

Diversity and Linkage of Genes in the Self-Incompatibility Gene Family in *Arabidopsis lyrata*

Deborah Charlesworth,^{*,1} Barbara K. Mable,^{*,2} Mikkel H. Schierup,[†] Carolina Bartolomé* and Philip Awadalla^{*,3}

^{*}Institute of Cell, Animal and Population Biology, University of Edinburgh, Ashworth Laboratories, Edinburgh EH9 3JT, United Kingdom and [†]Department of Ecology and Genetics, University of Aarhus, DK-8000 Aarhus C, Denmark

Manuscript received July 5, 2002
Accepted for publication March 19, 2003

ABSTRACT

We report studies of seven members of the S-domain gene family in *Arabidopsis lyrata*, a member of the Brassicaceae that has a sporophytic self-incompatibility (SI) system. Orthologs for five loci are identifiable in the self-compatible relative *A. thaliana*. Like the Brassica stigmatic incompatibility protein locus (*SRK*), some of these genes have kinase domains. We show that several of these genes are unlinked to the putative *A. lyrata SRK*, *Aly13*. These genes have much lower nonsynonymous and synonymous polymorphism than *Aly13* in the S-domains within natural populations, and differentiation between populations is higher, consistent with balancing selection at the *Aly13* locus. One gene (*Aly8*) is linked to *Aly13* and has high diversity. No departures from neutrality were detected for any of the loci. Comparing different loci within *A. lyrata*, sites corresponding to hypervariable regions in the Brassica S-loci (*SLG* and *SRK*) and in comparable regions of *Aly13* have greater replacement site divergence than the rest of the S-domain. This suggests that the high polymorphism in these regions of incompatibility loci is due to balancing selection acting on sites within or near these regions, combined with low selective constraints.

IN Brassica, control of pollen-stigma interactions at the stigmatic interface involves highly polymorphic recognition genes of the self-incompatibility (SI) system. It is of interest to understand which regions of the proteins that these genes encode have recognition functions, how this affects the polymorphism in the coding sequence and surrounding genome regions, and how the two recognition genes maintain their coadaptation to produce functional incompatibility types. To understand the evolution of the self-incompatibility loci, it will be helpful to study them in the context of the gene families to which they belong. Doing this allows one to evaluate the possibility of exchanges between loci by gene conversion. It also makes it possible to compare sequence evolution of loci that are involved in incompatibility, and are thus under balancing selection, with similar sequences not under such selection.

The S-locus region contains loci belonging to two distinct gene families. "S-domain genes," members of the plant receptor-like protein-kinase gene family (BOYES and NASRALLAH 1993; WALKER 1994), are important for the stigma recognition functions, while pollen specific-

ties are controlled by the *SCR* gene (S-locus cysteine rich, also called *SP11*), a member of the pollen coat protein *PCP* family (STEPHENSON *et al.* 1997; DOUGHTY *et al.* 1998; SCHOPFER *et al.* 1999; TAKAYAMA *et al.* 2000; VANOOSTHUYSE *et al.* 2001). These different genes are linked in a region whose length differs among different haplotypes (BOYES and NASRALLAH 1993; GORING and ROTHSTEIN 1996; YU *et al.* 1996; CONNER *et al.* 1998). In the Brassica S-locus region, there are two S-domain genes, *SLG* (S-locus glycoprotein) and *SRK* (S-receptor kinase), with expression chiefly in stigma epidermal cells. The *SRK* gene is essential for self-incompatibility (GORING and ROTHSTEIN 1996; CUI *et al.* 2000; TAKASAKI *et al.* 2000), while *SLG*, a closely linked S-domain gene without a kinase domain, is homologous to exon 1 of *SRK* and is nonessential for recognition, although it may have a role in the incompatibility phenotype (TAKASAKI *et al.* 2000).

Further S-domain genes are known in Brassica and related plants (LUU *et al.* 2001), most of them not linked to the S-locus (KAI *et al.* 2001), although linked ones are found in some haplotypes (SUZUKI *et al.* 1997, 1999; KUSABA *et al.* 2000), some of them apparently pseudogenes (YU *et al.* 1996; KAI *et al.* 2001). Most of the functional members of this gene family are presumably involved in processes other than pollination, although in Brassica some encode secreted stigma glycoproteins (NASRALLAH and NASRALLAH 1993) and some S-domain proteins are necessary for correct pollen-stigma adhesion (LUU *et al.* 1997; TAKAYAMA *et al.* 2000). In *Arabi-*

¹Corresponding author: Institute of Cell, Animal and Population Biology, University of Edinburgh, Ashworth Laboratories, King's Bldgs., West Mains Rd., Edinburgh EH9 3JT, United Kingdom.
E-mail: deborah.charlesworth@ed.ac.uk

²Present address: Department of Botany, University of Guelph, Guelph, ON N1G 2W1, Canada.

³Present address: Center for Population Biology, University of California, Davis, CA 95616.

dopsis thaliana the S-domain gene family has ~40 members (SHIU and BLEECKER 2001), and the Brassica *SRK* sequences are most similar to the *A. thaliana* *Ark* genes (PASTUGLIA *et al.* 2002). Other members of this gene family are not linked to the Slocus.

Studies of sequence diversity of Brassica Slocus genes have until recently concentrated on the *SLG* gene, but some data from the S-domains and a portion of the kinase domain of *SRK* of some haplotypes have been published (KUSABA *et al.* 1997; NISHIO *et al.* 1997). The broader view of the Slocus genes as members of a gene family has not been emphasized in Brassica, although some data are from loci other than *SLG* and *SRK*, and it is clear that both *SRK* and *SLG* genes are much more polymorphic than other S-domain genes that have been studied (DWYER *et al.* 1991; HINATA *et al.* 1995; KUSABA *et al.* 1997; WATANABE *et al.* 1997, 1998; SAKAMOTO *et al.* 1998). For loci other than *SLG* and *SRK*, sample sizes are very small. *SCR* also appears to be highly polymorphic (WATANABE *et al.* 2000; KIMURA *et al.* 2002), although no comparison of diversity with other members of this gene family has been published. The Brassica data are mainly from cultivars, not from random samples from natural populations.

Here we report results of population genetic studies of several S-domain loci in natural populations of *A. lyrata*. We characterize diversity at several different S-domain loci for comparison with *SRK* to establish whether *SRK* indeed has an unusually polymorphic S-domain, as expected for a gene under balancing selection. Balancing selection is not expected for S-domain genes that are not involved in SI (although they could have experienced other forms of selection, for instance, directional selection subsequent to gene duplication in the evolution of the gene family).

A. lyrata is a self-incompatible, predominantly diploid member of the Brassicaceae, but distantly related to Brassica (ROLLINS 1993). Silent site divergence of putatively orthologous genes between *A. lyrata* or *A. thaliana* and Brassica ranges from ~0.2 to >1 without Jukes-Cantor correction (reviewed in WRIGHT *et al.* 2002). We previously described several S-domain loci in *A. lyrata* (CHARLESWORTH *et al.* 2000; SCHIERUP *et al.* 2001). We refer to these as the *Aly* loci. Of these, *Aly13* is highly polymorphic within *A. lyrata*, and segregation with incompatibility groups in families suggests that it is the ortholog of the Brassica *SRK* gene (CHARLESWORTH *et al.* 2000; SCHIERUP *et al.* 2001). The same gene was identified by KUSABA *et al.* (2001), who isolated the complete sequence, including the kinase domain, from stigma mRNA of an *A. lyrata* plant heterozygous for two S-alleles and also showed cosegregation of sequence variants with the two progeny incompatibility groups in self-fertilized progeny; they named the gene *A. lyrata* *SRK*. Two sequences in our families closely match those of the two alleles in the plant studied by KUSABA *et al.*

(2001); *Aly13-13*'s S-domain is almost identical to that of the *SRKa* allele, while *Aly13-20* matches allele *SRKb*.

As expected for the Slocus, *Aly13* sequences are exceptionally polymorphic at both synonymous and replacement sites (SCHIERUP *et al.* 2001). S-domain diversity is even higher than in the S-domains of Brassica *SRK* or *SLG* loci (HINATA *et al.* 1995; CHARLESWORTH and AWADALLA 1998). To evaluate the variability, it is important to compare it with that of other loci. As currently few diversity data are from natural populations of plants, including *A. lyrata*, we have obtained new diversity data by studying other S-domain loci, which are the ideal "reference" loci for tests of whether *Aly13* is more polymorphic than other loci.

Data from other *Aly* loci also allow us to compare levels of selective constraint in different regions of the S-domain. It is often suggested that the hypervariable (HV) regions in the extracellular S-domain are the most important for recognition (NASRALLAH and NASRALLAH 1989; KIMURA *et al.* 2002). However, this remains uncertain, and effects of amino acid differences elsewhere in the domain are also likely (NASRALLAH 1997; MIEGE *et al.* 2001). Variability in the Slocus region will be affected by a number of factors, and it is important to distinguish the possibilities clearly.

First, in regions of the protein where amino acid variants alter specificities, balancing selection will promote variation, so we expect high nonsynonymous diversity in the genomic sequences, as observed in both Brassica (*e.g.*, NASRALLAH and NASRALLAH 1989) and *A. lyrata* (SCHIERUP *et al.* 2001). Second, closely linked synonymous sites will also have high diversity, because balancing selection maintains different functional alleles for long time periods (*e.g.*, WRIGHT 1939; VEKEMANS and SLATKIN 1994), allowing sequence differentiation of alleles (*e.g.*, STROBECK 1972; HUGHES *et al.* 1990; TAKAHATA 1990; NORDBORG *et al.* 1996); this prediction also fits the data on S-alleles (NASRALLAH and NASRALLAH 1989; CHARLESWORTH and AWADALLA 1998; SCHIERUP *et al.* 2001). Amino acid variants may be affected in the same way, so that such variants need not all be associated with specificity differences. For the same reason, other genome regions closely linked to the Slocus may also have high polymorphism. A number of cases of additional S-domain genes located close to the Brassica Slocus are known, and it is of interest to see whether they are highly polymorphic (as expected if linkage to the Sloci is very close) or have low diversity (which would imply that recombination occurs and that diversity is high only at sites that are physically very close to sites under balancing selection). We have already mentioned the highly polymorphic *SLG* locus in the Brassica Slocus region, but some other linked genes have been reported to have low diversity (HINATA *et al.* 1995). Among the *Aly* loci studied here, we find no *A. lyrata* ortholog of *SLG* (in agreement with KUSABA *et al.*

2001), but we observe high polymorphism in a different linked S-domain gene (*Aly8*).

A third important influence on diversity is that selective constraints may differ between different regions of the protein (NASRALLAH 1997). Low selective constraints may allow certain regions to have particularly high polymorphism due to linkage to the sites under balancing selection. Different selective constraints can be detected using comparisons with reference loci, and we use this approach to test whether this could contribute to the high variability of the hypervariable regions. A sample of seven putative alleles showed that the *Aly13* S-domain sequences have peaks of replacement site polymorphism in the regions equivalent to the Brassica Slocus HV regions (SCHIERUP *et al.* 2001). To test whether this is because the same regions are important for recognition functions in both species, we must exclude the possibility that these regions are low constraint regions.

