Phenotypic and Genotypic Expression of Self-incompatibility Haplotypes in *Arabidopsis lyrata* Suggests Unique Origin of Alleles in Different Dominance Classes

Nadia L. Prigoda, Annette Nassuth, and Barbara K. Mable

Department of Botany, University of Guelph, Guelph, Ontario, Canada

The highly divergent alleles of the *SRK* gene in outcrossing *Arabidopsis lyrata* have provided important insights into the evolutionary history of self-incompatibility (SI) alleles and serve as an ideal model for studies of the evolutionary and molecular interactions between alleles in cell-cell recognition systems in general. One tantalizing question is how new specificities arise in systems that require coordination between male and female components. Allelic recruitment via gene conversion has been proposed as one possibility, based on the division of DNA sequences at the *SRK* locus into two distinctive groups: (1) sequences whose relationships are not well resolved and display the long branch lengths expected for a gene under balancing selection (Class A); and (2) sequences falling into a well-supported group with shorter branch lengths (Class B) that are closely related to an unlinked paralogous locus. The purpose of this study was to determine if differences in phenotype (site of expression assayed using allele-specific reverse transcription–polymerase chain reaction) or function (dominance relationships assayed through controlled pollinations) accompany the sequence-based classification. Expression of Class A alleles was restricted to floral tissues, as predicted for genes involved in the SI response. In contrast, Class B alleles, despite being tightly linked to the SI phenotype, were unexpectedly expressed in both leaves and floral tissues; the same pattern found for a related unlinked paralogous sequence. Whereas Class A included haplotypes in three different dominance classes, all Class B haplotypes were found to be recessive to all except one Class A haplotype. In addition, mapping of expression and dominance patterns onto an S-domain–based genealogy suggested that allelic dominance may be determined more by evolutionary history than by frequency-dependent selection for lowered dominance as some theories suggest. The possibility that interlocus gene conversion might have contributed to allelic diversity is discussed.

**Introduction**

Cell-cell recognition systems as diverse as those involved in immune function in animals (e.g., involving genes at the major histocompatibility complex [MHC]) and prevention of self-fertilization in plants (e.g., genes controlling self-incompatibility [SI]) share many features related to maintaining high numbers of unique specificities to ensure maximum recognition of invading pathogens or maximum number of potential mating partners, respectively. Both processes share the necessity of distinguishing self from nonself and both are under the control of multiple genes in tight physical linkage. Theoretical population geneticists have long been interested in how the impressively high diversity is generated in both types of systems and how forces such as balancing selection act to maintain this diversity. Both processes share the necessity of distinguishing self from nonself and both are under the control of multiple genes in tight physical linkage. Theoretical population geneticists have long been interested in how the impressively high diversity is generated in both types of systems and how forces such as balancing selection act to maintain this diversity.

In SI systems, recognition between the male and female components of the same plant must be maintained for the system to work; a fact that has intrigued and perplexed population geneticists (e.g., Charlesworth 2000; Uyenoyama 2000b). This necessity is thought to explain why pollen and stigma genes are physically linked in a gene-dense region characterized by limited recombination and has made it difficult to explain how new dual specificities can arise simultaneously. For sporophytic SI systems (SSI), where the SI phenotype of a pollen grain is determined by the diploid genome of its parent, the system is also complicated by dominance interactions among alleles in both female and male components.

Diversity in both female (S-related kinase [SRK]; Stein et al. 1991) and male (S-cysteine rich [SCR]; Schopfer, Nasrallah, and Nasrallah 1999) genes has been investigated extensively in cultivated *Brassica* (Brassicaceae) (e.g., Kusaba et al. 2000; Nishio and Kusaba 2000; Sato et al. 2002; Choukajorn et al. 2004). Both genes are members of large gene families: SRK is a member of an extremely large family of serine-threonine receptor kinases (Edelman, Blumenthal, and Krebs 1987; Dwyer et al. 1994; Walker 1994), characterized by an extracellular glycoprotein domain involved in recognition of ligand proteins (such as pollen coat proteins) and a kinase domain that, upon binding of the ligand (e.g., a self-similar pollen protein) to the receptor domain, is activated by phosphorylation and then initiates a signal cascade; SCR is a member of a gene family known as plant defensins, which are small, secreted, cysteine-rich proteins with a variety of functions (Doughty et al. 1998; Takayama et al. 2000). In some haplotypes of *Brassica*, a gene encoding an extracellular glycoprotein (S-locus glycoprotein [SLG]; Nasrallah et al. 1985) with high sequence similarity to the extracellular domain (the S-domain) of the *SRK* variant found in that haplotype is also known to be physically linked to *SRK* (Boyes and Nasrallah 1993), but its function is not yet completely clear. In addition, some haplotypes are known to contain multiple copies of *SRK* or *SCR*, both within and outside the *S*-locus region (Boyes et al. 1997; Kusaba et al. 2001; Shiba et al. 2003), and a number of other paralogous loci have been identified with relatively high sequence similarity to *SRK* (Dwyer et al. 1994). Due to
the high divergence among SRK alleles, distinguishing which sequences are actually allelic and involved in the SI reaction cannot be based simply on relative similarity and requires genetic confirmation through analysis of cosegregation of genotypes with SI phenotypes or transformation studies. Most studies of SI have assumed that alleles involved in SI will only be expressed in reproductive tissues (e.g., Lalonde et al. 1989; Hatakeyama et al. 2001), in contrast to other gene family members that are known to be expressed more ubiquitously (Dwyer et al. 1994; Kumar and Trick 1994; Doughty et al. 1998; Pastuglia et al. 2002; Smythe and Ayscough 2003); this criterion has sometimes been used to exclude some sequence types as putative alleles. However, this assumption has not been rigorously tested in natural systems.

