed using the BLAST program at the National Center for Biotechnological Information (Altschul et al., 1990). Genomic regions showing significant matches (usually >100) were further studied, using GenScan for Arabidopsis, with a suboptimal exon cutoff of 1.00 (Burge and Karlin, 1998). Sequences were aligned with ClustalW (Thompson et al., 1994) by using the Blosum weight matrix, a gap opening penalty of 10.0, and a gap extension penalty of 0.05.

ACKNOWLEDGMENTS

We thank Dr. Shoba Sivasankar for critical reading of the manuscript; Dr. Yongzhong Wu and Mei Wang for help with nuclei preparation; Dr. Ung-Jin Kim of the California Institute of Technology for the fosmid vector; Angela Hollis of the Guelph Molecular SuperCentre for sequencing service; and Ian Smith for help in preparing figures. Y.C. was supported in part by an Ontario Graduate Scholarship. This work was funded by the Natural Sciences and Engineering Research Council of Canada.

Received June 25, 1999; accepted September 14, 1999.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215,** 403–410.

Ausubel, F.M., Brent, R., Kingston, R., Moore, D.M., Seidman, J.G., Smith, J.A., and Struhl, K., eds (1988). Current Protocols in Molecular Biology. (New York: Green Publishing Associates and Wiley-Interscience).

Bairoch, A. (1993). The PROSITE dictionary of sites and patterns in proteins—Its current status. Nucleic Acids Res. **21,** 3097–3103.

Bateman, A.J. (1955). Self-incompatibility systems in angiosperms. III. Cruciferae. Heredity **9,** 52–68.

Bennetzen, J.L. (1996). The contributions of retroelements to plant genome organization, function and evolution. Trends Microbiol. **4,** 347–353.

Bevan, M., et al. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana.* Nature **391,** 485–488.

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

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

Bull, J.J. (1983). Evolution of Sex-Determining Mechanisms. (Menlo Park, CA: Benjamin/Cummings).

Burge, C.B., and Karlin, S. (1998). Finding the genes in genomic DNA. Curr. Opin. Struct. Biol. **8,** 346–354.

Cabrillac, D., Delorme, V., Garin, J., Ruffio-Chable, V., Giranton, J., Dumas, C., Gaude, T., and Cock, J.M. (1999). The S_{15} self-incompatibility haplotype in *Brassica oleracea* includes three *S* gene family members expressed in stigmas. Plant Cell **11,** 971–986.

Camargo, L.E.A., Savides, L., Thormann, C.E., and Osborn, T.C. (1997). Location of the self-incompatibility locus in an RFLP and RAPD map of *Brassica oleracea.* J. Hered. **88,** 57–60.

Carthew, R.W., and Rubin, G.M. (1990). *Seven in absentia*, a gene required for specification of R7 cell fate in the Drosophila eye. Cell **63,** 561–577.

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.

Conner, J.A., Conner, P., Nasrallah, M.E., and Nasrallah, J.B. (1998). Comparative mapping of the Brassica *S* locus region and its homeolog in Arabidopsis: Implications for the evolution of mating systems in the Brassicaceae. Plant Cell **10,** 801–812.

Delorme, V., Giranton, J.L., Hatzfield, Y., Friry, A., Heizmann, P., Ariza, J., Dumas, C., Gaude, T., and Cock, J.M. (1995). Characterization of the *S* locus genes, *SLG* and *SRK*, of the Brassica S_3 haplotype: Identification of a membrane-localized protein encoded by the *S* locus receptor kinase gene. Plant J. **7,** 429–440.

de Nettancourt, D. (1977). Incompatibility in Angiosperms. (Berlin: Springer-Verlag).

Diatchenko, L., Lau, Y.-F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D., and Siebert, P.D. (1996). Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc. Natl. Acad. Sci. USA **93,** 6025–6030.

Fedoroff, N.V. (1989). About maize transposable elements and development. Cell **56,** 181–191.

Ferris, P.J., and Goodenough, U.W. (1994). The mating-type locus of *Chlamydomonas reinhardtii* contains highly rearranged DNA sequences. Cell **76,** 1135–1145.

Friedman, A.M., Long, S.R., Brown, S.E., Brikema, W.J., and Ausubel, F.M. (1982). Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. Gene **18,** 289–296.

