Comparative Mapping of the Brassica S Locus Region and Its Homeolog in Arabidopsis: Implications for the Evolution of Mating Systems in the Brassicaceae

Joann A. Conner,1 Patrick Conner,2 Mikhail E. Nasrallah, and June B. Nasrallah3
Section of Plant Biology, Division of Biological Sciences, Cornell University, Ithaca, New York 14853

The crucifer family includes self-incompatible genera, such as Brassica, and self-fertile genera, such as Arabidopsis. To gain insight into mechanisms underlying the evolution of mating systems in this family, we used a selective comparative mapping approach between Brassica campestris plants homozygous for the S8 haplotype and Arabidopsis. Starting with markers flanking the self-incompatibility genes in Brassica, we identified the homeologous region in Arabidopsis as a previously uncharacterized segment of chromosome 1 in the immediate vicinity of the ethylene response gene ETR1. A total of 26 genomic and 21 cDNA markers derived from Arabidopsis yeast artificial and bacterial artificial chromosome clones were used to analyze this region in the two genomes. Approximately half of the cDNAs isolated from the region represent novel expressed sequence tags that do not match entries in the DNA and protein databases. The physical maps that we derived by using these markers as well as markers isolated from bacteriophage clones spanning the S8 haplotype revealed a high degree of synteny at the submegabase scale between the two homeologous regions. However, no sequences similar to the Brassica S locus genes that are known to be required for the self-incompatibility response were detected within this interval or other regions of the Arabidopsis genome. This observation is consistent with deletion of self-recognition genes as a mechanism for the evolution of autogamy in the Arabidopsis lineage.

INTRODUCTION

The crucifer family (Brassicaceae) is well suited for the study of the evolution of plant mating systems. It includes self-fertilizing species, such as members of the genus Arabidopsis, as well as outcrossing species, such as members of the genus Brassica, in which self-pollination is prevented by a self-incompatibility (SI) system. In this family, SI is controlled genetically by haplotypes of a multifunctional gene complex, the S locus, which may span several hundred kilobases of DNA (Boyes et al., 1997). Among the genes contained at the locus are two highly polymorphic genes that encode proteins expressed specifically in the epidermal cells of the stigma, which is a structure at the tip of the female reproductive organ that is specialized for capturing and interacting with pollen. The two genes encode the cell wall-localized S locus glycoprotein (SLG) and the plasma membrane-spanning receptor protein kinase SRK (S locus receptor kinase). The predicted extracellular domain of the SRK shares extensive sequence similarity with the SLG (reviewed in Nasrallah and Nasrallah, 1993; Nasrallah et al., 1994b).

The SLG and SRK cell surface receptors fulfill an SI-specific function because plants carrying mutations that disrupt or downregulate the SLG and SRK genes are self-fertile (M.E. Nasrallah et al., 1992; Goring et al., 1993; J.B. Nasrallah et al., 1994a; Conner et al., 1997). In addition to the stigmatic SLG and SRK genes, the S locus complex is thought to encode at least one gene required for the SI phenotype in pollen. Indeed, S locus-encoded sequences that are expressed specifically in anthers have been isolated (Boyes and Nasrallah, 1995; Yu et al., 1996), although a role for these genes in the SI response has not been established.

Current evidence supports the view that SI evolved multiple times during the diversification of flowering plants and that SI has a single origin within any given plant family (Weller et al., 1993; Uyenoyama, 1995), with its recognition genes having been recruited early in the family lineage by duplication and modification of preexisting genes that perform other functions in the plant. These conclusions are based on the lack of relatedness of S locus genes from different plant families and on the observation that these genes typically belong to large gene families that include members expressed in nonreproductive tissues (Walker, 1993; Dwyer et al., 1994; Green, 1994). The emergence of SI would thus represent a gain of a self-recognition function(s) in plants already containing the full suite of genes required for successful pollen–pistil interactions and fertilization.
It is generally assumed that within plant families that include both outcrossing and self-fertilizing genera, SI is the ancestral mating system and self-fertility the derived condition (Stebbins, 1957; Lewis and Crowe, 1958; de Nettancourt, 1977; Charlesworth, 1985). This conclusion remains debatable because the mechanisms underlying transitions in mating system have not been resolved. Nevertheless, mathematical modeling has shown that under appropriate conditions (i.e., absence of inbreeding depression), a gene causing a plant to self-fertilize could spread rapidly in an initially outbreeding population (e.g., see Fisher, 1941; Charlesworth and Charlesworth, 1979; Lande and Schemske, 1985). Thus, the evolution of autogamy from SI has been proposed to occur through replacement of SI alleles at the S locus by alleles that confer self-fertility (Lewis and Crowe, 1958) or as a result of the emergence of alleles at modifier loci that affect the strength of SI (Charlesworth et al., 1990; Levin, 1996). However, molecular data pertaining to such an evolutionary transition and its genetic basis are not available.

To address this issue, we conducted a high-resolution comparative analysis of the organization of the S locus region in one genotype of B. campestris (synonym B. rapa) and a region of the Arabidopsis genome that we identified as being homeologous to the Brassica S locus. This comparative approach was spurred by the fact that extensive homeologies and microcolinearity (microsynteny) of markers are often found between the genomes of related species in animal systems (Dietrich et al., 1995) as well as in plants (Tankkelsy et al., 1992; Ahn et al., 1993; Dunford et al., 1995; Kilian et al., 1995; Chen et al., 1997) and by recent successes in the use of comparative mapping to facilitate the isolation and analysis of loci of interest (e.g., see Darling and Abbott, 1992). In the case of Arabidopsis and Brassica spp., it has been shown that the two genomes do contain islands of conserved organization, despite the fact that the Brassica genome is highly duplicated and rearranged relative to Arabidopsis (Kowalski et al., 1994; Lagercrantz et al., 1996; Sadowski et al., 1996). The results of our comparative study revealed a high degree of microsynteny between the Brassica S locus region and its homeolog in Arabidopsis, with the notable exception that sequences related to the Brassica SI genes were missing from this interval of the Arabidopsis genome. Deletion of self-recognition genes is proposed as one mechanism for the evolution of autogamy in the Arabidopsis lineage.