A final reason for studying other loci is that the analysis of sequence data to infer selection is complicated by population subdivision. To assess the effects of demographic and historical processes that can generate patterns that may be mistaken for evidence of selection, it is necessary to have reference loci that are not under strong balancing selection, but instead are evolving more or less neutrally. For example, genetic differentiation between populations can cause haplotype structure that may be difficult to distinguish from balancing selection unless other data are available to show the true situation (*e.g.*, CHARLESWORTH *et al.* 1997). Tests for selection such as TAJIMA's (1989a) test are thus affected by population subdivision and will often be unable to detect selection when samples from such populations are pooled (SCHIERUP *et al.* 2000). When there is subdivision, this test cannot distinguish between balancing selection and differentiation between populations, because Tajima's *D* is affected by both selection and population subdivision (which, like balancing selection, causes positive *D* values; TAJIMA 1989a,b). Directional selection at a locus causes negative *D* values, and pooling is therefore conservative (ignoring subdivision may obscure this form of selection). Data on diversity at the *Aly13* locus (the putative *A. lyrata* *SRK*) must therefore be interpreted in the light of such information. Finally, a locus experiencing balancing selection is expected to show less population subdivision than loci not experiencing such selection (SCHIERUP *et al.* 2000).

Here we describe analyses of diversity and selection at several *A. lyrata* S-domain loci and assess the implications for our understanding of balancing selection and its effects on sequence diversity within S-loci and in their genomic neighborhood.

MATERIALS AND METHODS

A. *lyrata* plant material and DNA preparation: Seeds were collected from four populations of *A. lyrata* (see details in

CHARLESWORTH *et al.* 2000). *A. lyrata* populations are widely distributed in northern North America, and European populations formerly classified as *A. petraea* are now considered to be the same species as *A. lyrata*, so both are referred to here as *A. lyrata* (following KOCH *et al.* 2000). The populations, and the abbreviations we shall use for them, were as follows: Two samples from North America (kindly provided by R. Mauricio) were from North Carolina (NC) and Indiana Dunes, Indiana (IN), and two from Europe were from Braemar, Scotland (kindly provided by R. Ennos), and the Reykjanes Peninsula, Iceland (some of them provided by E. Thorhallsdottir). Individual plants were grown in the greenhouse from seeds from these populations. DNA was extracted from leaves using a CTAB protocol (JUNGHANS and METZLAFF 1990).

Primers, amplification, cloning, and sequencing: *S-domain primers:* Primers were designed on the basis of sequence alignments of Brassica *SLG* and *SRK* loci (Table 1) and used to amplify *A. lyrata* genomic DNA. Because *SLG* and *SRK* are members of a gene family, our initial primers were based on the most conserved regions of the Brassica S-domain and should amplify multiple *A. lyrata* S-domain genes, particularly those most similar to *SRK*. The S-domains of most Brassica *oleracea* and *B. campestris* *SLG* and *SRK* alleles have no introns (TANTIKANJANA *et al.* 1993; HATAKEYAMA *et al.* 1998; CABRILLAC *et al.* 1999), so *A. lyrata* S-domains should be similar in length to those of Brassica. To distinguish between sequences from different loci, specific primers were designed from the *A. lyrata* sequences amplified. The primer sequences and amplification conditions are given in CHARLESWORTH *et al.* (2000).

Kinase domain reverse primers: To test whether each S-domain sequence had a kinase domain downstream from the S-domain, we used specific forward primers for the loci identified in *A. lyrata* with reverse primers based on either Brassica *SRK* locus kinase domains (*srk4r* and *srk5r*) or an *A. lyrata* *SRK* kinase sequence kindly provided by J. B. Nasrallah (*srknasr1*, *srknasr4*, and *srknasr3*; see Table 1).

Cloning and sequencing: Because some primers amplify more than one locus, and also because of the high variability of some of the putative loci (see below), PCR products of the expected size were generally cloned before sequencing [using the Invitrogen (San Diego) TOPO TA cloning kit]. To detect sequence variants and differentiate between loci (see below), the cloned amplification products were digested with four- and six-cutter restriction enzymes and fragments were separated electrophoretically.

Sequences were obtained using standard cycle sequencing protocols for the Applied Biosystems (Foster City, CA) model 377 sequencer; with the Big Dye sequencing kit, using M13 universal primers for clones; or by direct sequencing using primers specific to the original amplified product. All *Aly3*, *Aly9*, *Aly10.1*, and *Aly14* were sequenced directly. Sequences of the more variable loci were sequenced from cloned PCR products. In most cases, at least two clones were sequenced per individual to check for PCR errors. However, since this was not always done, some diversity values may be slightly overestimated, and an excess of singletons may have been produced; this does not affect our general conclusions. Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under POPSET accession nos. AY186752–AY186777. The full population sequence set can be obtained from the authors by request.

Sequence alignments and analyses: Sequences were aligned using ClustalX (THOMPSON *et al.* 1994), followed by manual adjustments based on the inferred amino acid alignments using Se-Al v. 1.0 (RAMBAUT 1996). *A. lyrata* S-domain sequences were aligned to two representative *B. oleracea* *SRK* alleles from class I (pollen dominant: *SRK9* and *SRK45*) and

class II (pollen recessive: *SRK5* and *SRK15*) nucleotide sequences and to several *A. thaliana* S-domain genes (see Table 2 for accession numbers and chromosomal locations). The nucleotide sequence of the apparent *A. thaliana* ortholog of *Aly13* (KUSABA *et al.* 2001) was also included (T6K22.100: accession no. AL031187). Conserved amino acid residues common to S-domain loci (KUSABA *et al.* 1997) were used to anchor the alignments. For visual comparison of relative divergence within and between sequence types, phylogenetic trees were reconstructed with PAUP* 4.0b10 (SWOFFORD 2002), using the minimum evolution function under an HKY85 model of substitution (HASEGAWA *et al.* 1985). Heuristic searches were conducted with initial trees obtained by simple stepwise addition, followed by branch swapping using the tree bisection-reconnection routine implemented in PAUP*. Relative support for individual nodes was assessed by bootstrapping (1000 replicates) using neighbor joining (SAITOU and NEI 1987). For these analyses, sequences from *Aly14* and their putative ortholog *At14* (see discussion below) were excluded because of the short length of the *Aly14* sequences. For the other sequence types, all unique sequences were included in the analysis, except for *Aly13*, for which a subset of 10 alleles with known linkage relationships (SCHIERUP *et al.* 2001) was included. The analysis was based on the region from bp 496 to 1365 of the alignment.

To estimate nucleotide divergence between sequences, synonymous (K_s) and nonsynonymous substitutions per site (K_a) were calculated using the method of NEI and GOJOBORI (1986) with the MEGA2 program (KUMAR *et al.* 2000). For divergence between pairs of paralogous loci within species or for divergence between species, all regions present in both sequences were included. The comparisons therefore involve slightly different lengths of sequence (see Figure 2 and Table 2).

Nonsynonymous and synonymous diversity values (π_a and π_s) within species were estimated using a set of putative alleles of the different loci sequenced from a common small sample of individuals from the four populations. (Some individuals did not yield sequences for some loci, sometimes because the DNA sample was used up, so we were unable to obtain exactly the same samples for all loci; Table 5 shows the sample sizes.) The MEGA2 software was used for diversity estimation. Proseq v. 2.9 (FILATOV and CHARLESWORTH 1999) was used to do TAJIMA'S D (1989a) tests and to test for population subdivision in *A. lyrata* using the K_t statistic (HUDSON *et al.* 1992). These analyses used all sequences available from the study populations for each locus; *i.e.*, sequences were included that were excluded by the sampling scheme used to estimate within-population diversity. The critical values for K_t were obtained by 1000 random permutations of the sequences between the populations (HUDSON *et al.* 1992).

Recombination in the *Aly* loci was tested by two types of analysis. Correlation analyses were done with the r2 program written by M. H. Schierup (<http://www.brics.dk/~compbio/r2>). Only segregating sites with frequencies >0.1 were included. Where only one or two sequences had gaps, the site was included (although the sequences with gaps were excluded from the analysis); otherwise, gap regions were excluded from all the sequences. Significance of the correlation coefficients of two measures of linkage disequilibrium with distance (r^2 or D') was determined using 5000 random permutations of the variable sites. The second analysis used the composite likelihood finite sites extension to HUDSON'S (2000) method (MCVEAN *et al.* 2002), implemented in the LDhat program (<http://www.stats.ox.ac.uk/~mcvean>). All sites were included in this analysis. Maximum-likelihood estimates of the scaled recombination rate were obtained from the likelihood curve evaluated from 20 points between $\rho = 0$ and $\rho = 50$. Significance against $\rho = 0$ was tested by 1000 random permuta-

tions. Minimum numbers of recombination events were estimated by HUDSON and KAPLAN'S (1985) estimator using DNAsp v. 3.5 (ROZAS and ROZAS 1999).

To test whether the parts of the S-domain sequences that are hypervariable in the *Aly13* sequences also evolve unusually in the other *A. lyrata* S-domain genes, we compared the levels of divergence and polymorphism in different regions of *Aly13* with diversity in the sequences of each putative locus. Following KUSABA *et al.* (1997), the positions of the HV regions in the S-domains of the Brassica *SLG* and *SRK* type I alleles initially recognized by DWYER *et al.* (1994) correspond to the following positions in our alignments: HV1 corresponds to amino acid residues 215–236; HV2 to 293–331; HV3 to 354–368; and the C-terminal region to residues 446–456. McDonald-Kreitman tests (MCDONALD and KREITMAN 1991) were used to test for selection acting on the *A. lyrata* S-domain loci using DnaSP v. 3.5.

To test whether more changes have occurred in any of the *A. lyrata* *Aly* genes, compared with their orthologs in *A. thaliana* genes since their divergence from outgroup sequences, relative rates were evaluated using Tajima's one-parameter test (TAJIMA 1993), which does not assume any specific substitution model. As no sequence data are from orthologs of closely related species, the closest paralogous *A. thaliana* S-domain gene was used as an outgroup for each comparison (for *Aly8/Ark3*, this was *Ats1*; for *Aly9/Ats1*, this was *Ark3*; for *Aly10.1/Ark1*, this was *Ark2*; and for *Aly10.2/Ark2*, this was *Ark1*).

Tests for linkage between the *A. lyrata* self-incompatibility locus and the S-domain loci: To test *Aly* S-domain loci for linkage to the self-incompatibility locus, we used full-sib families in which the incompatibility groups of progeny plants had been determined by hand-pollinations between individuals or in which *Aly13* genotypes had been determined, so that it was known that one or both parent plants were *Aly13* heterozygotes (see SCHIERUP *et al.* 2001). The two parents and the progeny were also scored using PCR amplification with primers specific for any of the other *Aly* sequence types that sequencing showed to be heterozygous in one or the other parent. We used digestion with restriction enzymes to score variants for these putative loci; the details are described below.