Gene conversion has been suggested previously as a mechanism for generating diversity in SI systems (Fisher 1961), and emphasis has been on the role of intralocus conversion events or concerted evolution between the extracellular domain of SRK and SLG in the generation of new specificities (J. B. Nasrallah and M. E. Nasrallah 1989; Cabrillac et al. 1999; Sato et al. 2002). Investigation of SRK diversity in natural populations of Arabidopsis lyrata (a self-incompatible relative of Brassica and the self-compatible model plant Arabidopsis thaliana; van Treuren et al. 1997; Savolainen et al. 2000; Mitchell-Olids 2001) has also revealed that interlocus gene conversion could play a role in allelic diversification (Charlesworth et al. 2003a). Based on DNA sequence divergence in the kinase domain, alleles at this locus can be divided into two groups: (1) a diverse collection of sequences whose relationships are not well resolved by phylogenetic analysis and display the long branch lengths expected for a gene under balancing selection (Class A); and (2) a well-supported group with shorter branch lengths (Class B) that are closely related to two sequences that are not linked to the SI phenotype (Charlesworth et al. 2003a). Recent preliminary mapping evidence suggests that at least one of these sequences (Aly13-7) is located on a different chromosome than the S-locus (X. Vekemans, personal communication), suggesting that it represents a paralogous locus. Although such indirect evidence cannot distinguish whether the Class B alleles have arisen from this paralogous locus or whether the paralogous locus arose from the S-locus, the possibility of interlocus exchanges is intriguing.

The effective number of mating types in SSI systems is increased by dominance interactions in both pollen and stigma genes, but it is not known how changes in dominance occur. Classic genetic studies in cultivated Brassica demonstrated that dominance interactions can be highly complex, with differing dominance in pollen and stigma and nonlinear dominance hierarchies (Thompson and Taylor 1966; Ockendon 1975; Stevens and Kay 1989; Hatakeyama et al. 1998). Theory predicts that the relative frequency of S-alleles should vary by their dominance levels, with the expectation that recessive alleles should occur at higher frequencies and persist longer than dominant alleles (Schierup, Vekemans, and Christiansen 1997). However, the relative frequency of alleles in different classes is complicated by population substructure, pollination biology, and the particular model of dominance considered (Schierup 1998; Vekemans, Schierup, and Christiansen 1998), with some models predicting that all dominant alleles are “doomed” to become recessive whereas others make more complex predictions. This complexity was reflected in empirically derived estimates of relative allele frequencies in a population of A. lyrata from Iceland, where a single recessive allele was found at high frequency whereas frequency of other alleles did not appear to be related to their dominance level (Mable, Schierup, and Charlesworth 2003). Studies into the molecular mechanisms of dominance have, so far, focused predominantly on interactions between clearly recessive and clearly dominant haplotypes in Brassica (Hatakeyama et al. 2001; Shibata et al. 2002; Kakizaki et al. 2003) and A. lyrata (Kusaba et al. 2002). In Brassica, only two dominance types have been identified (Class I being pollen dominant to Class II) that are also clearly separable based on sequence divergence in the S-domain (Nasrallah, Nishio, and Nasrallah 1991).

SI alleles in A. lyrata show deeper divergence and more dominance classes than those in cultivated Brassica (Schierup et al. 2001; Mable, Schierup, and Charlesworth 2003; Bechsgaard, Bataillon, and Schierup 2004). However, it is not known how dominance relates to the sequence-based Class A and B division because only a single Class B allele has been used in previous studies (S3; Schierup et al. 2001). The purpose of this study was to examine dominance interactions and site of expression of alleles from the two classes in more detail to determine how sequence diversity relates to possible functional differences among alleles.

Materials and Methods

Genealogy Reconstruction

A gene tree similar to those described in Charlesworth et al. (2003a) and Mable, Beland, and Di Berardo (2004) was reconstructed using a Bayesian approach to demonstrate division of sequences (POPSET alignment gi:27545451) into previously defined “classes.” The SRK gene in A. lyrata was originally labeled Aly13 because it was discovered as part of a large gene family (Charlesworth et al. 2000; Schierup et al. 2001). Here, sequences that are linked to the SI phenotype will be labeled AISRK, whereas those that are not will retain the Aly13 designation. The tree was rooted with a paralogous gene from A. thaliana known to be expressed in vegetative tissues (Dwyer et al. 1994; Smythe and Ayscough 2003) and to play a role in development (ARK1), along with its putative ortholog in A. lyrata (Aly10.1) and two other orthologous pairs (ARK2/ Aly10.2; ARK3/Aly8) recently identified (Charlesworth et al. 2003b). The genealogy was reconstructed under a general time-reversible model with rate heterogeneity accounted for by estimating the proportion of invariant sites and the gamma-shape parameter (I + G), which was determined to be the most appropriate model of evolution according to Modeltest 2.0 (Posada and Crandall 1998) in conjunction with PAUP*: 4.0b10 (Swofford 1996.). A Bayesian tree was reconstructed using MrBayes version 3.01 (Huelsenbeck and Ronquist 2001) with the following parameter settings: number of generations = 1,000,000, sample frequency = 10, print frequency = 100, number
of chains = 4, and burn-in = 2,000 generations. Posterior probabilities were computed through consensus of resulting trees, after discarding those found during the initial burn-in period. The resulting consensus tree was visualized in PAUP* using likelihood settings as determined in MrModeltest.

Genotyping Plant Materials Used For Gene Expression and Phenotype Studies

Plants used for this study were produced from previously performed crosses between plants of known SI genotype (Mable, Schierup, and Charlesworth 2003, table 1). A “family” was defined as progeny grown from seeds of a single fruit resulting from a controlled cross, and each family was assigned a unique identifier. All families were the result of crosses between plants collected from the Rekjanes Peninsula in Iceland. SRK was used to infer S-genotypes because too few SCR sequences have yet been identified in A. lyrata to allow efficient screening using allele-specific primers. Linkage to the SI phenotype has previously been tested for all SRK sequences used in this study (Schierup et al. 2001; Mable, Schierup, and Charlesworth 2003).

Plants were grown in a greenhouse under 16 h light, 8 h darkness, with greenhouse temperatures fluctuating between 16°C and 27°C, depending on external meteorological conditions. Plants were fertilized weekly with a solution of 15–30–15 fertilizer (Miracle-Gro, Marysville, Ohio) to encourage flowering.