RESULTS

Identification of an Arabidopsis Chromosomal Segment as a Homeolog of the Brassica S Locus Region

Although it is not evident that a self-fertile crucifer, such as Arabidopsis, would contain a structurally intact S locus, we reasoned that markers flanking genes required for SI in Brassica might be conserved in Arabidopsis. On the further assumption of microsynteny for these markers on a sub-megabase scale, we proceeded to locate the Arabidopsis chromosomal region that is homeologous to the region encompassing the S locus in Brassica and to isolate clones spanning the region from available bacteriophage and large-insert genomic libraries of Arabidopsis DNA.

We had reported previously on a λ chromosome walk in the B. campestris S8 haplotype (Boyes et al., 1997). The cloned region was subsequently extended, as shown in Figure 1, to generate two λ contigs (hatched boxes in Figure 1) separated by a region of <15 kb (filled box in Figure 1) that we were unable to clone (Boyes et al., 1997). One contig spans a region of 75.5 kb, containing SLG and 3′ flanking DNA, to single-copy marker sp16; a second contig spans 48 kb, including SRK and 3′ flanking DNA, to single-copy marker sp400.

In an initial comparative study of Brassica and Arabidopsis, the single-copy probes sBH2.0, s298, s299, and s40.2-1.7 (the s prefix indicates their derivation from the Brassica S locus) were found to hybridize with unique restriction fragments on gel blots of Arabidopsis DNA (data not shown). In addition, DNA sequence analysis demonstrated that the Brassica s298 marker (Letham and Nasrallah, 1998) and s299 marker (A.L. Casselman and J.B. Nasrallah, unpublished data) were 86 and 83%, respectively, identical to their Arabidopsis counterparts within their protein-coding regions. Screening of an Arabidopsis λGEM11 genomic library with these probes resulted in the isolation of these sequences on overlapping λ clones that together comprised a 40-kb contig. Thus, these markers are clustered in Arabidopsis as they are in Brassica. These results suggest that this Arabidopsis genomic region is homeologous to the Brassica S locus region.

A single-copy end probe derived from this Arabidopsis λ contig identified a HinfI restriction fragment length polymorphism (RFLP) between the Columbia and Landsberg erecta ecotypes of Arabidopsis and was used to map this region with the Lister and Dean recombinant inbred lines (Lister and Dean, 1993; see Methods). This single-copy marker mapped to chromosome 1 between markers m315 and g4026. This is a region of the Arabidopsis genome that showed synteny to linkage groups 1 and 6 of B. oleracea in a previous comparative mapping study (Kowalski et al., 1994). This chromosomal location is distinct from the map positions of members of the Arabidopsis S gene family that were isolated by hybridization with Brassica SLG and SRK probes (Dwyer et al., 1992, 1994; Tobias et al., 1992). Based on mapping with the recombinant inbred lines, only ARK1 and ARK2, two closely linked and vegetatively expressed genes, were found to be located on chromosome 1 (Tobias, 1995) but ∼1 centimorgan (cM) away from the S locus homoeologous region. AtS1 is tightly linked to MS2 on chromosome 3, and ARK3 is located on chromosome 4 ∼4.3 cM from g3845 and 12.2 cM from m600 (and not on chromosome 3, as previously reported; Dwyer et al., 1994).
Figure 1. Comparative Map of the S8 Haplotype of B. campestris and the Homeologous Region in Arabidopsis.

The 275-kb Arabidopsis genomic region and its coverage by bacterial artificial chromosome (BAC) and yeast artificial chromosome (YAC) clones are shown at left, and the 510-kb SfiI fragment containing the S8 haplotype of B. campestris is shown at right. The arrows at the end of the YACs indicate that the inserts in these clones extend beyond the region shown on the map. The Arabidopsis map shows the location of the original 40-kb λ contig that hybridized with the B. campestris S locus–derived s probes (stippled rectangle) and the position of BamHI and EagI sites, as determined by analysis of the BAC and YAC clones. Molecular markers found in common between the Arabidopsis and Brassica regions are connected by dashed lines. Arabidopsis-derived markers that were found to be unlinked to the Brassica S8 haplotype by pulsed-field gel electrophoresis (PFGE) or restriction fragment length polymorphism (RFLP) analysis are enclosed in boxes, and those that could not be used for comparative mapping are indicated by filled circles. Markers indicated by asterisks are shown in their most probable order. On the B. campestris map, the regions encompassed by overlapping λ clones are indicated by the hatched rectangles, and the uncloned DNA between SLG8 and SRK8 is indicated by the filled box. The broken rectangles represent segments where the scale of the map changes. The positions of the Brassica-derived markers sp14 and sp400 that were used in PFGE analysis are shown by filled diamonds. The orientation of the SLG8 and SRK8 transcriptional units was determined in a previous study (Boyes et al., 1997) and is indicated by arrows. The rectangles at bottom right show the scale of the maps.
**Isolation of Molecular Markers from Arabidopsis**

To extend the cloned portion of this Arabidopsis chromosomal segment, the EW (Ward and Jen, 1990) and EG (Grill and Somerville, 1991) yeast artificial chromosome (YAC) libraries were screened with the single-copy end probe from the λ contig, resulting in the isolation of the 300-kb EG21B10 and the 180-kb EW13G7 YAC clones. Neither of these clones had been placed on the YAC contig of Arabidopsis chromosome 1. In addition, the Texas A & M University (TAMU; College Station, TX) Arabidopsis bacterial artificial chromosome (BAC) library (Choi et al., 1995) was screened with the single-copy Brassica s298, s6B-1.2, and s40.2-1.7 probes and with a probe (y14) derived from YAC EW13G7. Nine BAC clones (TAMU 1A8, 3H22, 4O24, 6F5, 7L7, 7K6, 8F5, 2J 21, and 2P16) were identified in this screen. Three of these clones, 1A8, 6F5, and 2P16, with inserts of 99, 73, and 96 kb, respectively, contained most of the chromosomal region of interest, except for a 6-kb gap between 65F and 2P16 as well as a 3.5-kb gap between 65F and 1A8 (see Methods). The arrangement of BAC clones 1A8, 6F5, and 2P16 relative to the YAC clones is shown in Figure 1.