RESULTS

Amplification of S-domain sequences: Several combinations of primers designed to match conserved regions within the first exon (the S-domain) of the Brassica S-gene family (see MATERIALS AND METHODS) were used in the initial screening of *A. lyrata* genomic DNA for S-domain sequences. Amplifications from a single individual from Scotland yielded PCR products of the size predicted from Brassica S-gene sequences. Five different sequence types were initially identified using the six-cutter restriction enzymes *EcoRI*, *HindIII*, and *BamHI*, and these were sequenced (*Aly7*, *Aly8*, *Aly9*, *Aly10*, and *Aly13*; CHARLESWORTH *et al.* 2000; SCHIERUP *et al.* 2001).

BLAST searches of these sequences showed homology to Brassica and *A. thaliana* S-domain loci (see below), and they could all readily be aligned with Brassica S-allele sequences and with members of the S-domain gene family in *A. thaliana*. The portions of the S-domains sequenced (see Figure 2) in all the *Aly* loci identified, with the exception of *Aly7* and some of the *Aly10.1* alleles (see below), are open reading frames. There are

TABLE 1
Primers used

Locus	Primer name	Description	Primer sequence
S-domain primers			
<i>Aly7</i>	7–for	Used with primer 7R. Specific for the 7– haplotype (the last position of the primer spans the single base pair present in 7+ but absent in 7–)	5′ taagtccatgggacggcgtag 3′
<i>Aly7</i>	7+rev	7+ haplotype specific; within 9-bp insertion in the 7+ haplotype	5′ aagctgttgtagcggatccca 3′
<i>Aly8</i>	Ap8f	With SLGR, also amplifies <i>Aly10.1</i> , -10.2	5′ acgggaactctcaaaatattcg 3′
<i>Aly10.1</i>	10F1	With SLGR, also amplifies <i>Aly10.2</i>	5′ tccatcaagcggcgatttctcga 3′
Kinase domain primers			
	srk4R	Exon 3, amplifies 10.1 only	
	srk5R	Exon 3, amplifies <i>Aly10.1</i> only	5′ acagcttctaactctatcaatgga 3′
	srknasr1	Second base pair of exon 4, amplifies only <i>Aly10.1</i> , some <i>Aly13</i> plus unidentified sequences	5′ tctgcccgcagtaaccttc 3′
	srknasr3	Base pair 157 of exon 5, amplifies some <i>Aly13</i> sequences	5′ ccatcccgaatccgatattct 3′
	srknasr4	Last base pair of intron 3 (2 bp before start of srknasr1), amplifies <i>Aly10.2</i> and <i>Aly8</i> sequences	5′ tgcccgcagtaaccttccc 3′

Other primers mentioned in the text are described in CHARLESWORTH *et al.* (2000).

no stop codons in any of the S-domains, and all indels, including the few that are polymorphic within loci (see Figure 2), are multiples of three nucleotides. No introns were found in any of the *A. lyrata* S-domain sequences.

Design of specific primers for S-domain sequence types and evidence that they represent different loci: Given the very high diversity of the *A. lyrata Aly13* sequences, which we ascribe to a single incompatibility locus, as explained above, it is important to investigate in detail the extent and nature of the S-domain gene family to check that the *Aly13* sequences truly come from a single locus. We therefore used sequence variants to help classify the S-domain sequences into sets belonging to different loci.

Specific primer pairs were designed for individual sequences (CHARLESWORTH *et al.* 2000; Table 1). All individuals tested by amplification of genomic DNA, representing all four populations sampled, have all of the different sequence types. These sequences therefore suggest at least five different loci. A further locus, *Aly10.2*, was subsequently found in amplifications using primers initially designed from *Aly10.1* sequences. Amplifications using primers designed to be specific for *Aly13* (13seq1F and SLGR; see SCHIERUP *et al.* 2001) yielded another sequence type, *Aly14* (accession no. AY186762; as only short sequences were obtained from a few plants, this sequence type is not included in most of the analyses below). This yields a total of seven loci, in addition to the *Aly13* locus. Although there is some diversity within most of the sequence types, which we describe in detail below, the different sequences are diverged at both silent and replacement sites (Table 2) and fall into strongly supported groups in a phylogenetic analysis (Figure 1), consistent with the other evidence that they represent distinct loci. The five *A. thali-*

ana gene sequences in Figure 1 and Table 2 will be discussed later. Unlike these putative S-domain loci, no single primer amplifies all sequences of the *Aly13* type (MABLE *et al.* 2003). The *Aly13* sequences are highly variable, consistent with other data suggesting that these sequences represent a single highly polymorphic *A. lyrata* S-locus, rather than several different loci (SCHIERUP *et al.* 2001).

A. thaliana orthologs of the Aly genes and structure of the loci: Comparing the *Aly* sequences with *A. thaliana* S-domain receptor kinase genes (Table 2), we can identify probable orthologs for five genes (Figure 1). For three loci (*Aly3*, *Aly7*, or *Aly9*), we could not identify kinase domains by PCR (see MATERIALS AND METHODS); for brevity, we refer to these as “nonkinase domain” sequences, although it is possible that a kinase domain exists but was not detected. No orthologs can be identified for two of these loci, *Aly3* and *Aly7*. For *Aly9*, *AtS1* is a potential ortholog (see Table 2 for accession numbers). This is the probable ortholog of *SLRI* in Brassica (Dwyer *et al.* 1994) and, like *Aly9*, appears not to have a kinase domain.

We tentatively identify orthologs of *Aly8*, *Aly10.1*, and *Aly10.2* as the three kinase domain loci *Ark3*, *Ark1*, and *Ark2*, respectively (Table 2). These three *A. lyrata* genes have quite similar S-domains, which amplify with the same forward primers (see Table 1), and kinase domains were detectable for all three. A possible ortholog of *Aly14* (which we have not tested for the presence of a kinase domain) is an anonymous kinase domain sequence, which we denote by *At14* (contig accession AL161566.2|ATCHRIV66, position 170101).

For the putative orthologous pairs, silent site divergence from *A. thaliana* for the S-domain ranges from 18 to 30%, and replacement site divergence is between

TABLE 2

Synonymous (above diagonal) and nonsynonymous divergence values (below diagonal) between *A. lyrata* S-domain *Aly* sequences and those of potentially orthologous *A. thaliana* S-domain kinases

Locus	Accession nos.	Length of sequence ^a	<i>Aly3</i>	<i>Aly7-</i>	<i>Aly8</i>	<i>Aly9</i>	<i>Aly10.1</i>	<i>Aly10.2</i>	<i>Ark1</i>	<i>Ark2</i>	<i>Ark3</i>	<i>Ats1</i>
<i>Aly3</i>	AY186752	701		0.945	1.383	0.735	1.380	1.456	1.380	1.097	0.936	0.637
<i>Aly7-</i>	AY186753-4	862	0.494		>2	1.469	>2	>2	>2	>2	1.955	1.721
<i>Aly8</i>	AY186755	772	0.304	0.480		0.849	0.354	0.380	0.648	0.511	<u>0.305</u>	0.884
<i>Aly9</i>	AY186756	918	0.445	0.501	0.499		0.721	0.684	0.994	0.791	0.865	<u>0.176</u>
<i>Aly10.1</i>	AY186757-60	938	0.392	0.521	0.187	0.535		0.159	<u>0.311</u>	0.275	0.482	0.877
<i>Aly10.2</i>	AY186761	713	0.416	0.586	0.188	0.538	0.124		0.403	<u>0.192</u>	0.452	1.020
<i>Ark1</i> (chromosome 1)	M80238		0.367	0.462	0.226	0.405	<u>0.067</u>	0.159		0.293	0.527	0.961
<i>Ark2</i> (chromosome 1)	AY045777		0.394	0.517	0.221	0.414	<u>0.132</u>	<u>0.049</u>	0.112		0.495	0.822
<i>Ark3</i> (chromosome 4)	AL031187		0.283	0.396	<u>0.032</u>	0.391	0.182	0.213	0.178	0.172		0.867
<i>Ats1</i> (chromosome 3)	S84921		0.419	0.470	0.390	<u>0.119</u>	0.502	0.487	0.423	0.416	0.379	

The divergence values involving *A. lyrata* sequences are means using all sequences obtained from each of the *Aly* loci and were calculated by the method of NEI and GOJOBORI (1986) with Jukes-Cantor correction. Divergence values between the putative *A. lyrata* genes and their *A. thaliana* orthologs are underlined.

^a The column of sequence lengths gives the longest sequence available for each locus. For comparisons with the *A. thaliana* sequences, only the portions of sequences available from both species were used, so that the lengths are not given separately for these loci.

3 and 12% (see Table 2). Divergence values in the three kinase domain sequence types were similar (based on the coding sequence of exons 1–7, synonymous and nonsynonymous divergence values between *Aly8* and *Ark3* were 0.34 and 0.049, respectively; for *Aly10.1* *vs.* *Ark1*, the values were 0.22 and 0.07, and for *Aly10.2* *vs.* *Ark2*, 0.19 and 0.065). The values for both the S- and kinase domains are higher than values for most orthologous sequence comparisons between these two species (WRIGHT *et al.* 2002), but within reasonable limits (silent divergence exceeding ~30% after Jukes-Cantor correction is unlikely for true orthologs). The high divergence for *Aly8* is in part due to its diversity within *A. lyrata* (discussed in more detail below).

Figure 2 summarizes the structure of the S-domain sequence types identified in *A. lyrata* and their putative *A. thaliana* orthologs within the regions that were sequenced for the eight putative *A. lyrata* loci. The inferred amino acid sequences all share the 12 cysteine residues present in the Brassica *SLG*, *SRK*, and *SLR* S-domain sequences of other Brassicaceae (KUSABA *et al.* 1997), as well as the *A. thaliana* *Ark* loci. We searched our sequences for the motif of 10 amino acids described in the S-domains of other plant receptor-like protein kinases (WALKER 1994; WQSFYPTDT in all Brassica *SRK* sequences). The motif is present in *Aly3* and *Aly9* and in several *Aly13* sequences (13-2, -3, -4, -7, -8, -14, and -20). The *Aly8*, *Aly10.1*, and *Aly10.2* sequences differ in the 6th amino acid of the motif (F replaces Y), and the same is true for *Aly13* sequences 13-1, -5, -9, -13, -15, and -23 (Figure 2).