Genomic DNA for plant genotyping was extracted from leaf tissues using the FastDNA kit procedure and a FASTPREP tissue homogenizer (Bio101 Systems, Q-BIOgene, Irvine, Calif.). In order to examine segregation of SRK alleles based on parental genotypes, all progeny within families were genotyped (according to the procedures described in Schierup et al. 2001) using polymerase chain reaction (PCR) with allele-specific primers located in the extracellular S-domain (exon 1) and a general reverse primer located at the 3’ end of this domain (table 2). S-domain primers were used because they have been shown to be most reliable in specific amplification of target SRK alleles (Schierup et al. 2001). PCR products were run on a 1% agarose gel. The presence of a band of the expected size (table 2) was considered a positive test for presence of the allele.

To confirm identity, at least one example of each allele was sequenced and aligned manually to all previously characterized A. lyrata SRK alleles (POPSET gi:27545451), using Se-Al 2.0 (Rambaut 1996). Sequencing was performed on PCR bands purified using QIAquick gel extraction kits (Qiagen, Mississauga, ON), using the Big Dye terminator sequencing kit (Applied Biosystems, Foster City, Calif.), with separate reactions performed using the same forward or reverse primer used in the initial PCR. Sequencing reactions were run on an ABI 377 sequencer. Sequences were manually checked to ensure accuracy of base-calling using Sequencer version 4.1 (Gene Codes Corp, Ann Arbor, Mich.).

Phenotypic Expression: Controlled Cross-Pollinations to Infer Dominance Relationships

Plants of known SRK genotypes were crossed within families, to examine segregation of S-phenotypes with

### Table 1
Summary of Genotypes of Plant Families Used for Pollination Studies and Molecular Expression Studies

<table>
<thead>
<tr>
<th>Maternal Plant</th>
<th>Maternal Genotype</th>
<th>Paternal Plant</th>
<th>Paternal Genotype</th>
<th>Family Label</th>
<th>Number of Offspring</th>
</tr>
</thead>
<tbody>
<tr>
<td>99R 19/2</td>
<td>S₁S₅</td>
<td>99R 35/5</td>
<td>S₂S₂₂</td>
<td>BM02 A-10</td>
<td>10</td>
</tr>
<tr>
<td>99R 9/5</td>
<td>S₁S₁₄</td>
<td>99R 35/5</td>
<td>S₂S₂₂</td>
<td>BM02 A-14</td>
<td>5</td>
</tr>
<tr>
<td>99R 35/5</td>
<td>S₂S₂₂</td>
<td>99R 9/5</td>
<td>S₁S₁₄</td>
<td>BM02 A-15</td>
<td>5</td>
</tr>
<tr>
<td>99R 14/1</td>
<td>S₁₂S₁₆</td>
<td>99R 19/2</td>
<td>S₁S₅</td>
<td>BM02 H-1</td>
<td>14</td>
</tr>
<tr>
<td>99R 14/1</td>
<td>S₁₂S₁₆</td>
<td>99R 14/1</td>
<td>S₁₂S₁₆</td>
<td>BM02 H-2</td>
<td>9</td>
</tr>
<tr>
<td>99R 19/2</td>
<td>S₁S₅</td>
<td>99R 14/1</td>
<td>S₁₂S₁₆</td>
<td>BM02 H-3</td>
<td>11</td>
</tr>
<tr>
<td>99R 35/5</td>
<td>S₂S₂₂</td>
<td>99R 14/1</td>
<td>S₁₂S₁₆</td>
<td>BM02 H-4</td>
<td>14</td>
</tr>
<tr>
<td>99R 28/1b</td>
<td>S₂₅S₂₅</td>
<td>99R 19/2</td>
<td>S₁S₅</td>
<td>BM02 H-5</td>
<td>30</td>
</tr>
<tr>
<td>99R 19/2</td>
<td>S₁S₅</td>
<td>99R 28/1b</td>
<td>S₁₂S₂₂</td>
<td>BM02 H-6</td>
<td>21</td>
</tr>
<tr>
<td>Not known</td>
<td>Not known</td>
<td>BM03 P3b</td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

a Number of plants that flowered sufficiently to be used in pollination studies.

b Individual with sequence 13-7 in addition to the two alleles at the S-locus.

c Individual grown from a field-collected seed from Pinery Provincial Park, Ontario, with the genotype S₁S₁₄.

### Table 2
Primer Sequences and Product Size

<table>
<thead>
<tr>
<th>Name</th>
<th>Primer Sequence (5’-3’)</th>
<th>Size (bp)</th>
<th>Product</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>13-1F</td>
<td>TTTCAACAACAACTCAACAAGA</td>
<td>839</td>
<td>AISRK1</td>
<td>A</td>
</tr>
<tr>
<td>13-3F</td>
<td>GTATAAATCTGAAATGCTAG</td>
<td>700</td>
<td>AISRK3</td>
<td>B</td>
</tr>
<tr>
<td>13-6F</td>
<td>CTTACCCGAGTTGTAATCTCTAC</td>
<td>387</td>
<td>AISRK6</td>
<td>B</td>
</tr>
<tr>
<td>13-7F</td>
<td>GTGACTTCTCCAAGAAAATTTG</td>
<td>478</td>
<td>Alys13-7</td>
<td>U</td>
</tr>
<tr>
<td>13-9F</td>
<td>GAGGTCACTGACACGTCCTCTG</td>
<td>535</td>
<td>AISRK9</td>
<td>A</td>
</tr>
<tr>
<td>13-12F</td>
<td>CCAGATATGCGACACATGGACC</td>
<td>606</td>
<td>AISRK12</td>
<td>A</td>
</tr>
<tr>
<td>13-14F</td>
<td>CAGAGATGCAAAAGATGGAAAA</td>
<td>591</td>
<td>AISRK14</td>
<td>B</td>
</tr>
<tr>
<td>13-16F</td>
<td>TTTGGTTTGTTCACCATCTTGG</td>
<td>605</td>
<td>AISRK16</td>
<td>A</td>
</tr>
<tr>
<td>13-20F</td>
<td>ACAATGTTGGATATCAGGTCCTC</td>
<td>576</td>
<td>AISRK20</td>
<td>A</td>
</tr>
<tr>
<td>13-22F</td>
<td>ACACGTCTTTAATGCGCACAACAG</td>
<td>444</td>
<td>AISRK22</td>
<td>A</td>
</tr>
<tr>
<td>SLGR</td>
<td>ATCTGACATAAAAGATCTTGACC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP8F</td>
<td>ACGGGAACTCCTCAAAATATCCG</td>
<td>1,200</td>
<td>Alys10.1</td>
<td>U</td>
</tr>
<tr>
<td>SRK5R</td>
<td>ACAGCTTCTAATCTCATCATAATGGA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Size of targeted PCR product amplified using the allele-specific forward primers and SLGR, a general reverse primer located at the 3’ end of the S-domain (exon 1).