Gierl, A. (1996). The *En/Spm* transposable element of maize. Curr. Top. Microbiol. Immunol. **204,** 145–159.

Glavin, T.L., Goring, D.R., Schafer, U., and Rothstein, S.J. (1994). Features of the extracellular domain of the *S*-locus receptor kinase from Brassica. Mol. Gen. Genet. **244,** 630–637.

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

Goring, D.R., Banks, P., Fallis, L., Baszczynski, C.L., Beversdorf, W.D., and Rothstein, S.J. (1992a). Identification of an *S* locus glycoprotein allele introgressed from *B. napus* ssp. rapifera to *B. napus* ssp. *oleifera.* Plant J. **2,** 983–989.

Goring, D.R., Banks, P., Beversdorf, W.D., and Rothstein, S.J. (1992b). Use of the polymerase chain reaction to isolate an *S*-locus glycoprotein cDNA introgressed from *B. campestris* into *B. napus* ssp. *oleifera.* Mol. Gen. Genet. **234,** 185–192.

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.

Gurskaya, N.G., Diatchenko, L., Chenchik, A., Siebert, P.D., Khaspekov, G.L., Lukyanov, G.L., Lukyanov, K.A., Vagner, L.L., Ermolaeva, O.D., Lukyanov, S.A., and Serdlov, E.D. (1996). Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: Cloning of Jurkat cell transcript induced by phytohemaglutinin and phorbol 12-myristate 13-acetate. Anal. Biochem. **240,** 90–97.

Jones, J.D.G., Dunsmuir, P., and Bedbrook, J. (1985). High-level expression of introduced chimeric genes in regenerated transformed plants. EMBO J. **4,** 2411–2418.

Kandasamy, M.K., Paolillo, D.J., Nasrallah, J.B., and Nasrallah, M.E. (1989). The *S*-locus specific glycoproteins of Brassica accumulate in the cell wall of developing stigma papillae. Dev. Biol. **134,** 462–472.

Kim, U.J., Shizuya, H., de-Jong, P.J., Birren, B., and Simon, M.I. (1992). Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucleic Acids Res. **20,** 1083–1085.

Kunst, F., et al. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis.* Nature **390,** 249–256.

Kusaba, M., and Nishio, T. (1999). Comparative analysis of *S* haplotypes with very similar SLG alleles in *Brassica rapa* and *Brassica oleracea.* Plant J. **17,** 83–91.

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 mechanism. Proc. Natl. Acad. Sci. USA **94,** 7673–7678.

Le Guen, L., Thomas, M., Bianchi, M., Halford, N.G., and Kreis, M. (1992). Structure and expression of a gene from *Arabidopsis thaliana* encoding a protein related to SNF1 protein kinase. Gene **120,** 249–254.

Letham, D.L.D., and Nasrallah, J.B. (1998). A *ClpP* homolog linked to the *Brassica* self-incompatibility (*S*) locus. Sex. Plant Reprod. **11,** 117–119.

Lewin, B. (1997). Genes VI. (New York: Oxford University Press).

Lovett, M. (1994). Fishing for complements: Finding genes by direct selection. Trends Genet. **10,** 352–357.

Meinnel, T., and Blanquet, S. (1994). Characterization of the *Thermus thermophilus* locus encoding peptide deformylase and methionyl-tRNA (formyltransferase). J. Bacteriol. **176,** 7387–7390.

Murphy, D.J., and Ross, J.H.E. (1998). Biosynthesis, targeting and processing of oleosin-like proteins, which are major pollen coat components in *Brassica napus.* Plant J. **13,** 1–16.

Nacken, W.K.F., Piotrowiak, R., Saedler, H., and Sommer, H. (1991). The transposable element *Tam1* from *Antirrhinum majus* shows structural homology to the maize transposon *En*/*Spm* and has no sequence specificity of insertion. Mol. Gen. Genet. **228,** 201–208.

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

Nasrallah, J.B., Kao, T.H., Goldberg, M.L., and Nasrallah, M.E. (1985). A cDNA clone encoding an *S*-locus specific glycoprotein from *Brassica oleracea.* Nature **318,** 617–618.

