Cell-cell communication between pollen and pistil is important for sexual reproduction in flowering plants. A compatible pollen grain hydrates and germinates on the surface of the stigma to produce a pollen tube, which then grows down through the transmitting tissue of the style to reach the ovary, where fertilization takes place. Self-incompatibility (SI) is a mechanism by which self-fertilization is prevented. In self-incompatible Brassica species, self- and nonself-pollen grains are discriminated by the papilla cell of the stigma and, as a result, self-pollen tubes cannot penetrate the papilla cell.

The SI interaction in Brassica is sporophytically controlled by a single polymorphic locus, termed the S locus (Bateman 1995). Two highly polymorphic S locus genes, SLG (encoding S locus glycoprotein) and SRK (encoding S-receptor kinase), are expressed exclusively in the papilla cell (Nasrallah et al. 1988; Stein et al. 1991) and thought to be involved in SI recognition by the stigma on the basis of analyses of transgenic plants and self-compatible mutants (Nasrallah et al. 1992, 1994; Goring et al. 1993; Shiba et al. 1995; Conner et al. 1997; Stahl et al. 1998). SRK is a membrane-spanning receptor-like kinase whose extracellular domain is similar to SLG and named the “S domain” (Stein et al. 1991, 1996; Watanabe et al. 1994; Delorme et al. 1995). The physical distance between the SLG and SRK genes is within 25 kb in self-incompatible B. napus (Yu et al. 1996) and within 20 kb in the S9 haplotype of B. campestris (Boyce et al. 1997). Because the S locus appears to be a multigene complex, “S allele” is referred to as “S haplotype” (Boyce and Nasrallah 1993; Nasrallah and Nasrallah 1993).

The structural features of SLG and SRK proteins suggest that they might function as the receptors of yet-unidentified pollen ligand(s) that determine(s) the S specificity of pollen. Binding of SLG/SRK to their ligand(s) would then elicit a signal transduction pathway involving a kinase cascade in the papilla cell, leading to the rejection of self-pollen. To elucidate the mechanism of self-pollen recognition in SI, it is imperative that the pollen ligand(s) be identified. The gene(s) encoding the pollen ligand(s) is believed to be located at the S locus and expressed sporophytically in the anther.

To date, several S-linked genes have been identified in the region downstream of the SLG genes. In the S1 haplotype of B. oleracea, SLA (for S locus anther) is located downstream of SLG2 and expressed specifically in the anther (Boyce and Nasrallah 1995). A large...
insertion in the SLA gene was identified in self-compatible lines of B. napus and B. oleracea (Boy es and Nas r a l l a h 1995; Past uglia et al. 1997), as well as in self-incompatible lines of B. oleracea (Past uglia et al. 1997), suggesting that a functional SLA gene is not required for the SI response. In the S910 haplotype of B. napus, two S-linked genes, SLL1 (for S-locus linked gene 1) and SLL2 (for S-locus linked gene 2), are located downstream of the SGL910 gene and expressed in the anthers (Yu et al. 1996). Although SLL1 and SLL2 are tightly linked to the S locus, neither is likely to be the pollen S gene because SLL1 does not show any allelic sequence difference and SLL2 is also expressed in the stigma.

Two additional S-linked genes, 298 and 299, are located in the 3′-flanking region of the SLG gene of the S1 haplotype of B. campestris (Boy es et al. 1997). The S99 gene encodes SLL2 and 298 encodes a ClpP homologue (Conner et al. 1998). These two genes are not polymorphic and are expressed in vegetative tissue. The S-linked ClpP gene also seems unlikely to function in SI recognition because of its expression pattern (L e t h a m and N as r a l l a h 1998).

We have previously characterized a downstream region of SLG in the S1 haplotype of B. campestris, and found the SLL2 gene to be located 5.8 kb from SLG9 and identified a new gene, named SAE1 (for S locus anther-expressed gene), located between SLG9 and SLL2 (Watanabe et al. 1999). In this study, we completely sequenced and characterized a 76-kb region that contained SLG and SRK of the S1 haplotype and identified 10 additional genes located at the S locus. This information will be useful for understanding the structure and organization of the complex S locus.

MATERIALS AND METHODS

Plant materials: Plant materials used were S1, S′, Sf, Sfl haplotypes of self-incompatible B. campestris, S1, S′, Sf haplotypes of self-compatible B. oleracea, and self-compatible B. napus cv. Westar.

Cloning and sequencing of a 76-kb region of the S locus: The 76-kb fragment (obtained from Mlu-digested genomic DNA of B. campestris S′ homozygotes) containing both SLG9 and SRK9 had previously been cloned into a P1-derived artificial chromosome (PAC) vector, and this PAC clone was designated E89 (Suzuki et al. 1997). The 76-kb insert of the E89 clone was digested with SpeI or BamHI and subcloned into pBluescript II plasmid vector. Other subclones were derived from λ-phage clones, A1, B1, and L1, which contained SLG9 or SRK9 (Suzuki et al. 1995).