b Class A (A), Class B (B), or unlinked sequence from a paralogous locus (U).

c The Alys10.1 product was amplified using the forward primer AP8F with a reverse primer located in exon 3 of the kinase domain; the expected size is shown for cDNA.
Table 3

| Alleles S_A, S_B, S_C, and S_D are codominant in both stigma and anther |
|-----------------|-----------------|-----------------|
| S_A^S_A | S_A^S_B | S_A^S_C | S_A^S_D | S_B^S_B | S_B^S_C | S_B^S_D | S_C^S_C | S_C^S_D | S_D^S_D |
| 0^a | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 |

a Recipient genotype is provided in columns and donor genotype in rows.

b Shaded cells indicate crosses with less than 50% compatibility.

S-genotypes (i.e., compatibility between plants of particular SRK genotypes) and to allow inference of dominance relationships among S-alleles. Plants were covered with fine mesh to exclude pollinating insects for at least 3 days before pollinations began (before buds formed). Pollinations were performed by removing an anther from an open flower (pollen donor) using jeweler’s forceps and liberally applying pollen from this anther to the stigmatic surface of the second flower (pollen recipient). Between each pollen forceps were cleaned with 95% ethanol to kill any remaining pollen. After pollination the anthers were removed from the recipient flower. When flower availability permitted, three reciprocal replicates per cross combination for all possible crosses within families and six replicates of self-pollinations on each plant were performed.

Pollinations were scored based on fruit set. Fruits were collected once they had matured (approximately 3–7 weeks after pollination) to enable examination of progeny in future studies. For each collected fruit, length from pedicel to the tip of the persistent style (Gleason and Cronquist 1991) was measured using digital calipers, and the number of seeds within fruits was counted in order to assess whether any partial or reduced compatibility occurred.

Plants initially were separated into putative SI groups within families based on SI genotypes. The proportion of positive pollinations (pollinations resulting in fruit containing seeds) was calculated for each genotype, using only crosses for which fruit length and seed number information was available because subjective scoring was often inaccurate. Pollinations were classified as “generally negative” (resulting in fewer than 50% positive fruits) or “generally positive” (greater than 50% positive fruits) to identify SI genotype groups. Dominance was inferred by comparing expected compatibility of particular genotypic combinations under a completely codominant SI system (where any individuals sharing at least one SI allele would result in an infertile pollination to observed patterns (table 3)). By combining the inferences drawn from each family with previously published dominance predictions (Schierup et al. 2001; Mable, Schierup, and Charlesworth 2003; Bechsgaard, Bataillon, and Schierup 2004), it was possible to establish a dominance hierarchy for haplotypes in relation to their placement on the gene genealogy.

**RNA Experiments**

Plants were selected for RNA extraction based on the SRK genotypes established above, to comprise as many combinations of ALSRK alleles as possible. RNA extraction was conducted predominantly on two tissue types: rosette leaves and floral buds 1 to 2 days from flower opening (Kusaba et al. 2002). For some alleles, isolated stigmas were also tested, but few plants flowered sufficiently to allow collection of the large numbers required for unambiguous screening. To distinguish expression in vegetative from reproductive tissues, some tests were also conducted using buds with stigmas removed. Tissues were collected between dawn and midday using ethanol-washed forceps and placed in 1.5-ml microcentrifuge tubes. Five leaves were collected on the first collection day, but floral tissues (buds and stigmas) were collected over several days in order to obtain sufficient tissue for RNA extraction (20–30 flower buds or 40–70 stigmas). Collected tissues were frozen in liquid nitrogen and then stored at −80°C until RNA was extracted.

Frozen tissue was ground to a fine powder in the presence of liquid nitrogen, and between 80 and 110 mg of ground tissue was extracted with 1 ml of Trizol reagent according to the manufacturer’s instructions (Invitrogen, Burlington, ON), category number 15596018. The RNA pellet was resuspended in 100 µl of RNase-free water. RNA integrity was assessed by running samples on a 1% agarose gel electrophoresis and quantity was assessed spectrophotometrically (assuming that 40 µg/ml of RNA gives an absorbance value of 1 at 260 nm).

**Molecular Expression: Allele-Specific Reverse Transcription–PCR**

RNA samples were treated with DNase A (Invitrogen) to remove any residual DNA prior to reverse transcription (RT) using dT16AC (Sigma-Genosys, Mississauga, ON) and SuperScript II RNase H− reverse transcriptase, according to the manufacturer’s recommendations (Invitrogen). The resulting complementary DNA (cDNA) was stored at 4°C until required. PCR was performed using cDNA templates with the allele-specific strategy described for genomic DNA (table 2). The resulting RT-PCR products were run on 1% agarose gel with a 1-kb DNA ladder (MBI Fermentas, Burlington, ON) as a size standard. The presence of an amplified fragment of the expected size (table 2) for the allele-specific product was considered to be an indication of the presence of allele transcripts in the tissue tested. In addition, representative samples for each fragment were sequenced to confirm allele identity.