***Aly7* sequences:** As mentioned previously, some *Aly7* sequences have a single base-pair insertion (at position

751 within a region of four TA repeats; see Figure 3, “*Aly7*(+)” sequences). This disrupts the reading frame and creates a downstream stop codon. A 9-bp insertion is also present in this set of sequences, relative to the in-frame, *Aly7*(–) sequences (Figure 3). Overall, 26% out of a total of 27 *Aly7* sequences classified (either by sequencing or by amplifying with primers specific to each haplotype; see Table 1) were of the 7(+) type. These sequences could represent a separate locus or could be allelic to the other *Aly7* sequences. The sequences with and without the insertion form two haplotypes. There is significant linkage disequilibrium between the two types, between sites separated by 550 nucleotides, and between closer sites (Figure 3). This might suggest two distinct loci, but there are few pairs of sites for which linkage disequilibrium is complete, so the sequences appear to have recombined. Alternatively, interlocus gene conversion may have occurred, so this does not conclusively rule out the possibility of two loci. Nucleotide similarity is otherwise high between the two sequence types. Removing the insertion from the *Aly7*(+) sequences, the mean divergence from the other *Aly7* sequences is 0.017 for synonymous sites, and slightly higher (0.019) for nonsynonymous sites, suggesting that the sequences are evolving neutrally. However, net divergence is very small, given the diversity within the *Aly7*(–) sequences. The sequence diversity of the *Aly7*(+) sequences is lower than that of *Aly7*(–), and only slightly lower than the divergence between the two, and again the sequences appear to be evolving neutrally (among *Aly7*(+) sequences, $\pi_s = 0.0098$ and $\pi_a = 0.0107$). Tajima’s *D* is significantly negative ($D = -1.65$, $P < 0.05$) for the *Aly7*(+) sequences, but is not

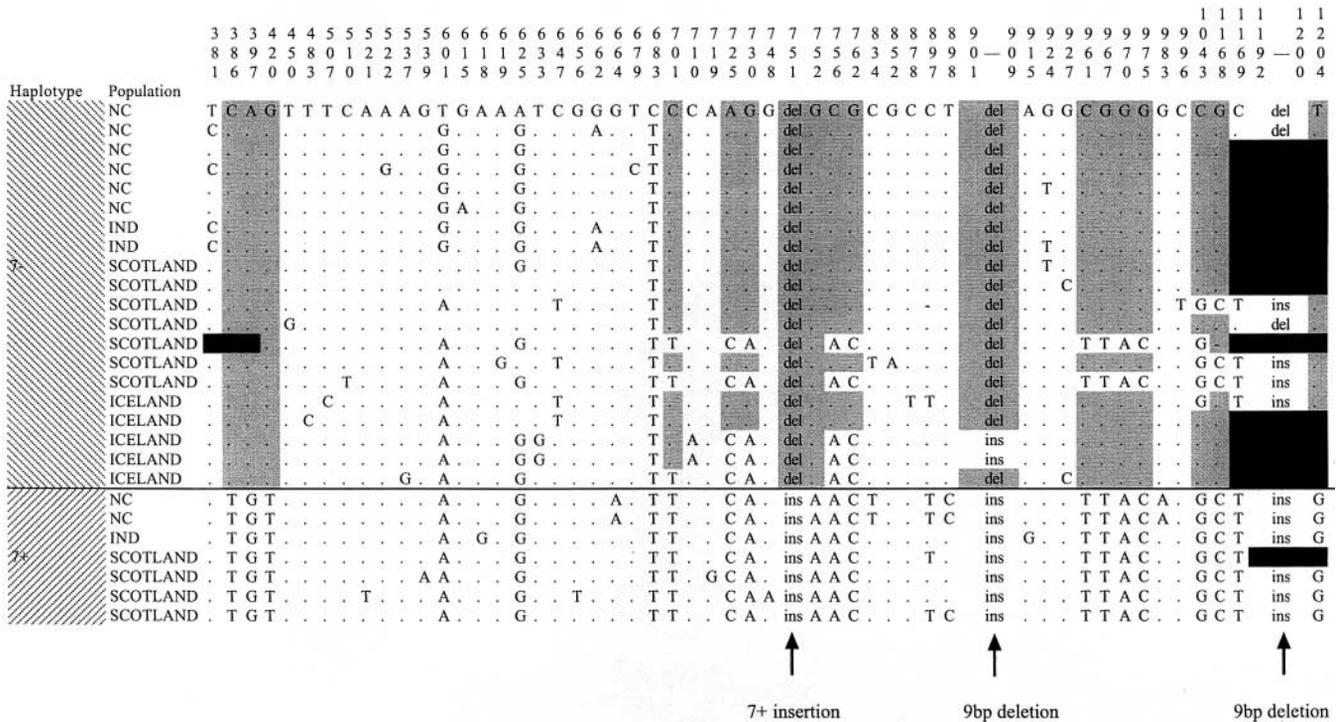


FIGURE 3.—Sequence variants of the *Aly7* locus. The numbers of the polymorphic sites are indicated at the top, and the rest of the figure shows the variants present in the sequences of the two haplotypes from the different populations. Sites in linkage disequilibrium are shaded, and regions of missing sequence are blacked out.

Tests for linkage between the *A. lyrata* self-incompatibility locus and the *S*-domain loci: We tested for linkage of the *Aly S*-domain loci and the self-incompatibility locus, using families whose parents were heterozygous for one or more *Aly* loci (see MATERIALS AND METHODS). Linkage between *Aly13* variants and the *S*-locus in both sibships, and in several other families, has already been reported (CHARLESWORTH *et al.* 2000; SCHIERUP *et al.* 2001). Since variants that do not segregate as alleles of course define different loci, we hoped that tests for linkage would help to show whether or not sequence variants are from a single locus, in addition to what can be deduced from similarities and differences between the sequences.

For the sibship 98E-15 (see CHARLESWORTH *et al.* 2000; SCHIERUP *et al.* 2001), direct sequencing reactions for *Aly3*, *Aly8*, *Aly10.1*, *Aly10.2*, and *Aly9* from the parents of this family showed that the parent 97F13-5 was heterozygous for a number of sites in two loci (Table 3). Two restriction enzymes, *Acl*I and *Bpu*AI, were used to test for segregation of *Aly3* alleles. Sequence variants at *Aly3* behave as allelic, and this locus frequently recombined with the *S*-locus. In the same family, a length polymorphism at the end of the *S*-domain of the *Aly10.1* locus segregates as expected if the variants are allelic (Table 3) and also recombine with the *S*-locus. The results for another sibship, 98G-23, confirmed this conclusion for *Aly10.1*.

For *Aly8*, however, some variants showed linkage. Ta-

ble 4 shows another sibship in which both parents were double heterozygotes and in which the *Aly8* variants again cosegregated with *Aly13* sequences. Linkage of variants was detected in several other sibships by scoring *Aly13* variants known from our previous work to show linkage to the *S*-locus (*Aly13*-4, -5, -9, -13, -16, and -22 from several different natural populations). These results for *Aly8* are consistent with the fact that *Ark3*, its *A. thaliana* ortholog, is linked to the putative *SRK* ortholog of this species, which is a pseudogene (KUSABA *et al.* 2001). However, other very similar sequences, in addition to two different linked variants, were sometimes amplified with primers for *Aly8*, suggesting that *Aly8* represents two separate loci. The linkage of the second locus with the *S*-locus is not known, and it is even possible that haplotypes vary in the numbers of *Aly8* genes linked to *Aly13*, similarly to the variable number of linked *SCR* genes in *A. lyrata* (KUSABA *et al.* 2001). Until physical maps of different haplotypes are available, this cannot be resolved.

To examine further the possibility of paralogous loci, we aligned all our *Aly8* sequences to test whether they cluster into two sets with fixed differences between them. However, we found no evidence for any such haplotype structure in the complete sequence data set. Moreover, we could not identify variants that characterize the set of linked or the set of recombining sequences from several families. In other words, there are no sites in linkage disequilibrium that allow us to define site

TABLE 3

Results of linkage tests between *Aly* loci and the self-incompatibility locus in family 98E-15

Plant	SI group	Genotypes at the loci tested		
		<i>Aly13</i>	<i>Aly3^a</i>	<i>Aly10.1</i>
Parent 97F-13/5	Not tested	1/13	aa	A
Parent 97F 15-3	Not tested	3/23	Aa	A/B
1	I	1/23	Aa	A
2	IV	3/13	aa	A
3	I	1/23	aa	A
4	II	13/23	aa	A
5	II	13/23	Aa	A
6	II	13/23	aa	A
7	III	1/3	Aa	A/B
8	III	1/3	aa	A/B
9	IV	3/13	aa	A
10	I	1/23?	Aa	A/B
11	IV	3/13	aa	A
12	II	13/23	aa	A/B
13	III	1/3	Aa	A/B
14	IV	3/13	Aa	A
15	IV	3/13	Aa	A

Sequences of *Aly3*, *Aly8*, and *Aly10.1* were amplified from genomic DNA of the parental and progeny plants, using primers specific for these sequence types (*Aly3* primers 3F + 3R and *Aly10.1* primers Ap8F + srk5R, respectively).

^aThe restriction enzyme *Bpu*AI was used to distinguish the "A" and the "a" sequence types.

states characteristic of the two putative loci and that might allow us to distinguish the loci on the basis of their sequences. This is also clear in Figure 1, in which there is no evident split into two *Aly8* types.

Finally, there is linkage disequilibrium between the *Aly8* and the *Aly13* loci. We studied a set of *Aly8* sequences that cosegregate with various incompatibility alleles of independent origins (scored using restriction enzyme digestion of *Aly13* PCR products to determine the *Aly13* sequence types). Four *Aly13* types were represented twice in the sample. There were a total of 38 nucleotide sites polymorphic among the *Aly8* sequences in this sample. Between pairs of *Aly8* sequences from the four pairs of haplotypes whose *Aly13* sequences match, the mean proportion of these variable sites that differ was 7% (range 0–17%). Between *Aly8*s from haplotypes with different *Aly13*s, the differences were much greater (the mean proportion of difference in the 21 comparisons is 41%, and the range is 20–71%). The similarities between the most similar *Aly8* sequences are underestimated, because some of the differences may be PCR errors (these sequences are from cloned PCR products, as this highly polymorphic locus cannot be directly sequenced).

Within- and between-population variability of the different *Aly* loci: *Within-population and total diversity:* We estimated sequence variability of seven S-domain *Aly* loci, at least two of which are not closely linked to the

TABLE 4

Results of linkage tests between *Aly8* variants and the self-incompatibility locus in family MS00-36

<i>Aly13</i> genotype	Nos. of plants with <i>Aly8</i> band			
	<i>Alu-1</i> , <i>Aly13-11</i>	<i>Alu-2</i> , <i>Aly13-9</i>	<i>Alu-3</i> , <i>Aly13-4</i>	<i>Msp-1</i> , <i>Aly13-16</i>
4/9	0	4	4	0
4/11	6	0	6	0
9/16	0	3	0	3
11/16	1	0	0	1

Segregation of the *Aly8* variants was scored with *Alu* and *Msp* digestion, and each column of the table is for one of the four alleles present in the parents (given at the top of the four columns in the table). The *Aly13* alleles in the maternal (Icelandic) and paternal (Scottish) parents were, respectively, 9/11 and 4/16, and each column also specifies which allele cosegregates with each of the *Aly8* alleles in this family.