Twenty-six independent subclones were isolated from libraries of EcoRl-HindIII fragments constructed from the EG21B10 and EW13G7 YACs, and 21 clones containing unique (non-cross-hybridizing) cDNAs were isolated by screening an Arabidopsis floral cDNA library with BAC clones 1A8, 6F5, and 2P16 (see Methods). The subset of sequences in clones 1A8, 6F5, and 2P16 relative to the YAC clones is shown in Figure 1.

Nine BAC clones (TAMU 1A8, 3H22, 4O24, 6F5, 7L7, 7K6, 8F5, 2J 21, and 2P16) were identified in this screen. Three of these clones, 1A8, 6F5, and 2P16, with inserts of 99, 73, and 96 kb, respectively, contained most of the chromosomal region of interest, except for a 6-kb gap between 65F and 2P16 as well as a 3.5-kb gap between 65F and 1A8 (see Methods). The arrangement of BAC clones 1A8, 6F5, and 2P16 relative to the YAC clones is shown in Figure 1.

Twenty-six independent subclones were isolated from libraries of EcoRI-HindIII fragments constructed from the EG21B10 and EW13G7 YACs, and 21 clones containing unique (non-cross-hybridizing) cDNAs were isolated by screening an Arabidopsis floral cDNA library with BAC clones 1A8, 6F5, and 2P16 (see Methods). The subset of sequences in clones 1A8, 6F5, and 2P16 relative to the YAC clones is shown in Figure 1.

The remainder of the Arabidopsis-derived markers did not hybridize with the S8 λ contigs, and they were positioned by using RFLP analysis of F2 plants that segregate for the S8 and S12 haplotypes and PFGE (see Methods). RFLP analysis revealed that the single-copy markers y81 and y35 (shown in boxes in Figure 1) segregated independently of the S haplotype, whereas the single-copy markers y14 and c31 cosegregated with the S haplotype (no recombinants detected in 67 F2 plants analyzed). Further positioning of these linked markers relative to SLG8 was achieved by PFGE, as shown in Figure 2 and Table 2, and summarized below.

On PFGE blots of DNA from plants homozygous for the S8 haplotype, a probe (HV8) specific for the SLG8/SRK8 gene pair hybridized with a 510-kb SflI restriction fragment and to a 610-kb NotI restriction fragment (Figure 2A), as did the flanking S8-derived markers sp14 and sp400 (data not shown). All three markers also hybridized consistently with a 780-kb SflI fragment that is, in all likelihood, a partially digested fragment (incomplete digestion is often observed upon treatment of Brassica DNA with SflI; Sadowski et al., 1996). Marker y14 hybridized with the same 510- and 780-kb SflI fragments as did the HV8 probe and with the same 97-kb BssHII and 115-kb Smal fragments as did sp400. However, in NotI digests, y14 hybridized with a 470-kb fragment and not with the 610-kb fragment containing the other S8 locus probes (Table 2). NotI-Smal digests confirmed that the NotI restriction site is located within the 115-kb Smal

---

**Comparative Mapping of Molecular Markers in Arabidopsis and Brassica**

The Brassica S locus–derived (s) markers and the Arabidopsis YAC-derived genomic (y) markers and BAC-derived cDNA (c) markers were used to generate a molecular map of the Arabidopsis chromosomal segment. The relative order of the markers (with the exception of cR, c2D, c2H, c2P, cJ, and cU) could be determined by hybridization with arrays of overlapping λGEM11 recombinant clones generated from each of the 1A8, 6F5, and 2P16 BAC clones by partial digestion with Sau3A1 (see Methods). Figure 1 shows a partial restriction map of the Arabidopsis chromosomal segment spanned by the BAC and YAC clones and the order of the markers within this segment.

For mapping in Brassica, we selected nine YAC-derived markers and seven cDNA markers because they were represented in relatively few copies in the Brassica genome, and they detected an RFLP between the parents of our B. campestris mapping population or produced adequate hybridization signals on pulsed-field gel electrophoresis (PFGE) blots of Brassica DNA.

Placement of markers c17, c3, y17, and c2K/y26, which are single-copy sequences in B. campestris, on the long-range map of the S8 haplotype was straightforward because these probes hybridized with the S8 λ contigs (Figure 1). Markers c2K and y26 hybridized with identical λ restriction fragments, and sequence analysis verified that they represent the genomic fragment and corresponding cDNA of the same gene. The position of marker c3 coincided with that of the S8-derived s6B-1.2; therefore, c3 is likely to be the Arabidopsis homolog of the Brassica s6B-1.2 sequence. Markers c20, c30, and y19, which correspond to multiple-copy sequences in Brassica, also hybridized with the S8 λ contigs (Figure 1), indicating that at least one copy of each sequence is located at the S locus region.

On PFGE blots of DNA from plants homozygous for the S8 haplotype, a probe (HV8) specific for the SLG8/SRK8 gene pair hybridized with a 510-kb SflI restriction fragment and to a 610-kb NotI restriction fragment (Figure 2A), as did the flanking S8-derived markers sp14 and sp400 (data not shown). All three markers also hybridized consistently with a 780-kb SflI fragment that is, in all likelihood, a partially digested fragment (incomplete digestion is often observed upon treatment of Brassica DNA with SflI; Sadowski et al., 1996). Marker y14 hybridized with the same 510- and 780-kb SflI fragments as did the HV8 probe and with the same 97-kb BssHII and 115-kb Smal fragments as did sp400. However, in NotI digests, y14 hybridized with a 470-kb fragment and not with the 610-kb fragment containing the other S8 locus probes (Table 2). NotI-Smal digests confirmed that the NotI restriction site is located within the 115-kb Smal
fragment, thus positioning the y14 marker between 39 and 139 kb 3’ of SRK (Figure 1).