Dideoxynucleotide sequencing was performed with double-stranded plasmid DNA templates. G + C contents of the sequence data were analyzed using GENETIX-MAC Ver. 10.0 software (Software Development Co., LTD., Tokyo). Homology searches were performed using the BLAST program (Altsch ul et al. 1990) available at the website of the National Institute of Genetics (http://www.ddbj.nig.ac.jp/searches.html) or DNASIS software (HITACHI Software Engineering, Yokohama, Japan). The possible existence of genes (exons) in the entire sequence of the 76-kb fragment was predicted by using GENSCAN program (Bur ge and Kar l in 1997) available at the website of the Pasteur Institute, Paris (http://bioweb.pasteur.fr/seqanal/interfaces/genscan.html).

Fluorescence in situ hybridization (FISH) on extended DNA fibers: Extended DNA fibers were prepared from mature leaves of B. campestris S′ homozygotes according to the protocol of Fr ans z et al. (1996) with minor modifications. The in situ hybridization procedure of O h mido and Fuk u (1997) and O h mido et al. (1998) was followed. pBIN-SLG9 and pBIN-SRK9 (Suzuki et al. 1996) were labeled with digoxigenin (dig)-11-dUTP, and E89 (Suzuki et al. 1997) was labeled with biotin-16-dUTP, using a Nick Translation Kit (Boehringer Mannheim, Mannheim, Germany). Dig-labeled pBIN-SLG9 and pBIN-SRK9 were detected using anti-dig rhodamine (Boehringer Mannheim), and the signal was amplified by Texas red-conjugated anti sheep IgG (Vector Laboratories, Burlingame, CA). For detection of the biotin-labeled E89, two layers of fluorescein avidin DCS (Vector Laboratories) and one layer of biotinylated antiavidin D (Vector Laboratories) were applied. Slides were counterstained with 4,6-diamidino-2-phenylindole (DAPI) and mounted with an antifade solution, Vectashield (Vector Laboratories).

High-resolution visual mapping of SLG9 and SRK9 on the E89 clone was carried out as follows: The E89 DNA was linearized with NotI, and the linearized DNA was mixed with dig-labeled pBIN-SLG9 and pBIN-SRK9 probes, denatured for 10 min at 60°C, and hybridized at 37°C overnight. Molecular combing (Bensimon et al. 1994) was performed without purification. Two μl of hybridized DNA was spotted onto an APS (3-aminopropyltriethoxysilane)-coated glass slide (Matsunami Glass, Japan), and the slide was covered with a coverslip (18 × 18 mm). DNA molecules were stretched and aligned according to receding air-water meniscus. The coverslip was removed with forceps and the slides were baked at 60°C for 6 hr after complete air-drying. For visualization of dig-labeled probes, two layers of antidig rhodamine (Boehringer Mannheim) and Texas red-conjugated antiseep IgG (Vector Laboratories) were used. The DNAs were counterstained with YOYO-1 (Molecular Probes, Eugene, OR) and mounted with Vectashield. Two or three fluorescent images were separately captured using a cooled CCD (charge-coupled device) camera (PXL-1400, Photometrics, Ltd., Tucson, AZ) and were merged into a single composite image.

Construction and screening of cDNA libraries: Flower buds of B. campestris S′ homozygous plants were collected and classified into four stages (stage 1 to 4; see the result section) of anther development. Poly(A) RNA was isolated from anthers of stages 2–4 and flower buds of stage 1 with a MicroFastTrack mRNAs isolation kit (Invitrogen, San Diego). cDNA synthesized from the poly(A)+ RNA using a CDNA synthesis kit (LK Instant, Uppsala, Sweden) was used for CDNA library construction in λ gt10 vector (Stratagene, La Jolla, CA). The libraries were screened by plaque hybridization with dig-labeled probes, which had been made from subclones of the E89 clone. The probes were prepared by random-primed DNA labeling using the digoxigenin DNA-labeling kit (Boehringer Mannheim). Hybridization and detection were carried out as described by Suzuki et al. (1995).

Reverse transcriptase PCR (RT-PCR): Double-strand cDNAs synthesized from poly(A)+ RNA isolated from flower buds of stage 1, anthers of stages 2 to 4, leaves, and pistils of stage 3 were used as templates for PCR amplification with primers specific to each S-linked gene. PCR was performed with Taq DNA polymerase (TaKaRa, Shuzo, Shiga, Japan) in a DNA thermal cycler (Perkin-Elmer, Norwalk, CT). Each of the 35 cycles consisted of 1 min of denaturation at 94°C, 1 min of annealing at 60°C, and 2 min of extension at 72°C (except for the last cycle in which the extension was for 7 min).