To test for the synthesis of cDNA (positive control), a reaction containing the rubisco primers Rubisco-H 680 and Rubisco-C 1322 (Nassuth et al. 2000) was included for each sample. To test for the absence of contaminating genomic DNA (negative control), reaction mixtures minus the RT enzyme were used. Only samples that tested positive and negative, respectively, were included in our analyses.
Results
Genealogy

The Bayesian genealogy was similar to a previous maximum likelihood tree described in Mable, Beland, and Di Berardo (2004), with strong posterior probability support for the clustering of Class B sequences with the unlinked sequences *Aly*13-2 and *Aly*13-7. The relatively short branch lengths within this group (due to the high similarity of nucleotide sequences) could suggest that they have originated and diverged more recently than the Class A sequences (fig. 1). As in previous analyses, relationships among Class A sequences were not as well resolved, possibly due to the very long branch lengths that separate them. However, two well-supported subgroups were apparent, one of which was more closely related to the Class B sequences than to the other Class A sequences (A3). The relative placement of *AISRK10* and *AISRK20* (both of which have been shown to be dominant alleles in previous studies: Schierup et al. 2001; Mable, Beland, and Di Berardo 2004) were not resolved.

Phenotypic Expression: Dominance Relationships Among S-haplotypes

Within-family crosses (table 4) showed a complex pattern of compatibility. In each set of reciprocal families there was an observable deviation from the expected pattern under a model of complete codominance of alleles (where only plants sharing no alleles would be compatible, table 3), with some evidence of nonreciprocal differences and an indication that viability differences beyond the SI reaction might be operating in some allelic combinations or in some individuals (table 4). A detailed description of the results from each cross can be found in Prigoda (2004); only the most relevant results will be pointed out here. Because the SI phenotype results from the combined
expression of male and female components, the standard S-haplotype notation will be used (which refers to the combined specificity of male and female components).

Previous studies have demonstrated that $S_1$ (Class A) is recessive to all other haplotypes with which it has been tested (Schierup et al. 2001; Mable, Schierup, and Charlesworth 2003; Bechsgaard, Bataillon, and Schierup 2004; Mable, Beland, and Di Berardo 2004). Presumably due to the high inbreeding depression observed in A. lyrata (Karkkainen et al. 1999), homozygous testers to clarify dominance relationships have been difficult to produce. The single $S_1S_1$ homozygote that flowered in family BM02 H-5 (table 4A) allowed clear inference that $S_1$ was recessive to both $S_6$ (Class B) and $S_{25}$ (Class A) in both pollen and stigma and supported the hypothesis that it is recessive to all other S-haplotypes. Using this information, family BM02 A-10 (table 4B) indicated that $S_6$ and $S_{22}$ (Class A) were both dominant to $S_9$ (Class B) and $S_1$ (Class A) in pollen and stigma, although there appeared to be some partial fertility (0.21) in crosses between a $S_{22}$ recipient and a $S_6S_{22}$ donor. Similarly, families BM02 H-1 and H-3 (table 4C) indicated that $S_{12}$ and $S_{16}$ (Class A) were dominant to $S_1$ (Class A) and $S_{10}$ (Class B) in pollen and stigma, but there appeared to be reduced fertility of compatible crosses involving $S_6$ and $S_{16}$ (below 70%). Families BM02 H-2 and H-4 (table 4D) included four different Class A haplotypes and showed a more complex pattern of compatibility. For example, $S_9$ and $S_{12}$ were coexpressed in stigmas, but either $S_{12}$ was dominant to $S_6$ in pollen or $S_{16}$ was dominant to $S_9$ in pollen but recessive or coexpressed in stigmas. Similarly, $S_{22}$ and $S_{12}$ were coexpressed in both pollen and stigma but showed some partial compatibility in both directions. $S_{22}$ was dominant to $S_{16}$ but there appeared to be some partial reduction in fertility in stigmas. Few plants flowered sufficiently to be tested in BM02 A-14 and A-15, but table 4E does show that $S_{14}$ (Class B) was recessive in pollen and stigma to $S_9$ and/or $S_{22}$ but dominant to $S_1$.

Molecular Expression: Allele-Specific RT-PCR

The assay was first tested using Aly101.1, whose ortholog in A. thaliana (ARK1) is known to be expressed vegetatively and to play a role in development (Tobias and Nasrallah 1996; Pastuglia et al. 2002), and AlSRK20, whose expression is known to be restricted to reproductive tissues (referred to as SRKb; Kusaba et al. 2001). The results were as expected: Aly10.1 showed strong expression in leaves and buds (vegetative tissues) but weaker expression in stigmas (reproductive tissues), whereas expression of AlSRK20 was restricted to buds.

Expression of all other Class A alleles was restricted to reproductive tissues (buds and stigmas), as previously predicted for SI sequences (table 5, fig. 2A). In contrast, Class B SRK alleles showed strong expression in both leaves and buds (or stigmas): the identical pattern to that seen for the unlinked sequence Aly13-7 (table 5, fig. 2B). Expression patterns were not altered when dominant and recessive alleles were found together (i.e., both were expressed; fig. 2A), and the recessive allele AlSRK1 showed equally strong expression as that seen for the dominant alleles tested. Alletic identity was confirmed by sequencing for all except AlSRK9 and AlSRK22 (table 5).

### Discussion

**Phenotypic Expression: Differences in Dominance of Class A and B Haplotypes**

Synthesis of results from cross-pollinations presented in this paper and others (Schierup et al. 2001; Mable, Schierup, and Charlesworth 2003; Bechsgaard, Bataillon, and Schierup 2004) emphasize the high level of complexity in dominance relationships among S-haplotypes in A. lyrata, and support a hypothesis of division of S-haplotypes into

---

**Table 4**

Percent Compatibility of Diallel Crosses Within Full-Sib Families Described in Table 1

(A) BM02 H-5 Family

<table>
<thead>
<tr>
<th></th>
<th>S1S1</th>
<th>S1S6</th>
<th>S1S22</th>
<th>S6S22</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1S_1$</td>
<td>0.00b</td>
<td>0.89</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$S_1S_6$</td>
<td></td>
<td></td>
<td>0.06</td>
<td>0.91</td>
</tr>
<tr>
<td>$S_1S_{22}$</td>
<td>1.00</td>
<td>0.26</td>
<td>0.86</td>
<td>0.38</td>
</tr>
<tr>
<td>$S_6S_{22}$</td>
<td>1.00</td>
<td>0.90</td>
<td>0.16</td>
<td>0.38</td>
</tr>
</tbody>
</table>