In the case of the other S locus–linked marker, c31, PFGE analysis (Figure 2B) demonstrated that it hybridized with 145-kb BssHII and 97-kb EagI fragments, which also contain the S locus–derived sp14 marker (Figure 2B) and s6B-1.2 marker (Table 2). The c31 probe hybridized with a 48-kb Smal fragment (Figure 2B), placing c31 distal to the 120-kb Smal fragment containing sp14 and the s6B-1.2 markers. In addition, c31 hybridized with the 510- and 780-kb SfiI

Table 1. Properties of Molecular Markers Used in This Study

<table>
<thead>
<tr>
<th>Markera</th>
<th>Arabidopsis</th>
<th>Brassica</th>
<th>Intensity of Hybridization with Brassica DNAc</th>
<th>Sequence Similarity to Known Genesd</th>
</tr>
</thead>
<tbody>
<tr>
<td>sSLG</td>
<td>0</td>
<td>1</td>
<td>+++++</td>
<td>Nasrallah et al. (1985)</td>
</tr>
<tr>
<td>sSRK</td>
<td>0</td>
<td>1</td>
<td>+++</td>
<td>Stein et al. (1991)</td>
</tr>
<tr>
<td>s299</td>
<td>1</td>
<td>1</td>
<td>++</td>
<td>SLL2 (Yu et al., 1996)</td>
</tr>
<tr>
<td>s40.2-1.7</td>
<td>1</td>
<td>1</td>
<td>++++</td>
<td>NDe</td>
</tr>
<tr>
<td>sBH2.0</td>
<td>1</td>
<td>1</td>
<td>+++</td>
<td>ND</td>
</tr>
<tr>
<td>s298/CA</td>
<td>1</td>
<td>1</td>
<td>+++</td>
<td>Clp protease (Letham and Nasrallah, 1998)</td>
</tr>
<tr>
<td>sd39/cl</td>
<td>4</td>
<td>4</td>
<td>+++</td>
<td>Nonef</td>
</tr>
<tr>
<td>s6.8-1.2/c3</td>
<td>1</td>
<td>1</td>
<td>++</td>
<td>None</td>
</tr>
<tr>
<td>y81</td>
<td>1</td>
<td>1</td>
<td>ND</td>
<td>Arginyl-tRNA-transferase</td>
</tr>
<tr>
<td>y12</td>
<td>1</td>
<td>4</td>
<td>++</td>
<td>None</td>
</tr>
<tr>
<td>y14</td>
<td>1</td>
<td>1</td>
<td>++++</td>
<td>Kinase</td>
</tr>
<tr>
<td>y17</td>
<td>1</td>
<td>1</td>
<td>+++</td>
<td>None</td>
</tr>
<tr>
<td>y19</td>
<td>1</td>
<td>4</td>
<td>++</td>
<td>DNA ligase</td>
</tr>
<tr>
<td>y26/c2K</td>
<td>1</td>
<td>1</td>
<td>++</td>
<td>ND</td>
</tr>
<tr>
<td>y35</td>
<td>1</td>
<td>1</td>
<td>+++</td>
<td>ND</td>
</tr>
<tr>
<td>y39</td>
<td>2</td>
<td>3</td>
<td>++</td>
<td>ND</td>
</tr>
<tr>
<td>y49</td>
<td>1</td>
<td>3</td>
<td>++</td>
<td>ND</td>
</tr>
<tr>
<td>cD</td>
<td>4</td>
<td>7</td>
<td>+</td>
<td>Sucrose proton symporter</td>
</tr>
<tr>
<td>c2D</td>
<td>3</td>
<td>—</td>
<td>—</td>
<td>None</td>
</tr>
<tr>
<td>cF</td>
<td>3</td>
<td>9</td>
<td>++++</td>
<td>Wilm’s tumor suppressor</td>
</tr>
<tr>
<td>c3H</td>
<td>1</td>
<td>1</td>
<td>++</td>
<td>None</td>
</tr>
<tr>
<td>cJ</td>
<td>3</td>
<td>13</td>
<td>+</td>
<td>None</td>
</tr>
<tr>
<td>c4J</td>
<td>9</td>
<td>9</td>
<td>++++</td>
<td>None</td>
</tr>
<tr>
<td>c2P</td>
<td>1</td>
<td>4</td>
<td>+</td>
<td>None</td>
</tr>
<tr>
<td>cN</td>
<td>2</td>
<td>—</td>
<td>—</td>
<td>None</td>
</tr>
<tr>
<td>cR</td>
<td>2</td>
<td>5</td>
<td>+</td>
<td>Kinase</td>
</tr>
<tr>
<td>cU</td>
<td>3</td>
<td>19</td>
<td>++</td>
<td>None</td>
</tr>
<tr>
<td>cX</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>Drosophila seven-in-absentia isolog</td>
</tr>
<tr>
<td>c17</td>
<td>1</td>
<td>1</td>
<td>+++</td>
<td>Fructokinase</td>
</tr>
<tr>
<td>c20</td>
<td>5</td>
<td>10</td>
<td>++</td>
<td>Calmodulin</td>
</tr>
<tr>
<td>c30</td>
<td>2</td>
<td>4</td>
<td>++</td>
<td>RNA binding protein</td>
</tr>
<tr>
<td>c31</td>
<td>1</td>
<td>1</td>
<td>++</td>
<td>ETR1 (Chang et al., 1993)</td>
</tr>
<tr>
<td>c52</td>
<td>1</td>
<td>1</td>
<td>+</td>
<td>None</td>
</tr>
<tr>
<td>c59</td>
<td>3</td>
<td>5</td>
<td>++</td>
<td>DNA binding protein</td>
</tr>
</tbody>
</table>

aMarker designations: s denotes markers derived from the S8 haplotype, y denotes markers generated from Arabidopsis YAC clones, c denotes Arabidopsis cDNA markers isolated by screening a floral cDNA library with BAC clones.

bThe number of hybridizing fragments for Arabidopsis is based on the number of bands detected in either an EcoRI or HindIII digest, whichever was lower. The number of hybridizing fragments for Brassica is based on the number of bands detected in an HindIII digest of B. campestris S8S8 homozygous plants. Dashes indicate no or barely detectable hybridization signals.

cBased on visual inspection of Arabidopsis and Brassica DNA gel blots, the Arabidopsis-derived probes produced hybridization signal intensities in Brassica that varied from equivalent to those produced in Arabidopsis (++++) to barely detectable (−).

dBased on database searches with partial DNA sequences.

eND, sequence not determined.

fNone, sequence did not exhibit similarity to known genes at >1 × 10−10 probability.
fragments and with the 610-kb NotI fragment that contain the $S_8$ haplotype–derived markers (Table 2). These results place $c31$ 315 kb 3' of $SLG_8$ and 60 kb distal to $sp14$ (Figure 1).