RNA gel blot analysis: Total RNA was isolated from anthers,
stigmas, and mature pollen grains of the S\(^t\) haplotype of B. campestris with ISOGEN (Nippongene, Tokyo). After denaturation in glyoxal, 10 \(\mu\)g of RNA was fractionated by electrophoresis on a 1.0\% agarose gel and transferred to nylon membranes (Nytran; Schleicher & Schuell, Dassel, Germany). Hybridization and detection were carried out as described by Suzuki et al. (1996). Loading of equal amounts of RNA was checked by staining of tRNA with ethidium bromide. These results suggest that the direction of transcription of the SLG and SRK genes is not conserved among S haplotypes. The SLG\(^{\text{t}}\) gene was located near the middle of the 76-kb region, so the E89 clone contained \(~15\) kb upstream of SRK\(^{\text{t}}\) and \(~37\) kb downstream of SLG\(^{\text{t}}\).

Dig-labeled pBIN-SLG9 and pBIN-SRK9 clones containing SLG\(^{\text{t}}\) and SRK\(^{\text{t}}\) genomic clones, respectively (Suzuki et al. 1996), and the biotin-labeled E89 clone were hybridized to extended DNA fibers (EDFs) of the genomic DNA of S\(^t\) homozygotes (Figure 2A); red signals corresponded to SLG\(^{\text{t}}\) and SRK\(^{\text{t}}\), and green signals corresponded to the entire 76-kb region contained in the E89 clone. The pBIN-SLG9 and pBIN-SRK9 regions are indicated on the schematic map of the E89 clone (Figure 2C). It should be noted that the degrees of stretching of the two EDFs in Figure 2A were not the same. We also hybridized dig-labeled pBIN-SLG9 and pBIN-SRK9 to the E89 clone extended with molecular combing technique (Figure 2B). The green fluorescent

![Figure 1](image1.png)

**Figure 1.** Gene map of the 76-kb genomic fragment of E89 clone of the S\(^t\) haplotype of Brassica campestris. (A) Location of 14 genes, a transposon-like sequence and three open reading frames (ORFs). Vertical bars indicate positions of SpI sites. Striped boxes at both ends of the map indicate regions of the PAC vector adjacent to the 76-kb genomic fragment. Filled boxes denote locations of the 14 genes, and horizontal arrows above the genes indicate directions of their transcription. Three long ORFs (a, b, and c) and a transposon-like sequence are represented by open boxes labeled with “a,” “b,” “c,” or “Transposon.” Open circles show positions of the box elements that are conserved in the SLG and SRK promoters. Thick bars below the map indicate the regions of the 76-kb fragment used as a probe for cDNA library screening. (B) G + C contents (%) of the E89 region. The average span used for calculation is 500 bp. The dotted line represents 40\% G + C content. The A + T-rich region in the SRK\(^{\text{t}}\) gene corresponds to the first and second introns. (C) Predicted protein coding regions. Boxes indicate the positions of potential exons as predicted by the GENSCAN program; arrows indicate the direction of transcription of the predicted genes. The boxes shown above and below the horizontal line correspond to genes that are transcribed in the direction indicated by the corresponding arrow.

![Figure 2](image2.png)

**Figure 2.** Visual mapping of SRK\(^{\text{t}}\) (pBIN-SRK9: 11 kb) and SLG\(^{\text{t}}\) (pBIN-SLG9: 7 kb) in the E89 region. (A) Two streams of the FISH signals of SLG\(^{\text{t}}\) and SRK\(^{\text{t}}\) (red) and the E89 (green) on the extended DNA fibers of B. campestris genomic DNA of S\(^t\) homozygotes at varying degrees of stretching. Signals of SLG\(^{\text{t}}\) and SRK\(^{\text{t}}\) appear as single or a few dots. (B) High-resolution mapping of E89 PAC clone by molecular combing. Red fluorescent signals at both ends indicate vector DNA sequences that allow evaluation of the entire sequence of the E89 region. Other major red signals indicate the position of SLG\(^{\text{t}}\) and SRK\(^{\text{t}}\), and a minor signal indicates the SLG/ SRK-promoter-like sequence (box elements) that was identified upstream of SRK\(^{\text{t}}\). Stretching degree of the DNA molecule by this combing method (~3 kb/\(\mu\)m) is close to Watson and Crick’s DNA model. (C) Cloning map of the E89 region. Bars, 5 \(\mu\)m (A, B); 15 kb (C).
signal represented the YOYO-1 stained extended DNA fiber of the E89 clone, and the two strong red signals corresponded to SLG\(^8\) and SRK\(^9\). Because the labeled pBIN-SLG\(^9\) and pBIN-SRK\(^9\) DNA contained the vector and promoter sequences, red signals were also observed at both ends of the E89 clone (vector) and in a duplicated SLG/ SRK-promoter-like region (promoter) identified 7.5 kb upstream of SRK\(^9\) (see the last section of results). The positions of SLG\(^8\) and SRK\(^9\) on the E89 clone as detected by EDF-FISH (Figure 2B) were consistent with those shown on the restriction map of the E89 clone (Figure 2C), and with those obtained by the EDF-FISH analysis using the S\(^9\) genomic DNA (Figure 2A).

