

# Uneven segregation of sporophytic self-incompatibility alleles in *Arabidopsis lyrata*

J. BECHSGAARD,\* T. BATAILLON† & M. H. SCHIERUP\*

\*Department of Ecology and Genetics, University of Aarhus, Aarhus, Denmark

†INRA-UMR 1097 'Diversité et Génomes des Plantes Cultivées', Domaine de Melgueil, Mauguio, France

## Keywords:

*Arabidopsis lyrata*;  
gametic selection;  
segregation distortion;  
sporophytic self-incompatibility.

## Abstract

Self-incompatibility in *Arabidopsis lyrata* is sporophytically controlled by the multi-allelic S-locus. Self-incompatibility alleles (S-alleles) are under strong negative frequency dependent selection because pollen carrying common S-alleles have fewer mating opportunities. Population genetics theory predicts that deleterious alleles can accumulate if linked to the S-locus. This was tested by studying segregation of S-alleles in 11 large full sib families in *A. lyrata*. Significant segregation distortion leading to an up to fourfold difference in transmission rates was found in six families. Differences in transmission rates were not significantly different in reciprocal crosses and the distortions observed were compatible with selection acting at the gametic stage alone. The S-allele with the largest segregation advantage is also the most recessive, and is very common in natural populations concordant with its apparent segregation advantage. These results imply that frequencies of S-alleles in populations of *A. lyrata* cannot be predicted based on simple models of frequency-dependent selection alone.

## Introduction

Self-incompatibility (SI) is widespread among hermaphroditic plants (de Nettancourt, 1977). SI prevents self-fertilization and reduces mating between closely related species (Vekemans *et al.*, 1998). Homomorphic SI systems are divided into gametophytic and sporophytic systems (de Nettancourt, 1977). An incompatible reaction occurs when the incompatibility type of the pollen matches the incompatibility type of the ovule. In gametophytic SI (GSI) systems, the incompatibility type of the pollen is determined by the Self-incompatibility allele (S-allele) of its haploid genome, whereas in sporophytic SI (SSI) systems it is determined by the diploid genotype of the paternal parent. Dominance between S-alleles exists in both pollen and stigma in SSI systems, whereas both alleles are always expressed in the stigma in GSI systems. The molecular genetics of GSI and SSI systems share many characteristics (Nasrallah, 2000). The SI locus

consists of at least one 'female' and one 'male' gene expressed respectively in the stigma and pollen tissues (the male gene has not yet been identified in GSI systems, but is expected to be present). In most SI systems, these genes are so closely linked that they behave effectively as a single Mendelian locus, also known as the S-haplotype (Takasaki *et al.*, 2000; Sato *et al.*, 2002).

Focusing on the SSI system of Brassicaceae, about which most is known, the stigma gene (SRK, Schopfer *et al.*, 1999) acts as a receptor for the pollen ligand (SCR, Takayama *et al.*, 2000; Takayama *et al.*, 2001). The sequences of SRK and SCR have been determined in more than 40 S-haplotypes in the closely related species *Brassica rapa* and *B. oleracea* (Sato *et al.*, 2002). In *Arabidopsis lyrata*, the orthologues of these two genes have also been identified (Kusaba *et al.*, 2001; Schierup *et al.*, 2001) and 23 different putative SRK-alleles have been sequenced (Schierup *et al.*, 2001; Charlesworth *et al.*, 2003). The SRK and SCR genes are sufficient to make the closely related self-compatible *Arabidopsis thaliana*, self-incompatible as shown by direct transformation (Nasrallah *et al.*, 2002). The S-haplotype appears to have reduced levels of recombination (Casselman

Correspondence: Mikkel H. Schierup, Bioinformatics Research Center, Department of Ecology and Genetics, University of Aarhus, Ny Munkegade, Building 540, Dk-8000 Aarhus C, Denmark.  
Tel.: +45 8942 3231; fax: +45 8942 3077;  
e-mail: mheide@birc.dk

*et al.*, 2000; Charlesworth *et al.*, 2003) and some haplotypes span up to 200 kb in both Brassica and *A. lyrata* (Casselman *et al.*, 2000; Kusaba *et al.*, 2001). The haplotype harbours many additional genes of known and unknown function. In Brassica this includes the SLG gene, which has been shown to enhance the specificity of the S-locus and the SI response in Brassica (Dixit *et al.*, 2000).

Assuming that the SI-system behaves as a single Mendelian locus, the population genetics of such systems are relatively simple. Models have assumed frequency-dependent selection as the only selective force affecting the dynamics of S-alleles in a population, resulting in equal frequencies of incompatibility types at equilibrium (Wright, 1939; Uyenoyama, 2000). In GSI systems this equates to equal S-allele frequencies, as there is no dominance between the S-alleles. This means that, if one samples a population and types the S-alleles, one immediately knows if the S-phenotypes are present in equal frequencies. Observations from natural populations of species possessing GSI systems show that equal S-phenotypes frequencies are often, but not always, found (see review by Lawrence, 2000).