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

Nasrallah, M.E., Kandasamy, M.J., and Nasrallah, J.B. (1992). A genetically defined *trans* acting locus regulates *S*-locus in Brassica. Plant J. **2,** 497–506.

Ockendon, D.J. (1974). Distribution of self-incompatibility alleles and breeding structure in open-pollinated cultivars of Brussels sprouts. Heredity **33,** 159–171.

Ozeki, Y., Davies, E., and Takeda, J. (1997). Somatic variation during long-term subculturing of plant cells caused by insertion of a transposable element in a phenylalanine ammonia-lyase (PAL) gene. Mol. Gen. Genet. **254,** 407–416.

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

Pelissier, T., Tutois, S., Deragon, J.M., Tourmente, S., Genestier, S., and Picard, G. (1995). *Athila*, a new retroelement from *Arabidopsis thaliana.* Plant Mol. Biol. **29,** 441–452.

Pelissier, T., Tutois, S., Tourmente, S., Deragon, J.M., and Picard, G. (1996). DNA regions flanking the major *Arabidopsis thaliana* satellite are principally enriched in *Athila* retroelement sequences. Genetica **97,** 141–151.

Pereira, A., Schwarz-Sommer, Z.S., Gierl, A., Peterson, P.A., and Saedler, H. (1985). Genetic and molecular analysis of the *Enhancer* (*En*) transposable element system of *Zea mays.* EMBO J. **4,** 17–23.

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

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Nacharov, D., Melake-Berhan, A., Sringer, P.S., Edwards, K.J., Lee, M., Avramova, Z., and Bennetzen, J.L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science **274,** 765–768.

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L. (1998). The paleontology of intergene retrotransposons of maize. Nat. Genet. **20,** 43–45.

Sato, T., Thorsness, M.K., Kandasamy, M.K., Nishio, T., Hirai, M., Nasrakkah, 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.

Scott, R., Dagless, E., Hodge, R., Paul, W., Soufleri, I., and Draper, J. (1991). Patterns of gene expression in developing anthers of *Brassica napus.* Plant Mol. Biol. **17,** 195–207.

Slocum, M.K., Figdore, S.S., Kennard, W.C., Suzuki, J.Y., and Osborn, T.C. (1990). Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea.* Theor. Appl. Genet. **80,** 57–64.

Song, K.M., Suzuki, J.Y., Slocum, M.K., Williams, P.H., and Osborn, T.C. (1991). A linkage map of *Brassica rapa* synonym *campestris* based on restriction fragment length polymorphism loci. Theor. Appl. Genet. **82,** 296–304.

Stahl, R.J., Arnoldo, M., Glavin, T.L., Goring, D.R., and Rothstein, S.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.

Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H., and DeLong, E.F. (1996). Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragments from a planktonic marine archaeon. J. Bacteriol. **178,** 591–599.

Stephens, R.S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R.L., Zhao, Q., Koonin, E.V., and Davis, R.W. (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis.* Science **282,** 754–759.

Thompson, H.L., Schmitt, R., and Dean, C. (1996). Identification and distribution of seven classes of middle-repetitive DNA in *Arabidopsis thaliana* genome. Nucleic Acids Res. **24,** 3017–3022.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22,** 4673–4680.

Tomkinson, A.E., Roberts, E., Daly, G., Totty, N.F., and Lindahl, T. (1991). Three distinct DNA ligases in mammalian cells. J. Biol. Chem. **266,** 21728–21735.

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.

Ullrich, A., and Schlessinger, J. (1990). Signal transduction by receptors with tyrosine kinase activity. Cell **61,** 203–212.

von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14,** 4683–4690.

Wessler, S.R., Bureau, T.E., and White, S.E. (1995). LTR-retrotransposons and MITEs—Important players in the evolution of plant genomes. Curr. Opin. Genet. Dev. **5,** 814–821.

Wright, D.A., and Voytas, D.F. (1998). Potential retrovirus in plants: *Tat1* is related to a group of *Arabidopsis thaliana Ty3/gypsy* retrotransposons that encode envelope-like proteins. Genetics **149,** 703–715.

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

Zhang, H.-B., Zhao, X., Ding, X., Paterson, A.H., and Wing, R.A. (1995). Preparation of megabase-size DNA from plant nuclei. Plant J. **7,** 175–184.