(B) BM02 A-10

<table>
<thead>
<tr>
<th></th>
<th>S1S9</th>
<th>S22S9</th>
<th>S22S9</th>
<th>S22S9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1S_9$</td>
<td>0.03</td>
<td>0.83</td>
<td>0.08</td>
<td>0.91</td>
</tr>
<tr>
<td>$S_1S_{22}$</td>
<td>0.86</td>
<td>0.00</td>
<td>0.78</td>
<td>0.36</td>
</tr>
<tr>
<td>$S_22S_9$</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$S_22S_{22}$</td>
<td>0.85</td>
<td>0.21</td>
<td>1.00</td>
<td>0.08</td>
</tr>
</tbody>
</table>

(C) BM02 H-1 and H-3

<table>
<thead>
<tr>
<th></th>
<th>S1S12</th>
<th>S1S16</th>
<th>S22S12</th>
<th>S22S16</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1S_{12}$</td>
<td>0.90</td>
<td>0.03</td>
<td>0.77</td>
<td>0.90</td>
</tr>
<tr>
<td>$S_1S_{16}$</td>
<td>0.06</td>
<td>0.85</td>
<td>0.05</td>
<td>0.83</td>
</tr>
<tr>
<td>$S_{12}S_{16}$</td>
<td>0.67</td>
<td>0.05</td>
<td>0.63</td>
<td>0.04</td>
</tr>
</tbody>
</table>

(D) BM02 H-2 and H-4

<table>
<thead>
<tr>
<th></th>
<th>S22S12</th>
<th>S22S16</th>
<th>S22S16</th>
<th>S22S16</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{22}S_{12}$</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
<td>0.74</td>
</tr>
<tr>
<td>$S_{22}S_{16}$</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
</tr>
<tr>
<td>$S_{22}S_{12}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$S_{22}S_{16}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

(E) BM02 A-14 and A-15

<table>
<thead>
<tr>
<th></th>
<th>S1S9</th>
<th>S1S22</th>
<th>S1S22</th>
<th>S1S22</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1S_9$</td>
<td>0.45</td>
<td>—</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>$S_1S_{22}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$S_{16}S_9$</td>
<td>0.11</td>
<td>—</td>
<td>0.19</td>
<td>0.92</td>
</tr>
<tr>
<td>$S_{16}S_{22}$</td>
<td>0.73</td>
<td>—</td>
<td>0.91</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*a* Recipient genotype is provided in columns and donor genotype in rows.

*b* Number of individuals that flowered sufficiently to be included in pollinations.

*c* Shaded cells indicate crosses with less than 50% compatibility.
four dominance groups (fig. 3). These dominance groups have been labeled A1, A2, A3, and B to emphasize the finding that the divergent and disperse Class A SRK sequences encompass three of these groups, whereas all Class B sequences tested so far are recessive to all Class A haplotypes except AISRk1 (see below). The division of Class A sequences into two dominant haplotypes is based on the finding that A2 haplotypes tend to be codominant with one another and dominant to all haplotypes in the other three groups, whereas A3 haplotypes show more complex dominance relationships, with different dominance in pollen and stigma, partial or reduced compatibility in some combinations, and clear dominance to the A1 and B groups.

A finding that was not evident based on previous studies is that this dominance hierarchy may be related to S-domain–based sequence divergence, as has been found for cultivated Brassica (Chen and Nasrallah 1990; Gaude et al. 1995; Kusaba et al. 1997). In Brassica, Class I alleles are pollen dominant to Class II alleles, with a clear and well-supported separation of the types based on synthesis of information presented here and in previous studies (S9, S13, S16, S22, S23, S25, and S28 were tested in Schierup et al. 2001; those of S34, S29, S30, and S33 were tested in Mable, Beland, and Di Berardo 2004). Haplotypes indicated in boldface are those with uncertain dominance relationships but whose dominance classification is predicted based on their relative position in the genealogy (see fig. 1).

Table 5
Summary of Allele-Specific RT-PCR Expression Assays

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Number of Individuals</th>
<th>Site of Expression</th>
<th>Sequence type</th>
</tr>
</thead>
<tbody>
<tr>
<td>AISRk3</td>
<td>1</td>
<td>Leaf and bud</td>
<td>Class B</td>
</tr>
<tr>
<td>AISRk6</td>
<td>6</td>
<td>Leaf*, bud (2), and stigma (4)</td>
<td>Class B</td>
</tr>
<tr>
<td>AISRk14</td>
<td>4</td>
<td>Leaf and bud*</td>
<td>Class B</td>
</tr>
<tr>
<td>AISRk1</td>
<td>5</td>
<td>Bud* (4) and stigma (1)</td>
<td>Class A</td>
</tr>
<tr>
<td>AISRk9</td>
<td>2</td>
<td>Bud</td>
<td>Class A</td>
</tr>
<tr>
<td>AISRk12</td>
<td>3</td>
<td>Stigma*</td>
<td>Class A</td>
</tr>
<tr>
<td>AISRk16</td>
<td>1</td>
<td>Bud</td>
<td>Class A</td>
</tr>
<tr>
<td>AISRk20</td>
<td>3</td>
<td>Bud*</td>
<td>Class A</td>
</tr>
<tr>
<td>AISRk22</td>
<td>1</td>
<td>Bud</td>
<td>Class A</td>
</tr>
<tr>
<td>Ayl13-7</td>
<td>2</td>
<td>Leaf and bud</td>
<td>Unlinked paralog</td>
</tr>
<tr>
<td>Ayl10.1</td>
<td>2</td>
<td>Leaf</td>
<td>Unlinked paralog</td>
</tr>
</tbody>
</table>

* Number of individuals screened for expression in both leaves and buds (or stigmas); for alleles where only a single individual was used, multiple independent RT-PCR reactions were performed.