Slocus (see above). Table 5 summarizes the results for each of the four populations studied, as well as for the total sample. The mean within-population synonymous site diversity values (π_s) are mostly <1%. π_a/π_s values, based on the within-population diversity values (Table 5), are mostly rather high for the loci studied here and for the *Aly13* locus (although the very high *Aly9* value is based on few variable sites). The extremely high diversity for *Aly8* (species wide $\pi_s = 7.5\%$ and $\pi_a > 1\%$) is partly, but not entirely, attributable to the fact that this sequence type could represent more than one locus, as just explained. This is discussed further below.

Between-population diversity: When all sites were used in the analysis, tests for spatial structure (HUDSON *et al.* 1992) detected significant population subdivision in *A. lyrata* for several loci. The exceptions are *Aly10.1* (with low diversity), the highly polymorphic *Aly13* locus, and *Aly8* (Table 5). *Aly8* shows some evidence of population structure for silent sites (a moderate K_s value, significant at the 5% level), but this is difficult to interpret, given our inability to assign different sequences to individual loci. Overall, the data therefore indicate isolation between the geographically distant populations studied. Clearly, however, the high variability at the *Aly8* locus is not merely a consequence of population subdivision. This sequence type is highly variable within all four populations studied (see above).

Recombination and linkage disequilibrium: In an effort to test whether the putative alleles of each of the loci identified from the sequences are truly allelic, we tested for recombination in S-domain loci other than *Aly13*, where polymorphism levels are high and balancing selection is likely, which violates the assumptions of the analysis. Except for the *Aly9* locus, which has little variability, both tests used suggest recombination (or some other form of exchange) in all the putative loci (Table 6). In the *Aly8* sequences, many exchange events are detected using HUDSON and KAPLAN's (1985) esti-

TABLE 5
Nucleotide diversity estimates and summary statistics for seven *S*-domain loci sampled from four *A. lyrata* populations

Locus	Diversity measure	Population				Global (π_T)	Mean within-population (π_s)	K_{st}		Tajima's <i>D</i>	π_a/π_s
		North Carolina	Indiana	Scotland	Iceland			All sites	Silent sites		
<i>Aly3</i>	π_s	0.0040	0	0.0033	0.0080	0.0053	0.0033	0.60 ^a	0.78 ^a	026	0.825
	π_a	0.0052	0	0.0023	0.0022	0.0044					
	<i>n</i>	8	6	4	3	21					
<i>Aly7</i>	π_s	0.0182	0	0.0138	0.0171	0.0212	0.0146	0.13 ^a	0.13 ^a	-1.52	0.712
	π_a	0.0074	0.0049	0.0110	0.0127	0.0151					
	<i>n</i>	6	2	7	5	20					
<i>Aly8</i>	π_s	0.0821	0.0812	0.0800	0.0619	0.0795	0.0750	0.02	0.10 ^a	-0.9	0.153
	π_a	0.0132	0.0117	0.0127	0.0094	0.0122					
	<i>n</i>	3	8	7	8	26					
<i>Aly9</i>	π_s	0.0201	0.0025	0	0	0.0054	0.0058	0.13 ^a	0.26	-2.34	0.889
	π_a	0.0071	0.0006	0.0032	0	0.0048					
	<i>n</i>	5	6	5	4	20					
<i>Aly10.1</i>	π_s	0	0.0021	0.0109	0.003	0.0138	0.0048	0.08	0.02	-1.43	0.159
	π_a	0.0011	0.0016	0.0085	0.0078	0.0022					
	<i>n</i>	7	7	10	5	29					
<i>Aly10.2</i>	π_s	0	0.0220	0.0243	0.0127	0.0227	0.0163	0.08 ^a	0.374	1.76	0.374
	π_a	0.0090	0.0090	0.0092	0.0052	0.0085					
	<i>n</i>	4	4	7	4	19					
<i>Aly13</i>	π_s	0.2915	0.3486	0.4467	0.3144	0.3612	0.3720	0.10	0.554	-0.19	0.554
	π_a	0.1722	0.2005	0.2306	0.1695	0.2000	0.2022				
	<i>n</i>	6	10	15	7	38					

Sample sizes are given in the rows labeled *n*. Note that the *Aly7* alleles with the inserted nucleotide are excluded from these estimates.

^a Significant population structure $P < 0.05$.

mator, even though these sequences probably come from at least two loci (see above). This suggests the possibility of gene conversion between the different loci.

Tests for selection within *A. lyrata*: *Tajima's tests*: Tajima's *D* statistic (Tajima 1989a) was calculated for each putative locus to check whether the sequences appear to be evolving neutrally. Since the sample sizes are too small to perform the test within populations, we pooled the sequences from the different populations. For most loci, the Tajima's *D* values were negative, but did not differ significantly from zero, although a significant neg-

ative value ($P < 0.01$) was found for *Aly10.1*. Only *Aly3* gave a positive Tajima's *D* value, but this is nonsignificant and may be attributable to population subdivision, consistent with the high K_{st} for this locus. There is thus no evidence from this test for balancing selection acting at any of the loci. This includes the *Aly8* sequences and the highly polymorphic *Aly13* locus, which is likely to be the *A. lyrata* self-incompatibility locus.

Patterns of evolution in the S-domain sequences: Replacement site polymorphism in the putative *A. lyrata* self-incompatibility locus, *Aly13*, is significantly higher in

TABLE 6
Tests for recombination in the *S*-domains of the *Aly* loci

Gene	No. of sequences	Alignment length	Correlation analysis			LDhat analysis	
			Sites included	r^2	<i>D'</i>	Sites included	ρ
<i>Aly3</i>	21	597	11	-0.42*	-0.60*	13	10.5***
<i>Aly7</i>	20	768	10	-0.35**	-0.40**	35	27.8***
<i>Aly8</i>	28	535	25	-0.45***	-0.42***	73	23.8***
<i>Aly9</i>	20	827	7	0.02	-0.18	68	0
<i>Aly10.1</i>	20	713	13	-0.30**	0.06	23	13.1
<i>Aly10.2</i>	23	859	11	-0.41**	-0.13	46	5.26***

The estimates are based on pooling sequences from all populations. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

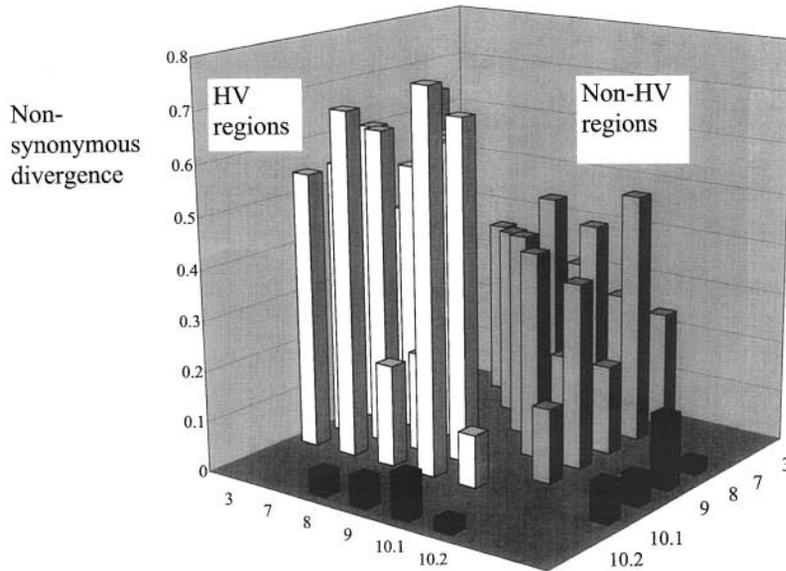


FIGURE 4.—Comparisons of divergence in HV *vs.* other regions of the *S*-domain sequences. The figure shows nonsynonymous site divergence values between the different paralogous *Aly* *S*-domain sequences in *A. lyrata*, and divergence of these sequences from the putative orthologous loci in *A. thaliana* (shown as solid). The total lengths of sequences compared can be seen in Table 2, and the HV regions are as described in MATERIALS AND METHODS and include \sim 200 bp.

the regions corresponding to the Brassica *SRK* and *SLG* hypervariable regions than in the rest of the sequence (SCHIERUP *et al.* 2001). This is not found for the further loci with *S*-domain sequences studied here. In these, including the *Aly8* sequences, variability is not especially high in these regions.

However, the low polymorphism at most of these loci (see above) makes it difficult to detect differences in diversity among different sequence regions. We therefore also estimated nonsynonymous divergence among the *A. lyrata* paralogs and among the three loci with putative orthologs in *A. thaliana* (*Aly8*, -10.1, and -10.2; see Table 2). Divergence for regions corresponding to the Brassica *SLG* and *SRK* and *Aly13* HV regions was compared with divergence elsewhere in the *S*-domain sequences (see MATERIALS AND METHODS for the positions assigned to these regions). Among *A. lyrata* paralogs, the regions that correspond to non-HV regions of the *S*-domain accumulate fewer substitutions per nonsynonymous site than do those corresponding to HV regions (the mean for HV is 60% higher than that for non-HV regions); synonymous divergence is saturated and comparisons are not informative for such sites. Of the 15 comparisons, 14 show HV nonsynonymous divergence greater than non-HV divergence (Figure 4, open and shaded bars, respectively; the difference is significant with $P < 0.0005$ by a paired sign test, although it must be realized that the tests are not independent). Thus either these regions are under lower selective constraint than the remainder of the sequence or directional selection has caused divergence in these regions, specifically, in the different loci. There is no such clear effect between the *Aly* loci and their *A. thaliana* orthologs (Figure 4, solid bars).

MCDONALD-KREITMAN (1991) tests did not detect evidence of directional selection driving divergence between paralogous loci specifically in the HV regions

(although this test indicated a significant excess of nonsynonymous polymorphic sites in the non-HV region of *Aly3*, when compared with divergence from several of the other loci; the reason for this result is unknown, but there is no evidence for balancing selection as this locus does not have high diversity). The relative rate tests also give no indication of any overall deviation from equal rates of evolution of the orthologous pairs of genes since divergence (data not shown). Overall, we conclude that the fairly high K_a/K_s values in the *S*-domains and the high nonsynonymous divergence in HV regions are due largely to low selective constraints, rather than to diversifying selection.

DISCUSSION

The *S*-domain gene family: The *S*-domain loci studied here clearly form part of an ancient gene family with members in all other angiosperms tested, including distantly related species such as maize (WALKER 1994; ANSALDI *et al.* 2000). The loci show different degrees of divergence, but most differ considerably. An exception is the two or more genes apparently required to account for the *Aly8* results, which could be a recent duplication, but further work is needed to clarify the situation with respect to these sequences. Five of the eight *A. lyrata* loci identified have plausible orthologs in *A. thaliana*, which is highly self-compatible, and not surprisingly, therefore, its *Aly13* ortholog is a pseudogene (KUSABA *et al.* 2001). Several of the genes are expressed both in nonflower tissues and in flowers in *A. thaliana* (we have no data on the expression of these genes in *A. lyrata*). *Ark1* is expressed in leaves, flower buds, and stigmas (TOBIAS *et al.* 1992; DWYER *et al.* 1994; TOBIAS and NASRALLAH 1996; SUZUKI *et al.* 1997, 1999); *Ark2* is expressed specifically during maturation of cotyledons, leaves, and sepals; and, on the basis of promoter

expression, *Ark3* is detected specifically in roots and in the root-hypocotyl transition zone (Dwyer *et al.* 1994). The nonkinase *AtS1* is expressed specifically in stigmatic papillary cells (ISOGAI *et al.* 1988; LALONDE *et al.* 1989).