**Genes located in the flanking regions of SLG\(^8\) and SRK\(^9\):** To identify and isolate all the anther-expressed genes located in the 76-kb genomic fragment, we screened five different cDNA libraries using as a probe DNA fragments spanning all except the previously analyzed regions (Watanabe et al. 1999) of this fragment (Figure 1A). One of the cDNA libraries screened was a cDNA library of anthers at the uninucleate stage that had previously been constructed by Kitashiba and Toriyama (1997). The other four cDNA libraries were constructed in this study from anthers (or flower buds) of four different developmental stages (stage 1 to stage 4), covering all stages except the stage of mature anthers from open flowers. Definition of the anther developmental stages is as follows: stage 1 anthers were from flower buds <2.0 mm in length and contained microspores at the tetrad stage; stage 2 anthers were from flower buds 2.0-5.0 mm in length and contained microspores at the uninucleate and binucleate stages; stage 3 anthers were from flower buds 5.0-6.5 mm in length and contained microspores at the binucleate stage; stage 4 anthers were from flower buds >6.5 mm in length and contained microspores at the binucleate and trinucleate stages. The library of the earliest stage (stage 1) was made from flower buds because of the difficulty of isolating anthers; the libraries of stages 2, 3, and 4 were all made entirely from anthers. Seven genes, SP1, SP2, SP3, SP4, SP8, SP10, and SP11 (S-locus protein 1, 2, 3, 4, 8, 10, and 11, respectively), were identified from the screening of 100,000 cDNA clones of the uninucleate library, 400,000 clones of the stage 1 library, 200,000 clones of the stage 2 library, and 100,000 clones each of the stages 3 and stage 4 libraries. We also screened a stage 3 pistil cDNA library (total 100,000 clones) using the same probe, and found three additional genes, SP5, SP6, and SP7 (S-locus proteins 5, 6, and 7, respectively).

Table 1 shows the characteristics of the 14 genes located in the E89 region. Database searches revealed that five of them, SP2, SP5, SP7, SP10, and SP11, showed significant sequence homology to known proteins. SP2 was a ClpP homologue, which had been identified previously in the S\(^8\) haplotype of B. campestris (Le et al. 1998). Eight amino acid differences were found between ClpP-S\(^8\) and ClpP-S\(^1\) (SP2). SP7 was homologous to yeast antisilencing protein ASF1 (Le et al. 1997). ASF1 in yeast has been identified as the gene that causes derepression of the silent mating-type loci when overexpressed. The predicted SP7 protein was similar to the N-terminal domain of ASF1, but did not have the C-terminal acidic domain characteristic of ASF1. SP10 was homologous to the En/Spm-type transposon (see the last section of results).

The predicted amino-acid sequence of SP5 contained a domain similar to the J domain of Dnaj, the Escherichia coli ortholog of cochaperone Hsp40 (Kelley 1998). This domain of SP5 had the following features of the J domain: four predicted helices (I to IV); hydrophobic residues on helices I, II, and III; and a HPD (histidine-proline-aspartic acid) tripeptide in the loop between helices II and III (Figure 3A). Dnaj and related co-chaperones facilitate protein folding through their interactions with members of the Hsp70 class of chaperones, and the J domain is thought to mediate interaction with Hsp70. The putative J domain of SP5 was located in the middle of the sequence and the rest of SP5 showed no significant homology to Hsp40, indicating that SP5 is a member of a subclass of the Hsp40 family that contains only the J domain. The sequence homologous to SP5 was found in YUP8H12R YAC clone of Arabidopsis thaliana, which had been submitted to database (accession no. AC002986). The YUP8H12R YAC clone was known to map at the bottom of the right arm of chromosome 1 between the ATAHATPAS and m532 markers, which mapped >20 cm away from the S-locus homologous region of A. thaliana (Conner et al. 1998).