* Asterisks indicate products that were sequenced to confirm allelic identity.
As suggested in other studies, haplotype \( S_1 \), which has been found to have a high transmission advantage in a previous study of segregation ratios (Bechsgaard, Bataillon, and Schierup 2004), is the only representative found to date of the most recessive group (A1). \( S_1 \) also appears to be the most common allele, appearing in nearly all natural populations studied (Mable, Schierup, and Charlesworth 2003; Mable et al. 2005). A single, common “most recessive” allele is expected under some selection scenarios (Bateman 1947; Uyenoyama 2000a), although the situation is complicated by population substructure and other demographic factors (Schierup, Vekemans, and Christiansen 1998; Schierup, Vekemans, and Charlesworth 2000). Although \( S_1 \) clusters with Class A1 alleles in the genealogy (fig. 1), a kinase domain has not been found, suggesting that it might be unable to cause the downstream reactions that would lead to rejection of self-pollen. Nevertheless, the limited crosses performed with the \( S_1S_1 \) homozygote indicate that it is likely functional in a homozygous situation. A survey of the ovaries of flowers from an \( S_1S_1 \) self-pollination would confirm whether they had been fertilized, which could indicate that the lack of fruits from such self-pollinations were due to later-acting incompatibilities unrelated to the SI reaction occurring at the surface of the stigma (possibly related to inbreeding depression).

Previous investigations of the mechanisms for dominance in A. lyrata have focused on the relationships between Class A alleles in the two most dominant classes. The haplotypes tested by Kusaba et al. (2002) correspond to our alleles \( S_{20} \) (Sa), which is in the most dominant A2 class and \( S_{13} \) (Sb), which is in the A3 class. There has been no investigation of the mechanism of recessivity of the two most recessive classes. As there has been no indication to date of more than two dominance classes in Brassica, these two additional recessive classes in A. lyrata provide a novel opportunity to investigate dominance in multialele systems. It might also be possible that additional dominance classes in Brassica exist and have been overlooked because alleles in these classes do not conform to the expected expression patterns. The high level of complexity in the relationships between SI alleles in A. lyrata as compared to those of Brassica (Schierup et al. 2001) may indicate that loss of alleles in Brassica (due to the selection bottlenecks associated with artificial selection) may have obscured some of these relationships and emphasizes the potential advantages of studying evolution of multialele systems in noncultivated models. Comparison of recently described SRK sequences from Arabidopsis halleri (Castric and Vekemans 2004) suggests that the division of Class A and Class B sequences occurred before species separation, as representatives of both types have been found; some of these are nearly identical to the A. lyrata sequences and some are unique and could be used to test dominance predictions.

Molecular Expression: Different Sites of Expression for Class A and B Alleles

The expression of Class A SRK alleles followed the previously expected expression pattern for female-determining SI alleles; namely, expression in female floral tissues (stigmas) but not in vegetative tissues (leaves) (e.g., Hiscock et al. 2003). This is supported by similar results from previous expression studies (Kusaba et al. 2001) using SRK\( a \) (AISRK13) and SRK\( b \) (AISRK20). However, some alleles of the Brassica SRK gene also exhibit limited expression in anther tissues (Stein et al. 1991), indicating that exclusively pistil-specific expression may not be the case for all female-determining SI alleles.

Gene-expression profiles of Class B alleles are presented here for the first time. These alleles are tightly linked to the SI phenotype (Charlesworth et al. 2000; Schierup et al. 2001; Mable, Schierup, and Charlesworth 2003; Mable, Beland, and Di Berardo 2004), indicating that they represent alleles of SRK for these haplotypes. Amplification of their S-domain shows that expression of the Class B alleles studied was not limited to female floral tissues but was also present in vegetative tissues (leaves). This expression pattern, while different from the pattern seen with Class A alleles, is identical to the pattern seen with the unlinked sequence Aly13-7 to which they are closely related (fig. 1). More broadly, these results indicate that when tissue-specific expression is used as a criterion for identifying SI genes, some genes may be incorrectly discounted. For example, the search for SRK homologs in non-Brassica species, such as Senecio squarifolius and Ipomoea sp., has focused on genes that show expression in pistil tissues only and has rejected genes that appear similar in structure to SRK and show linkage to the S-phenotype but have unexpected expression patterns (Kowyaama, Tsuchiya, and Kakeda 2000; Hiscock et al. 2003). It is thought that A. lyrata does not contain SLG, but this conclusion is based on only two Class A haplotypes (\( S_{13} \) and \( S_{20} \) (Kusaba et al. 2001). Tests for transcripts containing the kinase-domain have so far been inconclusive (see Prigoda 2004). However, rather than detracting from the results presented here, if the haplotype structure itself is altered compared to that of Class A, this would emphasize that Class B sequences may have experienced a distinctive evolutionary history. Structural rearrangements among haplotypes are common in Brassica (Boyes et al. 1997; Ciu et al. 1999; Suzuki et al. 1999; Shiba et al. 2003) and have been described for the two haplotypes so far studied in A. lyrata (Kusaba et al. 2001).

Currently, no information exists regarding the structure of the cis-acting enhancers (promoters) of the SRK genes in A. lyrata. Such enhancer elements are known to direct timing and location of gene expression and are located on the same DNA strand as the gene they affect (Maniatis, Goodboum, and Fischer 1987; Atchison 1988). Because the Class B alleles have a similar expression pattern as Aly13-7, which is different from that of Class A alleles, it is tempting to speculate that a recombination event brought not only the coding region but also at least part of the region surrounding the S-genes from that containing the Aly13-7 allele. It is, of course, also possible that the Class B alleles gave rise to the Aly13-7 locus or that the two arose from an unidentified common ancestor. Because Class B alleles and an ortholog of the Aly13-7 sequence are also present in A. halleri (X. Vekemans, personal communication), confirming the expression patterns in this species could provide insights into whether the more ubiquitous
expression pattern arose in *A. lyrata* or whether it was present in the origination of the Class B types.