The linkage relationship to the $S_8$ haplotype of Arabidopsis-derived markers that were multicopy in Brassica was less straightforward to establish but nevertheless was accomplished for four of the markers. As shown by the PFGE blots in Figure 3A, the $c59$ probe hybridized with more than one fragment in digests with three different rare-cutting restriction enzymes, indicating that the various $c59$-related sequences are not tightly clustered in the Brassica genome.

However, it is evident that at least one $c59$-related sequence is physically linked to the $S_8$ haplotype. In each of the digests shown in Figure 3A, one of the restriction fragments identified by $c59$, a 145-kb BssHII fragment, a 97-kb EagI fragment, and a 120-kb Smal fragment, corresponded in size to the restriction fragment that hybridized with the $S$ locus-derived markers $sp14$ (Figure 3A) and $s6B-1.2$ (Table 2). In addition, the $c59$ probe hybridized with the 610-kb NotI fragment that contains $SLG_8$ and $SRK_8$ (Table 2).

These data place at least one of the $c59$-related copies 10 to 60 kb from $sp14$ and 70 to 120 kb from $SLG_8$ (Figure 1). Further refinement of the map position of this $c59$-related sequence was made possible by the analysis of genomic clones isolated from an $S_8$ xGEM11 library on the basis of hybridization with $c59$. A subset of these clones contained both a $c59$-related sequence and the $c31$ marker, placing these two markers within 10 kb of each other.

In contrast, none of the $y49$, $y39$, and $y12$-related sequences found in the Brassica genome mapped to the immediate vicinity of the $S_8$ haplotype. Of the three HindIII restriction fragments detected by $y49$ (Table 1), one fragment was polymorphic between plants homozygous for the $S_8$ and $S_{f2}$ haplotypes, and it segregated independently of the $S$ haplotype. In addition, the $y49$ probe did not hybridize with either the 610-kb NotI fragment or the 510- and 780-kb SfiI fragments identified by the HV8 probe (Figure 3B). Probes $y39$ and $y12$, which are located 3 kb on either side of $y49$ in Arabidopsis and hybridize with three and four HindIII fragments, respectively, in B. campestris, detected the same NotI and SfiI fragments as did $y49$ (data not

Table 2. Restriction Fragments Identified by PFGE Blot Hybridization Analysis of the $S_8$ Haplotype

<table>
<thead>
<tr>
<th>Probe</th>
<th>restriction Enzyme</th>
<th>BssHII</th>
<th>EagI</th>
<th>Smal</th>
<th>NotI</th>
<th>SfiI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HV8</td>
<td>145</td>
<td>ND</td>
<td>120</td>
<td>610</td>
<td>510</td>
<td></td>
</tr>
<tr>
<td>$s6B-1.2$</td>
<td>145</td>
<td>97</td>
<td>120</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>$sp14$</td>
<td>145</td>
<td>97</td>
<td>120</td>
<td>610</td>
<td>510</td>
<td></td>
</tr>
<tr>
<td>$c59$</td>
<td>145 (350/500)</td>
<td>97 (130/140)</td>
<td>120 (325)</td>
<td>610</td>
<td>610</td>
<td>ND</td>
</tr>
<tr>
<td>$sp400$</td>
<td>97</td>
<td>ND</td>
<td>115</td>
<td>470</td>
<td>510</td>
<td></td>
</tr>
<tr>
<td>$y14$</td>
<td>97</td>
<td>80</td>
<td>115</td>
<td>470</td>
<td>510</td>
<td></td>
</tr>
<tr>
<td>$y49$</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>700/950</td>
<td>370/590</td>
<td></td>
</tr>
</tbody>
</table>

Numbers indicate the lengths of restriction fragments in kilobases. ND, not determined in this study.

Our estimate of the length of the EagI fragment containing the $s6B-1.2$ marker differs significantly from the 150-kb value reported in Boyes et al. (1997), probably because of differences in PFGE conditions.

For the multicopy $c59$ marker, the length of the restriction fragment containing the $S$ locus-linked copy is followed within parentheses by the lengths of the fragments containing other $c59$ copies that are likely to map elsewhere in the Brassica genome.
shown). Thus, the markers y12, y49, and y39 are physically linked to one another in B. campestris as they are in Arabidopsis, but they apparently have been duplicated and displaced as a group to locations outside of the S haplotype.

Alignment of the high-resolution maps generated in this study thus reveals that extensive but not absolute (see below) microsynteny has been maintained between the Arabidopsis and Brassica homeologous regions since their divergence (Figure 1). It also suggests that evolution of this genomic interval has involved a differential expansion or contraction of subsegments within the region. Physical distances separating the markers are similar between the two genomes at some intervals (e.g., s40.2-1.7 and c2k are ~2 kb apart in Arabidopsis and ~4 kb apart in Brassica, and the c17 and c20 markers are contained on the same λ clone in both Arabidopsis and Brassica). However, they were dissimilar at other intervals, with the distances in Brassica being longer in some cases (e.g., sBH2.0 and s40.2-1.7 are <30 kb apart in Arabidopsis but are >50 kb apart in Brassica) and somewhat unexpectedly, given that the size of the B. campestris genome is approximately five times that of the Arabidopsis genome, shorter in other cases (e.g., sBH2.0 and s6B-1.2 are separated by ~10 kb in Brassica but by as much as 35 kb in Arabidopsis).

**Complement of Sequences Contained within the Two Homeologous Regions and Linkage to ETR1**

Partial nucleotide sequence was derived for the majority of the molecular markers generated in this study. Many of the Arabidopsis-derived sequences were novel and did not uncover any matches in the DNA and protein databases. The fact that no Arabidopsis expressed sequence tags (ESTs) are known for approximately half of the cDNAs that we isolated might reflect either the low abundance of the corresponding transcripts or the underrepresentation of tissues expressing these transcripts in the cDNA libraries used to generate the ESTs. In any case, a subset of our markers did exhibit sequence similarity to genes previously cloned from plants or other organisms: these included genes encoding kinases, the Clp protease, DNA ligase, RNA and DNA binding proteins, calmodulin, and fructokinase. Particularly noteworthy is the sequence identity of marker c31 to ETR1, a gene that codes for the ethylene receptor (Chang et al., 1993). The conclusion that c31 represents a cDNA of ETR1 was verified by the mapping of YAC EG2G11, which contains ETR1 (Chang et al., 1993), to the Arabidopsis S locus homeologous region (Figure 1).