There is no evidence for major birth and death of members of this gene family between *A. lyrata* and *A. thaliana*, since most loci can be identified in both species. However, duplication of the pollen-expressed *SCR* gene was found in one of the two haplotypes studied by KUSABA *et al.* (2001), and different numbers of copies of *SRK* are known in Brassica haplotypes (CABRILLAC *et al.* 1999). There may also be further members of this family that we have not studied. PCR amplifications occasionally yielded further S-domain sequences that either were very different from those described here or proved not to be closely linked to the S-locus (M. H. SCHIERUP, unpublished results). Furthermore, two of the *Aly* loci without detectable kinase domains (*Aly3* and *Aly7*) have no evident *A. thaliana* orthologs, and, as discussed in the next section, two loci, *Aly7* and *Aly10.1*, may be pseudogenes.

Pseudogenes: Pseudogene S-domain genes have been found in the Brassica S-locus region (SUZUKI *et al.* 1999), and this possibility must be considered for two of our putative loci. The *Aly7* sequences with the inserted nucleotide might suggest that these sequences merely represent a pseudogene, but our segregation evidence suggests that they are allelic (presumably null alleles) at the same locus as the *Aly7(-)* sequences. Moreover, both types of sequence appear to be quite old, since there are numerous fixed differences between them, and both include considerable diversity and are found in most, if not all, *A. lyrata* populations. This would argue against this locus being a pseudogene; *7(+)* appears to have more singletons than *7(-)*, consistent with its being a derived sequence type. The significantly negative Tajima's *D* for the null alleles [*7(+)* sequences] suggests a recent increase in frequency of this allele and is conservative, given the other evidence for population subdivision. However, there is more diversity among *7(+)* alleles than would be expected if this haplotype rose to high frequency by a recent selective sweep, which would imply low or zero diversity (HUDSON *et al.* 1994).

The *Aly10.1* sequences containing deletions may also be a pseudogene. Our diversity analysis included the different types of alleles of this locus, excluding the deletion regions, and nonsynonymous diversity was low, as was the π_a/π_s ratio (see Table 5), suggesting that loss of function occurred recently. It therefore seems most likely that the B1, B2, and B3 alleles (see Figure 2) are null alleles.

Other examples of polymorphic null alleles are known, sometimes at frequencies as high as those found for the *Aly7* and *Aly10.1* sequences (*e.g.*, OXTOBY *et al.* 1991; GIBSON *et al.* 1992; CHARMLEY *et al.* 1993; MOMBAERTS 2001). One example is the deletion of a disease

resistance gene in *A. thaliana*, for which there is evidence for balancing selection (STAHL *et al.* 1999), but there is no evidence for this at the *Aly7* or *Aly10.1* loci. Replacement polymorphism in the *Aly7* sequences without the insertion is only slightly lower than that at silent sites. This high π_a/π_s ratio might suggest a locus that has lost function or is in an early stage of doing so and is evolving neutrally; but this is not certain, since similar or even higher π_a/π_s values are found for other *Aly* loci (see Table 5). Loss of function could also explain the higher diversity in *Aly7* than in most of the other loci. We are unable to compare divergence of the nonfunctional and potentially functional *Aly7* alleles, since no *A. thaliana* ortholog can be identified. The absence of an ortholog is, however, consistent with this gene being a nonfunctional duplicate in *A. lyrata*.

Levels and patterns of diversity: The S-domain loci studied here have a range of nucleotide diversity values, including widely differing silent site diversity. Lack of evidence for balancing selection and only moderate diversity levels are also reported for the Brassica *SLR1*, *SLA*, and *SLB* loci, which are not linked to the incompatibility locus (although sample sizes are very small; HINATA *et al.* 1995; SAKAMOTO *et al.* 1998; WATANABE *et al.* 1998; LUU *et al.* 2001). For *A. lyrata*, the diversity estimate for 1.6 kb of the alcohol dehydrogenase gene (*Adh*) yielded a mean within-population π -value for all nucleotide sites of 0.1%; the total diversity, including three different populations, was 0.38% (SAVOLAINEN *et al.* 2000). Higher values are found at other loci, particularly in the non-U.S. subspecies *petraea* (WRIGHT *et al.* 2002). All except one of our S-domain loci also have diversity values higher than those of the *Adh*, but there is no reason to suspect balancing selection. Ratios of nonsynonymous-to-synonymous site diversity within species or divergence between species for plant nuclear genes generally range from 0.1 to 0.2 (LI 1997; LIU 1998; WRIGHT *et al.* 2002). Published values for Brassica S-domain genes are, however, much higher (HINATA *et al.* 1995), and the same is true in our data (see Table 5; for divergence between paralogs within *A. lyrata*, the mean value is 0.55 for the sequence regions corresponding to the *SRK* HV regions and 0.24 for the rest of the S-domain sequence). This suggests that selective constraints are low in the S-domain, especially in the parts that are hypervariable in *SRK*, which would accord with the ability of the S-domain to generate highly diverse *SRK* alleles.

The difference between the reference loci studied here and the otherwise similar S-domain *Aly13* sequences (the putative *A. lyrata* S-locus) therefore supports the view that the *Aly13* silent and amino acid diversity is unusually high due to the maintenance of the polymorphism of incompatibility alleles. Moreover, variation at the *Aly13* locus is similar in all populations, and K_{st} does not differ significantly from zero. This is as expected for loci experiencing balancing selection

(SCHIERUP *et al.* 2000) and is similar to what has been found in the fungus *Schizophyllum commune*, where no population structure was detectable for mating-type alleles (RAPER *et al.* 1958) in contrast to strong structure for polymorphic allozymes (JAMES *et al.* 1999). The same is true for *Aly8*, although analysis of the synonymous diversity suggests some differentiation between populations. In contrast, several of the other *Aly* loci are significantly differentiated among the populations studied here, indicating some degree of isolation between these geographically distant populations; the same is observed for several other non-S-domain loci (WRIGHT *et al.* 2002). Since subdivision obscures evidence of selection when samples are pooled (SCHIERUP *et al.* 2000), the absence of significant Tajima's test results for the *Aly13* sequences (see above) is not inconsistent with balancing selection. The main evidence for balancing selection on *Aly13* remains its extraordinarily high nonsynonymous diversity.

The lower diversity of the loci studied here, compared with that of *Aly13*, probably cannot be attributed to their location in regions of low recombination, which is a possible reason for low diversity (*e.g.*, BEGUN and AQUADRO 1992; STEPHAN and LANGLEY 1998). Recombination was detected in all loci, even using sequences pooled for all populations, which must lead to underestimation of its frequency, given the fact that there is evidence for subdivision, so that some sequences have no opportunity to recombine. However, the possibility of gene conversion means that recombination may be overestimated using sequence variants (LANGLEY *et al.* 2000; PRZEWORSKI and WALL 2001). Even if this were the only form of exchange, hitchhiking processes would still act to reduce diversity, as is observed in other genome regions where reciprocal recombination is rare (*e.g.*, LANGLEY *et al.* 2000).

In comparisons among paralogs within *A. lyrata*, the regions corresponding to the Brassica S-locus hypervariable regions are more diverged than the rest of the S-domain. This supports the suggestions made above that these regions have lower selective constraints than the rest of the S-domain protein. There is no tendency for any of the *Aly* loci studied here to have especially high variability in these regions, in contrast with the results for *Aly13*, in which variability is higher in the sites corresponding to Brassica HV regions than elsewhere in the S-domain sequence. The fact that the putative *A. lyrata* *SRK* gene, but not the other S-domain genes in the species, has excess diversity in the HV regions, is consistent with suggestions that these regions may be involved in recognition functions. However, it is quite plausible that elevated variability in these regions is due to linkage to sites under balancing selection elsewhere in the *SRK* gene, with low selective constraints in these parts of the sequence allowing higher amino acid diversity than elsewhere in the protein. With current evidence, it is difficult to exclude this possibility, and in

genes under balancing selection, high amino acid variability should therefore not be considered sufficient evidence for the sites of the selection. More direct analyses, including transgenic experiments, will probably be necessary to determine the recognition regions within the *SRK* sequence.

Diversity of the *Aly8* locus, a gene linked to the S-locus:

The high silent and replacement diversity at the *Aly8* locus can be accounted for in part if these sequences come from two loci, at least one of which we have shown to be linked to the S-locus. The situation could be similar to that in Brassica, in which some haplotypes have further S-domain loci in the S-locus region (YU *et al.* 1996; SUZUKI *et al.* 1997, 1999; KAI *et al.* 2001).

The existence of two paralogs is not the sole reason for the high *Aly8* diversity, however, because the variants are not found in different haplotypes that could be assigned to different loci, but appear instead to be from a recombining set of sequences. One possibility that we cannot at present exclude is that, in some haplotypes, but not others, the *Aly8* locus recombines with the S-locus. In the two *A. lyrata* haplotypes for which physical maps exist for the S-locus region, the ortholog of the *Ark3* gene is >5 kb away from the *SRK* gene in one, but at least 30 kb in the other haplotype (KUSABA *et al.* 2001); these distances, however, would not imply frequent recombination. Moreover, three different *Aly8* sequences have been found in PCR reactions of some individuals, which suggests two separate loci in some plants at least. It is unlikely that variability at the *Aly8* locus is maintained by balancing selection, and it is not concentrated in the regions that show hypervariability in the Brassica and *A. lyrata* *SRK* loci.

Much of the *Aly8* diversity is probably attributable to linkage to the highly polymorphic *Aly13* locus, which could cause elevated variability at synonymous and even nonsynonymous sites, similarly to the effects discussed above for the different regions of the S-domain. *Aly8* is orthologous to the *Ark3* gene, which is closely linked to the *A. thaliana* *SRK* pseudogene ortholog (KUSABA *et al.* 2001). In Brassica, low diversity has been suggested for the SLL1 and SLL2 non-S-domain loci that are linked to *SRK*. However, sample sizes were very small, and diversity was scored by a restriction fragment length polymorphism approach that would not detect all sequence variants, so the estimate of variability is not accurate (YU *et al.* 1996). It will be interesting to get more data on diversity at loci linked to the S-loci in both Brassica and *A. lyrata*.