Although recent evidence suggests that minor sequence variation in either male or female components may not alter specificity (Chookajorn et al. 2004) to maintain SI (i.e., to recognize and reject self-pollen), new specificities arising in SRK can only be functional if corresponding changes occur in the regions of SCR that bind to the appropriate recognition sites. One would predict, therefore, that gene conversion with a paralogous locus could only lead to new mating types if the entire haplotype was converted rather than exchange of information between single paralogs. Evidence in support of such a large-scale conversion event would require identification of a SCR homolog in proximity to the *Aly*13-7 SRK gene. Comparative mapping approaches, such as that recently used to align *A. lyrata* and *A. thaliana* chromosomal regions (Kuittinen et al. 2004) and/or chromosome painting (Lysak et al. 2001) could be used to evaluate this hypothesis. Whatever the nature of their origin, the relatively short branch lengths within the Class B group may reflect more recent recruitment to the S-locus, followed by diversification due to the pressures of balancing selection.

**Distribution of CpG Islands Is Consistent with Recombination Outside the S-Gene Region**

For genes at the MHC, the suggestion that polymorphism may be increased through intralocus and/or interlocus gene conversion has been somewhat controversial (e.g., Kuhner et al. 1991; Martinsohn et al. 1999; Richman et al. 2003; Reusch, Schaschi, and Wegner 2004), partly due to the difficulty of associating phenotypes with allelic variants at particular loci. SI systems, on the other hand, have a clearly recognizable phenotype (blocking or allowing pollen tube growth and subsequent fruit development) that can be associated with particular allelic variants through controlled crossing studies (e.g., Charlesworth et al. 2000) and thus could provide a better model to evaluate the role of gene conversion in cell-cell recognition systems. For MHC systems, it has been suggested that recombination (gene conversion) is statistically associated with areas rich in CpG islands (i.e., GC-rich regions containing a high concentration of CpG dinucleotides) (Högstrand and Böhme 1999; Reusch, Schaschi, and Wegner 2004). BAC clone T6K22 of the *A. thaliana* genome contains the pseudogene orthologs of the *S*-genes from *A. lyrata* (SRKψ and SCRψ; Kusaba et al. 2001; fig. 4). Preliminary analysis of the distribution of CpG dinucleotides and predicted CpG islands (found using the program CpG Island searcher http://www.cpgisland.com: Takai and Jones 2003) in this region in relation to the *S*-gene equivalents supports the view that recombination should be limited in this region. Overall, there appears to be a paucity of CpG islands in the region where the *S*-gene orthologs are contained (fig. 4), except at the end of the *S*-domain in SRKψ, where gene conversion has been implicated in homogenization of SLG to the corresponding SRK S-domain sequence (e.g., Cabrillac et al. 1999). Using the more rigorous threshold for detection used by others (Högstrand and Böhme 1999; Reusch, Schaschi, and Wegner 2004), the only areas detected were at either end of the BAC clone, which could indirectly support the hypothesis that gene conversion involving the entire haplotype region has occurred.

**Fig. 4.**—Histogram showing the distribution of CpG dinucleotides and predicted CpG islands across the sequence of *Arabidopsis thaliana* BAC clone T6K22, which contains pseudogene orthologs to *Arabidopsis lyrata* genes SRK and SCR (indicated as SRKψ and SCRψ, respectively). The region has one complete copy of SRK (SRKψ) and SCR (SCRψ1) and several fragmented copies of SCR (SCRψ2 and SCRψ3). The Web-based program CpG island searcher (http://www.cpgisland.com) was used to establish the frequency of CpG dinucleotides in adjacent windows of 800 bp along the sequence (histogram bars) and to predict the locations of CpG islands (GC-rich regions containing a higher proportion of CpG dinucleotides in a 200-bp region than expected) based on two stringency thresholds (*: lower limit of 50% GC; observed/expected CpG ratio 0.60; **: lower limit of 55% GC; observed/expected CpG ratio 0.67). The upper threshold has been used as the criterion in previous papers suggesting a statistical association between CpG islands and recombination frequency.
Conclusions

Combined mapping of dominance relationships and expression patterns onto a genealogy of relationships among S-alleles in *A. lyrata* provides provocative evidence that alleles at the S-locus may have experienced different types of origins and/or evolutionary histories that determine their current dominance interactions. The close sequence-based relationship of the highly recessive $S_1$ allele (A1) to the most dominant class A alleles (A2) could suggest that these dominance classes were in place before recruitment of the A3 and B groups. Although the precise mechanism of such recruitment remains a mystery, the possibility that large-scale gene conversion events could be involved deserves more attention. The involvement of other members of the serine-threonine kinase superfamily in cell-cell recognition roles such as development and pathogen response systems (Shiu and Bleecker 2001; Pastuglia et al. 2002; Smythe and Ayscough 2003) suggests that they might also require coordinated development and pathogen response systems (Shiu and Bleecker 2001; Pastuglia et al. 2002; Smythe and Ayscough 2003) suggests that they might also require coordinated


2003) suggests that they might also require coordinated

development and pathogen response systems (Shiu and Bleecker 2001; Pastuglia et al. 2002; Smythe and Ayscough 2003) suggests that they might also require coordinated

The involvement of other members of the serine-threonine kinase superfamily in cell-cell recognition roles such as development and pathogen response systems (Shiu and Bleecker 2001; Pastuglia et al. 2002; Smythe and Ayscough 2003) suggests that they might also require coordinated development and pathogen response systems (Shiu and Bleecker 2001; Pastuglia et al. 2002; Smythe and Ayscough 2003) suggests that they might also require coordinated

Acknowledgments

The authors thank B. Husband and L. Lukens for comments on a previous version of the manuscript and continued advice throughout the project. L. Witham, J. Beland, C. Di Berardo, P. Kron, M. Bannerman, S. Dart, and S. Richer provided assistance with plant rearing, pollinations, tissue collection, and genotype screening. The comments of two anonymous reviewers significantly improved the manuscript. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), The Canadian Foundation for Innovation (CFI), and the Ontario Innovation Trust (OIT).

Literature Cited


Schierup, M. H., B. K. Mable, P. Awadalla, and D. Charlesworth. 2001. Identification and characterization of a polymorphic


Brandon Gaut, Associate Editor

Accepted May 14, 2005