Significantly, none of the markers generated from the Arabidopsis YAC/BAC contig exhibited sequence similarity to SLG and SRK. Furthermore, no hybridization of the YAC and BAC clones spanning the region was detected with these Brassica probes, even under low-stringency hybridization conditions. These observations are consistent with previous studies suggesting the absence of an SLG/SRK-like gene pair in Arabidopsis (Dwyer et al., 1994).

**DISCUSSION**

By exploiting the resources generated by the Arabidopsis genome project in a targeted comparative mapping study,
we constructed high-resolution physical maps of a 510-kb chromosomal region that encompasses the S locus in B. campestris and of a 275-kb homeologous region in Arabidopsis. The value of comparative mapping for traits that are unique to one of the two species being compared has been questioned (Van Deynze et al., 1995). However, in our study, the comparative mapping approach was clearly useful, despite the fact that Arabidopsis lacks a functional SI system.

The close physical and genetic linkage of the ETR1 marker to the Brassica S locus identifies the Brassica homolog of this well-characterized Arabidopsis gene and its map position, and it provides a clear landmark for the genomic location of the S locus. Furthermore, due to the synteny exhibited by the two homeologous regions on a submegabase scale, the comparative study generated several molecular markers for the Brassica S locus region that should facilitate future analysis of this complex locus. Except for two small clusters of markers, all of the Arabidopsis-derived markers that were usable in our comparative study were found to map to the S locus region in B. campestris. Only the interval encompassing markers y81 and y35 and the interval encompassing markers y12, y49, and y39 were displaced to regions outside of the 510-kb genomic segment that includes the S4 haplotype of B. campestris. In addition, except for one relatively small inversion that involves markers s298, s299, and one of the sd39 sequences, the order of the markers was largely conserved between the two species.

It remains to be determined how many of the markers we identified represent expressed sequences (and potentially functional genes) in Arabidopsis and in Brassica. Among the 21 distinct classes of Arabidopsis cDNAs that were isolated by hybridization with BAC DNA, 10 hybridized with unique restriction fragments in Arabidopsis DNA and are therefore clearly derived from mRNAs encoded by sequences in this region. The number of potentially functional genes may be higher if at least some of the multiple-copy cDNAs are derived from the region or a fraction of the random genomic clones for which no cDNAs have been isolated do contain expressed sequences. For example, the single-copy genomic markers y12 and y14 (Table 1), respectively, might represent highly diverged functional members of the arginyl-tRNA transferase and kinase gene families.

In Brassica, only cDNAs corresponding to the s298 and s299/SLL2 markers have been isolated to date. In both cases, the high degree of nucleotide identity shared by the Arabidopsis- and Brassica-derived sequences suggests that they perform similar functions in the two species and might not fulfill a function specific to the SI response. However, the possibility cannot be excluded that some of the sequences held in common between the two homeologous regions might fulfill an SI-related function in Brassica while representing nonfunctional genes in Arabidopsis. Nevertheless, based on sequence similarity alone, it is reasonable to assume that a subset of potentially expressed sequences, such as c2K and c59, which encode a DNA ligase and a DNA binding protein, respectively, are unlikely to function in determining SI recognition functions.

The most striking difference between the two homeologous regions is that the Arabidopsis region does not contain any sequences related to SLG and SRK, as determined by DNA gel blot analysis of the YAC and BAC contigs. This result indicates that in Arabidopsis, the SI recognition genes were either deleted from the genome or displaced to another chromosomal location. As stated earlier, four genes have been isolated from Arabidopsis by hybridization with SLG and SRK probes (Dwyer et al., 1992, 1994; Tobias et al., 1992), none of which maps to the S locus homeologous region described in this study. Among these four genes, the SRK-like genes ARK1, ARK2, and ARK3 are not expressed in stigmas or anthers and are likely to be orthologous to some of several vegetatively expressed SRK-like genes from Brassica (Kumar and Trick, 1994; Pastuglia et al., 1997). Only the Arabidopsis AtS1 gene, which codes for a secreted glycoprotein, is expressed specifically in stigmatic papillar cells and thus potentially functions in pollen-stigma interactions (Dwyer et al., 1992, 1994). However, based on immunological and sequence relatedness, AtS1 appears to be the ortholog of SLR1, a papillary cell-specific Brassica gene that is unlinked to the S locus (Lalonde et al., 1989). Thus, although it is formally possible that in Arabidopsis, the SI specificity genes were displaced to another chromosomal location and subsequently diverged to assume different expression patterns and functions, it is evident that none of the members of the Arabidopsis S gene family isolated to date exhibits features of SI recognition genes.

The data are consistent with the notion that the SI specificity genes SLG and SRK were deleted from the ETR1-linked chromosomal position in the Arabidopsis lineage or that they were inserted at the homeologous position in the Brassica lineage. It is interesting that it is precisely at the position where the deletion/insertion of the SLG/SRK gene pair is predicted to have occurred, namely, the segment containing the s298, s299, and sd39 markers, that the two homeologous regions differ by an inversion and a duplication of the sd39 marker in Arabidopsis. Although it is not possible to infer with confidence the ancestral and derived configurations for this chromosomal segment, we favor the deletion hypothesis for various reasons. First, based on current estimates, the origin of Brassica S haplotypes at 20 to 40 million years ago (Uyenoyama, 1995) predates the divergence of Arabidopsis and Brassica (Kowalski et al., 1994). Second, the genus Arabis, which is very closely related to Arabidopsis (Price et al., 1994), contains self-incompatible species (Van Treuren et al., 1997). Third, in Brassica species, spontaneous deletion of the S locus genes has occurred, leading to the genetic breakdown of SI (Nasrallah et al., 1994a). Fourth, studies relating species dispersal with mating system (Baker, 1955; Stebbins, 1957) have demonstrated the prevalence of self-incompatible species at the center of distribution zones and the predominance of self-fertilizing species.
in peripheral zones, supporting the derived nature of self-fertility.