Our data show linkage disequilibrium between *Aly8* sequences and S-alleles, with independent individuals whose S-alleles are the same having very similar *Aly8* sequences. This would be expected if the high *Aly8* diversity were due to close linkage to a gene under balancing selection. There is also evidence for some exchange (either recombination or gene conversion)

between this locus and other *Aly8* sequences, making it impossible to distinguish the locus of origin of our sequences. Given that *Aly8* has a kinase domain, it may be helpful to study diversity in this region of the gene to search for sequences that distinguish the two or more different loci. This might allow tests for selection *vs.* neutrality of the *Aly8* variants and should help to disclose the extent of the region of linkage disequilibrium in the *S*-locus region. Further work is clearly needed to elucidate the number of loci contributing to the diversity of sequences that we classify as *Aly8* and to study their diversity individually, as well as that at other loci linked to the *S*-locus. If extensive linkage disequilibrium is confirmed, it would support the view that recombination does not occur in this region of the genome and suggest that the exchanges that can be detected (AWADALLA and CHARLESWORTH 1999; SCHIERUP *et al.* 2001) may be due to gene conversion. If this occurs, it may have important implications for the maintenance of diversity among *S*-alleles, since sequences from other *S*-domain genes may sometimes be introduced into alleles at the *S*-locus.

Conclusions: The chief conclusion from these studies is that high silent and replacement diversity in several sequences and regions may be attributable to linkage to a locus under balancing selection. This locus is presumably *Aly13*, the *A. lyrata* putative incompatibility locus. We have found high variability at the *Aly8* locus, which cannot entirely be accounted for in terms of paralogous loci and which displays linkage disequilibrium with *Aly13* alleles, just as predicted for regions close to a locus with a balanced polymorphism (see above). The *Aly13* sequences have elevated nonsynonymous diversity in the sites corresponding to the hypervariable regions of Brassica *SRK* *S*-domains, but our evidence that these regions diverge more than the rest of the *S*-domain suggests that this is largely because of low selective constraints of these portions of the proteins. The hypervariability cannot therefore be taken as evidence that sites within these regions themselves experience balancing selection. Among different regions linked to sites subject to balancing selection, those regions that are not important for the *S*-domain's signal transmission function are expected to have the greatest nonsynonymous divergence and polymorphism. They may thus be the regions most likely to evolve sequence differences that can be used in recognition of different specificities. This may explain the observation of hypervariability in similar regions in Brassica and *A. lyrata* *SRK* and in Brassica *SRK* and *SLG*. An implication is that it may thus be difficult to determine which amino acid sites determine the functional incompatibility types (and are therefore under balancing selection). One possible approach is to do experimental functional tests involving altering individual amino acid residues in *S*-allele sequences. Studies of *S*-allele sequence variability may also

help to indicate which amino acids can differ between alleles with the same incompatibility type.

We thank the staff at the University of Edinburgh for growing the plants and T. E. Thorhallsdottir, C. H. Langley, and R. Mauricio for the seeds used in this work. This work was supported by the Biotechnology and Biological Sciences Research Council of the UK. B. K. Mable was also supported by the Natural Sciences and Engineering Council of Canada, D. Charlesworth by the Natural Environment Research Council of Great Britain and Edinburgh University, M. H. Schierup by the Danish Natural Sciences Research Council (grant nos. 9701412 and 1262), and P. Awadalla by an Edinburgh University Faculty of Science and Engineering Scholarship.

LITERATURE CITED

- ANSALDI, R., A. CHABOUD and C. DUMAS, 2000 Multiple I gene family members including natural antisense transcripts are differentially expressed during development of maize flowers. *J. Biol. Chem.* **275**: 24146–24155.
- AWADALLA, P., and D. CHARLESWORTH, 1999 Recombination and selection at Brassica self-incompatibility loci. *Genetics* **152**: 413–425.
- BEGUN, D. J., and C. F. AQUADRO, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**: 519–520.
- BOYES, D. C., and J. B. NASRALLAH, 1993 Physical linkage of the *SLG* and *SRK* genes at the self-incompatibility locus of *Brassica oleracea*. *Mol. Gen. Genet.* **236**: 369–373.
- CABRILLAC, D., V. DELORME, J. GARIN, V. RUFFIO-CHABLE, J.-L. GIRANTON *et al.*, 1999 The *S*₁₅ self-incompatibility haplotype in *Brassica oleracea* includes three *S* gene family members expressed in stigmas. *Plant Cell* **11**: 971–985.
- CHARLESWORTH, B., M. NORDBORG and D. CHARLESWORTH, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided inbreeding and outcrossing populations. *Genet. Res.* **70**: 155–174.
- CHARLESWORTH, D., and P. AWADALLA, 1998 The molecular population genetics of flowering plant self-incompatibility polymorphisms. *Heredity* **81**: 1–9.
- CHARLESWORTH, D., P. AWADALLA, B. K. MABLE and M. H. SCHIERUP, 2000 Population-level studies of multiallelic self-incompatibility loci, with particular reference to Brassicaceae. *Ann. Bot.* **85** (Suppl. A): 227–239.
- CHARMLEY, P., K. WANG, L. HOOD and D. A. NICKERSON, 1993 Identification and physical mapping of a polymorphic human T-cell receptor V-beta gene with a frequent null allele. *J. Exp. Med.* **177**: 135–143.
- CONNER, J. A., P. CONNER, M. E. NASRALLAH and J. B. NASRALLAH, 1998 Comparative mapping of the Brassica *S* region and its homeolog in *Arabidopsis*: implications for the evolution of mating systems in the Brassicaceae. *Plant Cell* **10**: 801–812.
- CUI, Y., N. BRUGIÈRE, M. ARNOLDO and S. J. ROTHSTEIN, 2000 The *S* locus glycoprotein and the *S* receptor kinase are sufficient for self-pollen rejection in Brassica. *Proc. Natl. Acad. Sci. USA* **97**: 3713–3717.
- DOUGHTY, J., S. DIXON, S. J. HISCOCK, A. C. WILLIS, I. A. P. PARKIN *et al.*, 1998 PCP-A1, a defensin-like Brassica pollen coat protein that binds the *S* locus glycoprotein, is the product of gametophytic gene expression. *Plant Cell* **10**: 1333–1347.
- DWYER, K. G., M. A. BALENT, J. B. NASRALLAH and M. E. NASRALLAH, 1991 DNA sequences of self-incompatibility genes from *Brassica campestris* and *B. oleracea*: polymorphism predating speciation. *Plant Mol. Biol.* **16**: 481–486.
- DWYER, K. G., M. K. KANDASAMY, D. I. MAHOSKY, J. AXIAI, B. I. KUDISH *et al.*, 1994 A superfamily of *S* locus-related sequences in *Arabidopsis*: diverse structures and expression patterns. *Plant Cell* **6**: 1829–1843.
- FILATOV, D. A., and D. CHARLESWORTH, 1999 DNA polymorphism, haplotype structure and balancing selection in the *Leavenworthia PgiC* locus. *Genetics* **153**: 1423–1434.
- GIBSON, J. B., A. V. WILKS and A. AGROTIS, 1992 Molecular relationships between alcohol-dehydrogenase null-activity alleles from