SP11 encoded a small cysteine-rich protein with a putative signal peptide of 24 hydrophobic amino acid residues at its N terminus. The predicted mature protein (59 amino acids, 6.7 kD, pl 9.08) contained eight cyste-
Genomic Organization of the S Locus

Figure 3.—Amino-acid sequence alignment of the J domain of (A) SP5 with Escherichia coli DnaJ, and (B) SP11 with Brassica pollen coat protein. Colons and periods indicate identical and similar amino acids, respectively. Gaps introduced to maximize similarity are indicated by dashes. (A) The conserved HPD tripeptide is indicated by asterisks. The regions corresponding to the four helices of DnaJ are indicated by filled boxes below the alignment. The conserved sequences in Helix III are indicated by a bar above the alignment. (B) The putative signal peptides are denoted with boldface italics; arrowheads show the positions of the introns. The eight cysteine residues in the mature proteins are boxed.

ine residues, characteristic of members of the pollen coat protein (PCP) family of Brassica (Stanchev et al. 1996; Doughty et al. 1998a). Alignment of SP11 and PCP-A1 showed that seven of the eight cysteines were conserved between them (Figure 3B). Similar to two PCP genes, PCP-A1 and PCP1, SP11 contained an intron located in the region encoding the signal peptide; the location of the intron is conserved among these three genes. An alternative transcript of SP11, in which the intron had not been spliced out, was also identified in the cDNA cloning process (data not shown). From database searches, SP11 was also found to be similar to pl230 disease-resistance response protein of pea (40% similarity) and AmbtV pollen allergen of giant ragweed (50% similarity), both of which are small cysteine-rich proteins (Chiang and Hadwiger 1991; Ghosh et al. 1991).

All the genes we isolated were indeed located in the E89 region because their cDNA sequences were identical to the corresponding genomic sequences except for the intron sequences. The location and direction of transcription of these SP genes are shown in Figure 1A: SP5, SP6, and SP7 were located upstream of SRK³, SP11 was located between SRK³ and SLG³, and the rest were located downstream of SLG³. The ClpP homolog (SP2) was located immediately downstream of SSL2, which was located in the 3'-flanking region of SLG³. This gene arrangement is similar to that found in the S¹ haplotype of B. campestris (Boyes et al. 1997; Letham and Nasrallah 1998). In the S¹₀ haplotype of B. napus, another S-locus-linked gene, SSL1, is located between SLG¹₀ and SSL2 (Yu et al. 1996). Although there was no SSL1 gene at the corresponding position in the S³ haplotype, a fragmented SLL1-related sequence was observed between SAE1 and SLL2 (data not shown).

Expression of the S-linked genes: To investigate the expression of the genes located in the E89 region, we performed RT-PCR analysis for the 10 newly identified genes as well as for SLG/SRK and SAE1 (Figure 4). The amplified DNA fragments did not correspond to intron-containing genomic DNA fragments of SP2, SP3, and SP5 (data not shown), suggesting that the templates used for PCR amplification were not contaminated with genomic DNA. In the cases of SP2, SP3, SP7, and SP10, cDNA was amplified strongly from all the tissues examined: flower buds, anthers, pistils, and leaves. The ubiquitous expression of these genes suggests that they might have a housekeeping function. Amplified fragments of SP1 and SP4 cDNAs were detected in all tissues; however, for SP1, the intensity of the band detected in leaves was the lowest, and, for SP4, the intensity of the band...
detected in pistils was the lowest. For SP5, SP6, SP8, and SP11, no amplified fragments were detected in leaves. The fragments corresponding to the SP5 transcripts in flower buds of stage 1 and anthers of stage 2 were much less intense than in pistils of stage 3, suggesting predominant pistil expression of SP5. The SP6 and SP8 genes were expressed in pistils and anthers at early stages. Expression of SP6 and SP8 in the anther appeared to be developmentally regulated: the intensity of amplified fragments decreased at stage 3 and/or 4. The SP11 gene was expressed in anthers at stages 2, 3, and 4, but not in flower buds at stage 1, indicating the possibility that it might be a pollen-expressed gene. The amplified fragment for SP11 was also detected in pistils, but the intensity of the amplified fragment was less than that in anthers.

RNA gel blot analyses of the SP5 and SP11 genes were conducted using their corresponding cDNA clones as probes to determine whether these two genes showed stigma-specific or anther-specific expression. Transcripts of SP5 were detected in stigmas at late developmental stages, but not in anthers of any stage (Figure 5). These results suggest that SP5 is expressed specifically in stigmas, as are SLG and SRK. On the contrary, SP11 showed anther-specific expression (Figure 5): a 0.5-kb transcript was detected in anthers at stages 2, 3, and 4, with the level of the transcript being the highest in anthers at stage 3, but not in flower buds of stage 1, mature pollen grains, or stigmas. This expression pattern suggests that SP11 is expressed in anthers, as is PCP-A1 (Doughty et al. 1998a). The detection of the amplified fragment of SP11 cDNA in pistils (Figure 4) would be due to high sensitivity of RT-PCR. Longer transcripts of SP11 detected in anthers at stage 3 might correspond to the alternatively spliced transcript that was observed in the cDNA cloning.