Both SI and self-fertility can be of adaptive value to organisms. In this context, a best-of-two-worlds scenario has been proposed (Charlesworth et al., 1990; Levin, 1996) in which the evolution of self-fertility would occur via the spread of mutations at modifier loci that weaken or abolish the activity of SI alleles. Indeed, this class of mutations has been observed in some self-compatible strains of Brassica (Nasrallah, 1974; de Nettancourt, 1977; Hinata et al., 1983; Nasrallah et al., 1992). Such a mechanism would allow the benefits of selfing, including reproductive assurance in pollen-limited situations and perpetuation of adaptive genes, while maintaining the capacity to revert to an outcrossing mating system as warranted. In contrast, a mechanism based on deletion of the SI specificity genes, as we propose occurred in the Arabidopsis lineage, would represent an irreversible evolutionary change from an outbreeding mode to an autogamous mating system. A test of this deletion hypothesis and of the generality of such a mechanism for the evolution of self-fertilization in crucifers awaits the extension of the comparative analysis presented in this study to other lineages of the Brassicaceae.

METHODS

Genetic Mapping in Arabidopsis

Map position in Arabidopsis thaliana was determined by segregation data from the Lister and Dean (1993) mapping population. DNA from the first 100 lines was isolated using a DNA minipreparation procedure and digested with a restriction enzyme that reveals a restriction fragment length polymorphism (RFLP) for the probe used. Two-point segregation data from the Lister and Dean recombinant inbred lines were analyzed relative to 66 markers whose segregation and map positions were previously established using MAPMAKER (Lander et al., 1987).

Isolation of Arabidopsis Yeast Artificial and Bacterial Artificial Chromosome Clones

For the isolation of yeast artificial chromosome (YAC) clones, filters containing arrays of the EW (Ward and Jen, 1990) and EG (Grill and Somerville, 1991) YAC libraries were obtained from the Arabidopsis Biological Resource Center (ABRC) at Ohio State University (Columbus, OH) and probed with a single-copy Arabidopsis sequence from a λ contig isolated by hybridization to the Brassica campestris s298 and s299 probes. Bacterial artificial chromosome (BAC) clones were isolated from the Texas A & M University (TAMU; College Station, TX) BAC library (filters provided by the ABRC) by hybridization with the single-copy sequences s298, s68-1.2, s40.2-1.7, and YAC-derived marker y14 (see below). Nine BAC clones identified in this screen were ordered relative to the YAC clones based on hybridization of BAC end probes to BamHI-digested YAC DNA and to corresponding λ contigs.

Isolation and Sequencing of Molecular Markers

Three types of molecular markers were used in this study. Brassica S locus-derived sequences are prefaced with an s; Arabidopsis sequences derived from YAC clones or BAC clones are prefaced with a y and c, respectively.

s Molecular Markers

The s molecular markers are derived from the λ contigs of the S₈ haplotype. The placement of SLG₈, SRK₈, s298, s299 (designated as SLL2 in Yu et al. [1996]), s40.2-1.7, s68-1.2, and sBH2.0 has been described previously (Boyes et al., 1997). Markers sp14 and sp400 are single-copy sequences recently isolated by λ walking in the S₈ haplotype. HV8 is a 510-bp probe specific for the SLG₈/SRK₈ gene pair generated by polymerase chain reaction (PCR) using primers that flank a hypervariable region of SLG₈ (starting at nucleotide 710 and ending at nucleotide 1220 of the SLG₈ sequence; Dwyer et al., 1991).

y Molecular Markers

The y molecular markers were derived from Arabidopsis YAC DNA as follows. YAC DNA was isolated using a standard yeast DNA isolation protocol (Pasero and Marilley, 1993) and subjected to pulsed-field gel electrophoresis (PFGE) by using clamped homogeneous electric field (CHEF) technology in a CHEF-DR II electrophoresis apparatus (Bio-Rad) on a 1% (w/v) low-melting agarose gel in 0.5 × Tris-borate-EDTA at 175 V with a pulse-switching interval ramped from 5 to 50 sec or 20 to 40 sec for 24 hr. DNA was extracted from gel slices containing the YAC by digestion with Gelase (New England Biolabs, Beverly, MA), concentrated using Centricon-30 (Amicon, Beverly, MA), digested to completion with EcoRI and HindIII, and shotgun cloned into the pBluescript KS+ (Stratagene, La Jolla, CA) plasmid. Approximately 480 colonies were chosen for further analysis.

Clones containing inserts of previously subcloned sequences of the yeast vector or of repetitive yeast and plant DNA were identified by colony hybridization and were not analyzed further. The inserts from the remaining clones were labeled with phosphorus-32 by the random priming method of Feinberg and Vogelstein (1983) using the random primed DNA labeling kit from Boehringer-Mannheim. Duplicate inserts were eliminated by hybridization of the library colonies with the newly isolated inserts. The inserts from individual subclones were hybridized with gel blots containing DNA of the EG21B10 and EW13G7 YACs as well as from an unrelated negative control YAC clone. In all, 26 distinct classes of subclones were used in subsequent studies.

c Molecular Markers

The c molecular markers are cDNA markers generated by screening ~200,000 cDNA clones of an Arabidopsis × ZAP cDNA flower bud library (Dwyer et al., 1994) with BAC clones. Duplicate cDNAs from the primary phage picks were identified by a screen that combined PCR and hybridization with already isolated clones as follows. DNA was isolated using a method developed for PCR sequencing from phage DNA (Krishnan et al., 1991): 5 µL of the DNA was added to a 50-µL
PCR reaction mix containing 30 μM each T3 and T7 primers, 200 μM each deoxynucleotide triphosphate, 1 × PCR reaction buffer (Bethesda Research Laboratories), and 1.25 units of Taq polymerase (Bethesda Research Laboratories). After PCR (35 cycles at 94°C for 1 min, 46°C for 2 min, and 72°C for 2 min and 30 sec), half of the PCR reaction product was separated on a 1% (w/v) agarose gel and subjected to DNA gel blot analysis (Southern, 1975). The blots were hybridized with a mix of already isolated cDNA probes, allowing the identification of clones containing sequences related to previously isolated clones. By using this method, the number of unique (non-cross-hybridizing) clones was reduced from 180 primary BAC-hybridizing clones to 21 clones.