- natural-populations of *Drosophila melanogaster*. *Mol. Biol. Evol.* **9**: 250–260.
- GORING, D. F., and S. J. ROTHSTEIN, 1996 S-locus receptor kinase genes and self-incompatibility in *Brassica napus*, pp. 217–230 in *Signal Transduction in Plant Growth and Development*, edited by D. P. S. VERMA. Springer-Verlag, New York.
- HASEGAWA, M., K. KISHINO and T. YANO, 1985 Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**: 160–174.
- HATAKEYAMA, K., T. TAKASAKA, M. WATANABE and K. HINATA, 1998 Molecular characterization of S locus genes, SLG and SRK, in a pollen-recessive self-incompatibility haplotype of *Brassica rapa* L. *Genetics* **149**: 1587–1597.
- HINATA, K., M. WATANABE, S. YAMAKAWA, Y. SATTA and A. ISOGAI, 1995 Evolutionary aspects of the S-related genes of the Brassica self-incompatibility system: synonymous and nonsynonymous base substitutions. *Genetics* **140**: 1099–1104.
- HUDSON, R. R., 2000 A new statistic for detecting genetic differentiation. *Genetics* **155**: 2011–2014.
- HUDSON, R. R., and N. L. KAPLAN, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**: 147–164.
- HUDSON, R. R., D. D. BOOS and N. L. KAPLAN, 1992 A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **9**: 138–151.
- HUDSON, R. R., K. BAILEY, D. SKERECKY, J. KWIATOWSKI and F. J. AYALA, 1994 Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. *Genetics* **136**: 1329–1340.
- HUGHES, A., T. OTA and M. NEI, 1990 Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol. Biol. Evol.* **7**: 515–524.
- ISOGAI, A., S. TAKAYAMA, H. SHIOZAWA, C. TSUKAMOTO, T. KANBARA *et al.*, 1988 Existence of a common glycoprotein homologous to S-glycoproteins in 2 self-incompatible homozygotes of *Brassica campestris*. *Plant Cell Physiol.* **29**: 1331–1336.
- JAMES, T. Y., D. PORTER, J. L. HAMRICK and R. VILGALYS, 1999 Evidence for limited intercontinental gene flow in the cosmopolitan mushroom *Schizophyllum commune*. *Evolution* **53**: 1665–1677.
- JUNGHANS, H., and M. METZLAFF, 1990 A simple and rapid method for the preparation of total plant DNA. *Biotechniques* **8**: 176.
- KAI, N., G. SUZUKI, M. WATANABE, A. ISOGAI and K. HINATA, 2001 Sequence comparisons among dispersed members of the Brassica S multigene family in an S-9 genome. *Mol. Genet. Genomics* **265**: 526–534.
- KIMURA, R., K. SATO, R. FUJIMOTO and T. NISHIO, 2002 Recognition specificity of self-incompatibility maintained after the divergence of *Brassica oleracea* and *Brassica rapa*. *Plant J.* **29**: 215–223.
- KOCH, M., B. HAUBOLD and T. MITCHELL-OLDS, 2000 Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis* and related genera (Brassicaceae). *Mol. Biol. Evol.* **17**: 1483–1498.
- KUMAR, S., K. TAMURA, I. JACOBSEN and M. NEI, 2000 MEGA2: Molecular Evolutionary Genetics Analysis, Version 2.0. Penn State University/Arizona State University, University Park, PA/Tempe, AZ.
- KUSABA, M., T. NISHIO, Y. SATTA, K. HINATA and D. OCKENDON, 1997 Striking similarity in inter- and intra-specific comparisons of class I SLG alleles from *Brassica oleracea* and *Brassica campestris*: implications for the evolution and recognition mechanism. *Proc. Natl. Acad. Sci. USA* **94**: 7673–7678.
- KUSABA, M., M. MATSUSHITA, K. OKAZAKI, Y. SATTA and T. NISHIO, 2000 Sequence and structural diversity of the S-locus genes from different lines with the same self-recognition specificities in *Brassica oleracea*. *Genetics* **154**: 413–420.
- KUSABA, M., K. DWYER, J. HENDERSHOT, J. VREBALOV, J. B. NASRALLAH *et al.*, 2001 Self-incompatibility in the genus *Arabidopsis*: characterization of the S-locus in the outcrossing *A. lyrata* and its autogamous relative, *A. thaliana*. *Plant Cell* **13**: 627–643.
- LALONDE, B. A., M. E. NASRALLAH, K. G. DWYER, C. H. CHEN, B. BARLOW *et al.*, 1989 A highly conserved Brassica gene with homology to the S-locus-specific glycoprotein structural gene. *Plant Cell* **1**: 249–258.
- LANGLEY, C. H., B. P. LAZZARO, W. PHILLIPS, E. HEIKKINEN and J. M. BRAVERMAN, 2000 Linkage disequilibrium and the site frequency spectra in the *su(s)* and *su(w)* regions of the *Drosophila melanogaster* X chromosome. *Genetics* **156**: 1837–1852.
- LI, W.-H., 1997 *Principles of Molecular Evolution*. Sinauer Associates, Sunderland, MA.
- LIU, F., 1998 Genetic diversity in *Leavenworthia* populations with different inbreeding levels: the effect of breeding system on the level and pattern of molecular variation in plant populations. Ph.D. Thesis, University of Chicago, Chicago.
- LUU, D.-T., P. HEIZMANN, C. DUMAS, M. TRICK and M. CAPPADOCIA, 1997 Involvement of SLR genes in pollen adhesion to the stigmatic surface in Brassicaceae. *Sex. Plant Reprod.* **10**: 227–235.
- LUU, D. T., S. HUGUES, E. PASSELÈGUE and P. HEIZMANN, 2001 Evidence for orthologous S-locus related I genes in several genera of Brassicaceae. *Mol. Gen. Genet.* **264**: 735–745.
- MABLE, B. K., M. H. SCHIERUP and D. CHARLESWORTH, 2003 Estimating the number of S-alleles in a natural population of *Arabidopsis lyrata* (Brassicaceae) with sporophytic control of self-incompatibility. *Heredity* **90**: 422–431.
- MCDONALD, J. H., and M. KREITMAN, 1991 Accelerated protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.
- MCVEAN, G. A. T., P. AWADALLA and P. FEARNHEAD, 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.
- MIEGE, C., V. RUFFIO-CHABLE, M. H. SCHIERUP, D. CABRILLAC, T. GAUDE *et al.*, 2001 Intra-haplotype polymorphism at the Brassica S locus. *Genetics* **159**: 811–822.
- MOMBAERTS, P., 2001 The human repertoire of odorant receptor genes and pseudogenes. *Annu. Rev. Genomics Hum. Genet.* **2**: 493–510.
- NASRALLAH, J. B., 1997 Evolution of the Brassica self-incompatibility locus: a look into S-locus gene polymorphisms. *Proc. Natl. Acad. Sci. USA* **94**: 9516–9519.
- NASRALLAH, M. E., and J. B. NASRALLAH, 1989 The molecular genetics of self-incompatibility in Brassica. *Annu. Rev. Genet.* **23**: 121–139.
- NASRALLAH, M. E., and J. B. NASRALLAH, 1993 Pollen stigma signaling in the sporophytic self-incompatibility response. *Plant Cell* **5**: 1325–1335.
- NEI, M., and T. GOJOBORI, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418–426.
- NISHIO, T., M. KUSABA, K. SAKAMOTO and D. OCKENDON, 1997 Polymorphism of the kinase domain of the S-locus receptor kinase gene (*SRK*) in *Brassica oleracea* L. *Theor. Appl. Genet.* **95**: 335–342.
- NORDBORG, M., B. CHARLESWORTH and D. CHARLESWORTH, 1996 Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. *Proc. R. Soc. Lond. Ser. B* **263**: 1033–1039.
- OXTOBY, E., M. A. DUNN, A. PANCORO and M. A. HUGHES, 1991 Nucleotide and derived amino-acid-sequence of the cyanogenic beta-glucosidase (linamarase) from white clover (*Trifolium repens* L.). *Plant Mol. Biol.* **17**: 209–219.
- PASTUGLIA, M., R. SWARUP, A. ROCHER, P. SAINDRANAN, D. ROBY *et al.*, 2002 Comparison of the expression patterns of two small gene families of S gene family receptor kinase genes during the defence response in *Brassica oleracea* and *Arabidopsis thaliana*. *Gene* **282**: 215–225.
- PRZEWORSKI, M., and J. D. WALL, 2001 A genome-wide departure from the standard neutral model in natural populations of *Drosophila*. *Genetics* **156**: 257–268.
- RAMBAUT, A., 1996 Se-Al: Sequence Alignment Editor, Version 1 (<http://evolve.zps.ox.ac.uk>).
- RAPER, J. R., G. S. KRONGELB and M. G. BAXTER, 1958 The number and distribution of incompatibility factors in *Schizopyllum*. *Am. Nat.* **92**: 221–232.
- ROLLINS, R. C., 1993 *The Cruciferae*. Stanford University Press, Stanford, CA.
- ROZAS, J., and R. ROZAS, 1999 DnaSP version 3.0: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- SAKAMOTO, K., M. KUSABA and T. NISHIO, 1998 Polymorphism of the S-locus glycoprotein gene (*SLG*) and the S-locus related gene (*SLRI*) in *Raphanus sativus* L. and self-incompatible ornamental plants in the Brassicaceae. *Mol. Gen. Genet.* **258**: 397–403.
- SAVOLAINEN, O., C. H. LANGLEY, B. P. LAZZARO and H. FREVILLE,

- 2000 Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. *Mol. Biol. Evol.* **17**: 645–655.
- SCHIERUP, M. H., X. VEKEMANS and D. CHARLESWORTH, 2000 The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genet. Res.* **76**: 51–62.
- SCHIERUP, M. H., B. K. MABLE, P. AWADALLA and D. CHARLESWORTH, 2001 Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*. *Genetics* **158**: 387–399.
- SCHOPFFER, C. R., M. E. NASRALLAH and J. B. NASRALLAH, 1999 The male determinant of self-incompatibility in Brassica. *Science* **286**: 1697–1700.
- SHIU, S.-H., and A. B. BLEECKER, 2001 Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. USA* **98**: 10763–10768.
- STAHL, E. A., G. DWYER, R. MAURICIO, M. KREITMAN and J. BERGELSON, 1999 Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* **400**: 667–671.
- STEPHAN, W., and C. H. LANGLEY, 1998 DNA polymorphism in *Lycopodium* and crossing-over per physical length. *Genetics* **150**: 1585–1593.
- STEPHENSON, A. G., J. DOUGHTY, S. DIXON, C. ELLRMAN, S. HISCOCK *et al.*, 1997 The male determinant of self-incompatibility in Brassica is located in the pollen coating. *Plant J.* **12**: 1351–1359.
- STROBECK, C., 1972 Heterozygosity in pin-thrum plants or with partial sex linkage. *Genetics* **72**: 667–678.
- SUZUKI, G., M. WATANABE, N. KAI, N. MATSUDA, K. TORIYAMA *et al.*, 1997 Three members of the S multigene family are linked to the S locus of Brassica. *Mol. Gen. Genet.* **256**: 257–264.
- SUZUKI, G., N. KAI, T. HIROSE, K. FUKUI, T. NISHIO *et al.*, 1999 Genomic organization of the S locus: identification and characterization of genes in SLG/SRK region of S9 haplotype of *Brassica campestris* (syn. *rapa*). *Genetics* **153**: 391–400.
- SWOFFORD, D. L., 2002 PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. Sinauer Associates, Sunderland, MA.
- TAJIMA, F., 1989a Statistical method for testing the neutral mutation hypothesis. *Genetics* **123**: 585–595.
- TAJIMA, F., 1989b The effect of change in population size on DNA polymorphism. *Genetics* **123**: 597–601.
- TAJIMA, F., 1993 Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**: 599–607.
- TAKAHATA, N., 1990 A simple genealogical structure of strongly balanced allelic lines and trans-species polymorphism. *Proc. Natl. Acad. Sci. USA* **87**: 2419–2423.
- TAKASAKI, T., K. HATAKEYAMA, G. SUZUKI, M. WATANABE, A. ISOGAI *et al.*, 2000 The S receptor kinase determines self-incompatibility in Brassica stigma. *Nature* **403**: 913–916.
- TAKAYAMA, S., H. SHIBA, M. IWANO, H. SHIMOSATO, F.-S. CHE *et al.*, 2000 The pollen determinant of self-incompatibility in *Brassica campestris*. *Proc. Natl. Acad. Sci. USA* **97**: 1920–1925.
- TANTIKANJANA, T., M. E. NASRALLAH, J. C. STEIN, C.-H. CHEN and J. B. NASRALLAH, 1993 An alternative transcript of the S locus glycoprotein gene in a class II pollen-recessive self-incompatibility haplotype of *Brassica oleracea* encodes a membrane-anchored protein. *Plant Cell* **5**: 657–666.
- THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 1994 Clustal-W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- TOBIAS, C. M., and J. B. NASRALLAH, 1996 An S-locus-related gene in *Arabidopsis* encodes a functional kinase and produces two classes of transcripts. *Plant J.* **10**: 523–531.
- TOBIAS, C. M., B. HOWLETT and J. B. NASRALLAH, 1992 An *Arabidopsis thaliana* gene with sequence similarity to the S-locus receptor kinase of *Brassica oleracea*—sequence and expression. *Plant Physiol.* **99**: 284–290.
- VANOOSTHUYSE, V., C. MIEGE, C. DUMAS and J. M. COCK, 2001 Two large *Arabidopsis thaliana* gene families are homologous to the Brassica gene superfamily that encodes pollen coat proteins and the male component of the self-incompatibility response. *Plant Mol. Biol.* **46**: 17–34.
- VEKEMANS, X., and M. SLATKIN, 1994 Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics* **137**: 1157–1165.
- WALKER, J. C., 1994 Structure and function of the receptor-like protein-kinases of higher-plants. *Plant Mol. Biol.* **26**: 1599–1609.
- WATANABE, M., T. ONO, K. HATAKEYAMA, S. TAKAYAMA and K. HINATA, 1997 Molecular characterization of SLG and the S-related genes in a self-compatible *Brassica campestris* L. var. *yellow sarson*. *Sex. Plant Reprod.* **10**: 332–340.
- WATANABE, M., M. WATANABE, G. SUZUKI, H. SHIBA, S. TAYAMA *et al.*, 1998 Sequence comparison of four SLR1 alleles in *Brassica campestris* (syn. *B. rapa*). *Sex. Plant Reprod.* **11**: 295–296.
- WATANABE, M., A. ITO, Y. TAKADA, C. NINOMIYA, T. KAKIZAKI *et al.*, 2000 Highly divergent sequences of the pollen self-incompatibility (S) gene in class-I haplotypes of *Brassica campestris* (syn. *rapa*) L. *FEBS Lett.* **473**: 139–144.
- WRIGHT, S., 1939 The distribution of self-sterility alleles in populations. *Genetics* **24**: 538–552.
- WRIGHT, S. I., B. LAUGA and D. CHARLESWORTH, 2002 Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Mol. Biol. Evol.* **19**: 1407–1420.
- YU, K., T. L. GLAVIN, D. F. GORING and S. J. ROTHSTEIN, 1996 Molecular characterization of the S locus in two self-incompatible *Brassica napus* lines. *Plant Cell* **8**: 2369–2380.

Communicating editor: M. K. UYENOYAMA