**Figure 5.**—RNA gel blot analysis of SP5 and SP11 genes. Total RNA was isolated from flower buds of stage 1 (1), anthers and stigmas of stage 2, 3, and 4 (2, 3, and 4, respectively), and mature pollen grains (Po) of the B. campestris S9 haplotype. The cDNA clones of SP5 and SP11 were used as probes. The bottom panel shows the results of ethidium bromide-stained rRNA.

**Figure 6.**—Genomic DNA gel blot analyses of SP2, SP5, and SP11 genes. Total DNA was isolated from S8 (8), S9 (9), S48 (48) haplotypes of B. campestris (B.c.); S3 (3) and S4 (4) haplotypes of B. oleracea (B.o.); and self-compatible B. napus (B.n.) cv. Westar (W). Two micrograms of each DNA sample was digested with EcoRI and subjected to DNA gel blot analysis.

*Linkage genes in other S haplotypes:* Genomic DNA gel blot analyses using the cDNA clones as probes were carried out to examine the existence of the SP genes in S1, S3, and S48 haplotypes of B. campestris, in S1 and S4 haplotypes of B. oleracea, and in B. napus (cv. Westar). SP1, SP2, SP4, SP7, and SP8 cDNA probes each hybridized to a single polymorphic band in these haplotypes; the results for SP2 are shown in Figure 6. Therefore, each of these five genes appeared to be a single-copy gene in Brassica. SP3, SP5, and SP6 cDNA probes each hybridized to several polymorphic bands; the results for SP5 are shown in Figure 6. Thus, these genes belong to polymorphic multigene families. In the case of SP11, two strong bands (13 kb and 5.6 kb) were detected in the S4 haplotype of B. campestris, and a weak nonpolymorphic band (4.4 kb) was detected in all three haplotypes of B. campestris (Figure 6). No bands were detected in B. oleracea and B. napus. Of the two strong bands unique to the S4 haplotype, the 5.6-kb band contained the SP11 gene as inferred from the analysis (E989) and the 13-kb band probably contained an SP11-like sequence. The restriction fragment length polymorphism (RFLP) analysis of an F2 population (15 plants) segregating for S1 and S48 haplotypes showed that both the 13-kb fragment and the 5.6-kb fragment cosegregated with the S4 genotype (data not shown). Thus, the SP11-like
sequence contained in the 13-kb fragment is also linked to the S locus. Absence of the SP11 and SP11-like bands in the other haplotypes examined suggests that either the SP11 gene is highly polymorphic or it is not present in the other haplotypes. The SP10 probe hybridized with multiple fragments, resulting in smear signals (data not shown).

**Genomic structure of the S locus:** To dissect the genomic structure of the S locus, the sequence of the entire 76-kb fragment of the E89 clone was determined. Based on the sequence, the size of this fragment was 76,025 bp with an average G + C content of 32.6%. The G + C content across the entire sequence is shown in Figure 1B. The G + C contents of intergenic sequences were relatively low, and the exons of the genes were generally located in regions with >40% G + C contents.

A long open reading frame (ORF) homologous to the En/Spm-type transposon was identified in the region where the SP10 gene was located (Table 2). This S-linked transposon-like sequence was similar to Ps1 Ptta' of Petunia hybrida (74% similarity; Napoli and Salamini 1998). Inverse repeats (CACTACAAGAAAA), which might represent the borders of the transposon-like element, were identified at positions 54,171 bp–54,183 bp and 61,989 bp–62,001 bp from the SP7-end of the 76-kb region, suggesting that the transposon-like sequence might potentially encompass ~10% of the region (8 kb; Figure 1A). There were no retrotransposon-like sequences found in the 76-kb region.

Other long ORFs were identified in the region between SP5 and SP6 (ORF-a), in the region between SP2 and SP10 (ORF-b), and downstream of SP11 (ORF-c; Figure 1A, Table 2). The ORF-a and ORF-b were each located in a region with a high G + C content, and their predicted amino acid sequences showed no significant homology to any known proteins in the databases. A part of the predicted amino acid sequence of ORF-c was similar to an ORF contained in the BAC clone F21J9 of Arabidopsis (accession no. AC000103-24), but not to any other known proteins in the databases.

We also used the gene identification program, GENSCAN, to predict the locations of the genes in the E89 region. The results are shown in Figure 1C. All the genes we isolated on the basis of their corresponding cDNAs, except SP11, were found to be located in the regions that the GENSCAN program predicted as protein-coding regions. The regions corresponding to the transposon-like sequence and the three long ORFs (ORF-a, ORF-b, and ORF-c) were also predicted as protein-coding regions. There are probably no other protein-coding genes in the E89 clone.