Sequence Analysis of Clones

cDNA and YAC inserts of interest were end sequenced using the Sequence Analysis of Clones isolated clones. By using this method, the number of unique (non-identification of clones containing sequences related to previously bridged with a mix of already isolated cDNA probes, allowing the physical mapping of molecular markers in Arabidopsis.

Ordering of the YAC- and BAC-Derived Markers

BAC DNA was isolated essentially as previously described (Ioannou et al., 1994). DNA from each BAC was partially digested with Sau3AI, and DNA ranging from 15 to 50 kb was size selected by centrifugation on a 5 to 29% NaCl gradient. Partially digested insert DNA from each BAC was ligated to λGem11 BamHI arms (Promega, Madison, WI) and packaged using in vitro packaging extracts (Stratagene). Each library contained >50,000 clones.

Bacteriophage from each library were plated at low density for the recovery of individual plaques. Ninety-six clones from each library were picked at random into a 96-well microtiter plate and spotted with a multichannel pipettor onto a lawn of KW251 cells to create a master grid of plaques. Plaque lifts were prepared with Hybond filters (Amersham) that were processed, hybridized, and washed according to the manufacturer’s instructions. The s probes that hybridized with Arabidopsis DNA and the Arabidopsis-derived y and c probes were hybridized with the filters to determine marker order.

Restriction Enzyme Analysis

To construct a partial restriction map of the Arabidopsis chromosomal DNA spanned by the BAC and YAC clones, both BAC and YAC DNA were digested with BamHI and Eagl-NotI enzymes that cleave the DNA infrequently. BAC fragments were separated on a 0.5% (w/v) agarose gel. YAC fragments were fractionated by PFGE on a 1% (w/v) agarose gel in 0.5 × Tris-borate-EDTA at 175 V with a pulse-switching interval ramped from 5 to 15 sec for 17 hr. The gels were subjected to DNA gel blot analysis and hybridized with s probes that hybridize with Arabidopsis DNA and the y and c probes. By using the hybridization pattern of previously ordered molecular markers, a restriction map could be produced. The molecular sizes of the individual fragments were estimated by comparison to standard markers run alongside of the samples. Not every BamHI restriction site could be identified for the BAC 1A8 clone.

Linkage of Arabidopsis-Derived Molecular Markers to the Brassica S8 Locus

Three methods were used to place the Arabidopsis-derived molecular markers that produced a hybridization signal in Brassica DNA on the map of the S8 haplotype. The probes were first hybridized with the 123.5-kb S8 χ contigs. For probes that were not contained on the χ contigs, RFLP analysis or PFGE was used to establish genetic or physical linkage to the S8 haplotype.

To identify RFLPs, the probes were hybridized with DNA gel blots containing DNA isolated from S8S8 and S8Sf2 plants and digested with BamHI, EcoRI, HindIII, and SstI. The segregation of the RFLP was then analyzed in a population of F2 plants generated by selfing an S8Sf2 F1 plant derived from a cross between a plant homozygous for the S8 haplotype and another homozygous for the Sf2 haplotype. The haplotype composition of the F2 plants was determined using a HindIII RFLP that distinguishes the SLG genes of the Sf2 and S8 haplotypes, as identified by hybridization with an SLG cDNA probe. Several of the markers used in this study also detected a HindIII RFLP between plants homozygous for the Sf2 and Sf2 haplotypes. To determine linkage of a particular Arabidopsis-derived molecular marker to the S locus, its segregation in the F2 population was compared with that of the S haplotypes.

For PFGE analysis, high molecular weight DNA was isolated from the nuclei of S8 homozygous plants (Zhang et al., 1995). Digestion of agarose-embedded DNA with rare-cutting enzymes was conducted as described by Boyes and Nasrallah (1993). Restriction fragments were resolved by PFGE in a CHEF-DR II apparatus by using the following conditions. For separation of DNA in the range of 25 to 500 kb, a 1% (w/v) gel in 0.5 × Tris-borate-EDTA was run at 175 V with a pulse-switching interval ramped from 5 to 50 sec over 30 hr. For separation of DNA in the range of 350 to 1200 kb, a 1% (w/v) gel in 0.5 × Tris-borate-EDTA was run at 200 V with a pulse-switching interval ramped from 47 to 74 sec over 35 hr.

After electrophoresis, the DNA was transferred to a GeneScreen Plus (Du Pont, Boston, MA) membrane. DNA gel blots were hybridized at 65°C in 10% (w/v) dextran sulfate, 330 mM sodium phosphate, pH 7.0, 10 mM EDTA, and 5% (w/v) SDS. The blots were washed at 65°C in 2 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate) and 0.5% (w/v) SDS. When multiple probings were performed, the blots were stripped between probings and exposed to x-ray film to verify the absence of residual hybridization signal.

ACKNOWLEDGMENTS

We thank Amy Casselman for data relating to p299/SLL2 and for initial PFGE results and Kathleen Dwyer for providing the map location of the Arabidopsis ARK3 gene. This work was supported by a grant from the U.S. Department of Agriculture. J.A.C. was supported in part by a predoctoral fellowship from the Cornell National Science Foundation Plant Science Center, a unit in the U.S. Department of Agriculture-Department of Energy-National Science Foundation Plant Science Centers Program.
REFERENCES


Comparative Mapping of the *Brassica S* Locus Region and Its Homeolog in Arabidopsis: Implications for the Evolution of Mating Systems in the Brassicaceae
Joann A. Conner, Patrick Conner, Mikhail E. Nasrallah and June B. Nasrallah
*Plant Cell* 1998;10;801-812
DOI 10.1105/tpc.10.5.801

This information is current as of July 8, 2017

<table>
<thead>
<tr>
<th>References</th>
<th>This article cites 54 articles, 23 of which can be accessed free at: /content/10/5/801.full.html#ref-list-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>eTOCs</td>
<td>Sign up for eTOCs at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a></td>
</tr>
<tr>
<td>CiteTrack Alerts</td>
<td>Sign up for CiteTrack Alerts at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a></td>
</tr>
<tr>
<td>Subscription Information</td>
<td>Subscription Information for <em>The Plant Cell</em> and <em>Plant Physiology</em> is available at: <a href="http://www.aspb.org/publications/subscriptions.cfm">http://www.aspb.org/publications/subscriptions.cfm</a></td>
</tr>
</tbody>
</table>