Five box elements (boxes I to V), which are conserved in the stigma-specific SLG, SRK, and SLR1 (S-locus-related gene 1) promoters (Dzelskalns et al. 1993; Delsol et al. 1995; Suzuki et al. 1995; Stein et al. 1996), were identified 2.5 kb upstream of the SP5 gene as well as just upstream of SLG$^9$ and SRK$^9$. These box elements were more similar to the boxes of SRK$^1$ or SRK$^9$ of B. oleracea than to those of SLG$^9$ and SRK$^9$ in B. campestris (data not shown). No ORFs of significant length were observed just downstream of the boxes located upstream of the SP5 gene. This SLG/SRK-promoter-like sequence might affect the stigma-specific expression of the SP5 gene. Alternatively, it might be a trail of duplication events that have occurred multiple times in the Brassica genome (Song et al. 1991; Kianian and Quiros 1992; Truco and Quiros 1994).

**DISCUSSION**

We have extensively characterized a 76-kb chromosomal region (contained in a PAC clone named EB89) of the S$^6$ haplotype of B. campestris, which we previously found to contain SLG$^9$, SRK$^9$, SLL2, and SAE1 genes. From this study, we have identified 10 additional genes located in the S-locus complex. The average gene density in this region is one gene for every 5.4 kb. This density is similar to that reported for A. thaliana: 1 gene per 4.8 kb in a 1.9-Mb region of chromosome 4 and 1 gene per 4.7 kb in a 1.6-Mb region of chromosome 5 (Satoh et al. 1997; Bevan et al. 1998). Because the genome size of B. campestris is ~3 times that of A. thaliana (Arumuganathan and Earle 1991), the expected average gene density for B. campestris is 1 gene per 15 kb, if Brassica has the same number of genes as Arabidopsis. This calculation suggests that the S locus is located in a gene-rich region of the B. campestris genome. Such "gene space" is often observed in plant genomes (Bar a-

### TABLE 2

**Characteristics of the long ORFs located in the E89 region**

<table>
<thead>
<tr>
<th>ORF</th>
<th>Accession no.</th>
<th>Location$^a$</th>
<th>Similar protein in databases$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transposon-like</td>
<td>AB022082</td>
<td>54,606–56,717</td>
<td>Petunia Ps1 PttA'</td>
</tr>
<tr>
<td>ORF-a</td>
<td>AB022079</td>
<td>5,212–6,495</td>
<td>None</td>
</tr>
<tr>
<td>ORF-b</td>
<td>AB022080</td>
<td>50,666–51,280</td>
<td>None</td>
</tr>
</tbody>
</table>

$^a$ Location (bp) is numbered from the end of the SP7-end of the E89 insert.

$^b$ Database search was performed by BLAST program (http://www.ddbj.nig.ac.jp/searches.html).
kat et al. 1997, 1998). Gene clusters are often separated by gene-poor regions, and transposons have been found in the intergenic regions of gene clusters in the genome of some species in Poaceae. In the present study, we have also identified a transposon-like sequence encompassing ~8 kb of the 76-kb region.

Sequence determination of the entire 76-kb SLG/ SRK region has revealed that the average G + C content in this region is ≈33%, similar to that of the Arabidopsis genome that has an average G + C content of 35.8% in a 1.6-Mb sequence of chromosome 5 (Sato et al. 1997). Most of the genes in the E89 region are located in G + C rich regions. This is similar to the finding in Arabidopsis (Sato et al. 1997; Bevan et al. 1998) and humans (Guigo and Fickett 1995). In this respect, the 14 genes we have identified on the basis of their expression in anthers and/or pistils and the three ORFs (ORF-a, ORF-b, and ORF-c) most likely account for all the genes in the 76-kb region, because the rest of this region is A + T rich and not predicted to contain exons by the GENSCAN program. However, we cannot completely rule out the possibility that there are other genes in the A + T rich regions. In human genome, there is some evidence that A + T-rich isochores favor genes that are more tissue specific and less abundantly expressed, while G + C-rich isochores tend to contain housekeeping genes expressed constitutively (Bernardi 1989; Mouchiroud et al. 1991). In fact, the SP11 gene, which shows anther-specific expression, is located in an A + T-rich region (Figure 1).

The genomic DNA gel blot analysis for the S-linked genes shows that all the genes except SP10 and SP11 may exist in other S haplotypes of B. campestris, and in B. oleracea and B. napus. Our preliminary results of DNA gel blot analysis using pulsed-field gel electrophoresis (PFGE) indicate that some of the SP genes are also tightly linked to the S locus in other S haplotypes of B. campestris and in some S haplotypes of B. oleracea (G. Suzuki, M. Watanabe and T. Nishio, unpublished results). Thus, the genes in the E89 region may be used as markers for physical mapping of the S locus in other S haplotypes. Information on the nature of the S-linked genes from different S haplotypes and their relative placement at the S locus will be useful for understanding the evolution of the highly polymorphic S locus. In the present study, we have also physically mapped the SLG and SRK loci by the EDF-FISH analyses. By using the DNA-combing technique, we have been able to visualize and analyze single DNA molecules individually. The visual mapping system by EDF-FISH and the DNA-combing method will be useful for physical mapping, especially for rapid analyses of newly cloned large fragments.

Four of the SP genes identified in this study (i.e., SP5, SP6, SP8, and SP11) and the SAE1 gene previously identified (Watanabe et al. 1999) appear to be expressed specifically in reproductive organs. Therefore, the gene cluster around the S locus, in addition to controlling SI, may also be involved in developmental processes and/or cell-cell interactions in the reproductive organs. It is noteworthy that genes controlling floral traits have been mapped near the S locus in tomato, suggesting a possible gene complex participating in reproductive functions (Bernardi and Tanksley 1997), although the tomato S locus is not homologous to the Brassica S locus in the evolutionary sense. SP2, SP5, SP7, and SP11 are similar to known genes in the databases (Table 1). However, further study of these genes is necessary to determine their physiological functions and to reveal any additional functional roles the S-locus gene cluster may have.

The deduced amino acid sequence of SP5 contains the J domain, which is known to interact with Hsp70 chaperone (Kelley 1998). The J domain protein-Hsp70 complex associates with specific substrate(s) to carry out cellular functions using the chaperone activities. Because SP5 is expressed exclusively in the stigma and located near the SI genes, it may function in the stigma during cell-cell communication between the male and female organs, if genes of related functions are clustered together. In addition, it is of considerable interest that the genes encoding two chaperone-related proteins, the J-domain protein (SP5) and ClpP (SP2), are tightly linked to the S locus. The mechanism of SI recognition possibly requires the aid of chaperones to accomplish the complex signal transduction.

The SP11 gene encodes a small cysteine-rich protein that has the characteristics of the PCP family proteins localized on the surface of Brassica pollen. Proteins in PCP-A and PCP-B classes, some of which are known to bind SLG or SLR1 protein, have eight conserved cysteine residues; however, these cysteines are not conserved across different classes (Doughty et al. 1998a,b). SP11 most likely belongs to a novel class of the PCP proteins: (i) as with PCP proteins, SP11 is a basic protein and consists of a similar number of amino acids; (ii) similar to PCP proteins, SP11 has eight cysteine residues in the predicted mature protein; (iii) SP11’s gene is expressed predominantly in the anther at late developmental stages; and (iv) SP11’s gene contains an intron located at the same position as that of PCP genes. The pollen determinant of the SI recognition in Brassica is thought to be a small protein of the PCP family localized in the pollen coat (Stephenson et al. 1997). One of the most interesting findings in this study is that a PCP-like gene is tightly linked to the S locus and located just downstream from the SRK2 gene.

One of the goals of our characterizing a chromosomal region containing SLG and SRK is to identify the pollen gene(s) involved in SI. Among the 14 genes we have identified in the E89 region, only SAE1 and SP11 could be potential candidates for the pollen S gene(s). The expression pattern of SAE1 (Watanabe et al. 1999) is consistent with the predicted sporophytic expression of the pollen S gene(s) (Heslop-Harrison 1975), al-
though it may not encode a small protein. SP11 encodes a small cysteine-rich PCP-like protein which, as stated above, has been thought to be a likely candidate for the pollen determinant of the SI recognition. Furthermore, both SP11 and SAE1 are located in the immediate vicinity of SRK and SLG, respectively; SP11 is located in the immediate 3'-flanking region of SRK, and SAE1 is located in the immediate 3'-flanking region of SLG. Further analyses of the spatial expression patterns of SP11 and SAE1, cloning of allelic variants of these two genes for comparison of allelic sequence differences, and ultimately in vivo functional study using transgenic plants will have to be carried out to determine whether either of these two genes functions as the pollen S gene(s).

Self-incompatible lines of S1 and S4 homozygotes of B. oleracea were kindly provided by Dr. Dave Astley in Horticultural Research International. The authors thank Professor Teh-hui Kao for his critical reading of the manuscript. G.S. is a recipient of Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists. This work was supported in part by Grants-in-Aid for Special Research Areas (nos. 07281101, 07281102, and 07281103; Genetic Dissection of Sexual Differentiation and Pollination Processes in Higher Plants) from the Ministry of Education, Science, Culture and Sports, Japan.

LITERATURE CITED


Ohmido, N., and K. Fukui, 1997 Visual verification of close dispo-
tion between a rice A genome-specific DNA sequence (TrsA) and the telomere sequence. Plant Mol. Biol. 35: 963–968.


Communicating editor: M. K. Uyenoyama