In SSI systems, dominance relations makes it more difficult to set expectations for the frequencies of different alleles, because these will depend on the dominance relations between all pairs of alleles. A full description of the dominance between alleles does not yet exist for any species as it requires extensive and multigenerational artificial crossing. Analysis of simplified models of dominance shows that when equal frequencies of incompatibility types are assumed, recessive alleles are expected to reach higher frequencies than dominant alleles (Schierup *et al.*, 1997, 1998; Uyenoyama, 2000), but that recessive alleles have shorter persistence times in the population (Schierup *et al.*, 1997; Uyenoyama, 2000). Furthermore, genetic drift causes stochastic fluctuation of allele frequencies whose magnitude is increased by population subdivision (Schierup, 1998).

Selection occurring at sites completely linked to the S-locus may cause S-allele frequencies to deviate from those expected under equal SI phenotype proportions. One can imagine three different mutually nonexclusive types of selection: (1) recessive deleterious mutations, (2) dominant deleterious mutations and (3) segregation distorters.

Recessive deleterious mutations can be sheltered because of the enforced high heterozygosity at the SI-locus, particularly at the GSI locus where homozygotes cannot be formed. This effect will also affect heterozygosity at loci tightly linked to the S locus (Glémin *et al.*, 2001). It has been shown theoretically that such sheltering load may build up and cause substantial inbreeding depression and affect the gene genealogy of S-alleles (Uyenoyama, 2000, 2003). In sporophytic systems, less sheltering is possible, particularly for recessive alleles, as homozygotes can be formed. In the genetically simple

heterostylous systems, which may be viewed formally as a SSI system with 2 S-alleles 'pin' and 'thrum', Crosby (1949) showed indirectly in *Primula vulgaris* that homozygous homozygotes have lower fitness than homostylous heterozygotes.

Dominant deleterious mutations are not as likely to invade and fix within a SI-haplotype because they are immediately selected against within the specificity. However, mutations tightly linked to the S-locus are more likely to be fixed than variants occurring elsewhere in the genome, because of strong genetic drift within an allelic class. In GSI, drift within an allelic class can be quantified through the effective size as  $N_e/n_e$ , where  $N_e$  is the effective size of the whole population and  $n_e$  is the effective number of alleles in the population (Vekemans & Slatkin 1994). For SSI, the effective size of an S-allele depends on its dominance level, and more dominant S-alleles are expected to accumulate comparatively more linked deleterious variation.

Segregation distorters favouring the segregation of certain haplotypes are expected to invade an allelic specificity if they cause preferential transmission of some haplotype at meiosis against some or all of the remaining haplotypes. Once fixed within a given haplotype, the segregation distorter will lead to a higher equilibrium frequency of this haplotype than expected from the equal phenotype rule of SI-systems. However, negative frequency dependent selection against the frequent S-haplotype will quickly cancel out the advantage caused by the segregation distorter, thus preventing the fixation of the haplotype carrying the distorter. Although segregation distorters will often be fixed if occurring at an autosomal locus, they will remain polymorphic if closely linked to an SI-system until they can move between specificities through recombination.

Two papers have reported evidence that segregation is not always equal in homomorphic SI-systems: Lawrence & Franklin-Tong (1994) investigated the segregation ratios of the S-alleles of *Papaver rhoas*, a species possessing a GSI system, and found unequal frequencies of S-phenotypes in natural populations (Campbell & Lawrence, 1981). Equal proportions of the two alleles of the pollen or ovule parent in the progeny are, however, expected if the two parents do not share any S-alleles. Lawrence and Franklin-Tong found that segregation of the S-alleles deviated significantly from Mendelian expectations in two full-sib families, suggesting extra selection beyond the frequency-dependent selection. In *Solanum carolinense*, a species with GSI determined by different genetics, Judy Stone (personal communication) also found deviations from Mendelian segregation of the S-alleles in two out of seven crosses.

Here we present the first study of the segregation of S-alleles in a species with SSI, *A. lyrata*. The segregation of different combinations of S-alleles in 11 large full-sib families was investigated, and it was found that segregation deviates significantly from Mendelian expectations

in six of these. The strongest force contributing to these deviations seems to be some form of gametic selection. Furthermore, we found that recessive S-alleles generally show the highest transmission rates.

## Methods

### Plant material and pollination experiment

Eleven full-sib families were obtained from crossing plants originating from a single population near Reykjavik, Iceland. Reciprocal crosses were made between parental genotypes to obtain each family. The genotype of the parental genotypes as well as the dominance hierarchy between alleles used in this study are given Tables 1 and 2. Four to 15 pollinations were made in each reciprocal cross, depending on the number of mature flowers available. Pollinations were performed in a pollinator-free greenhouse to avoid contamination, and plants were spaced to avoid accidental contact of flowers between individuals. Pollinations were made by brushing the anthers gently on the surfaces of the stigmas. Seeds from each fruit were planted in different pots in a nutrition-poor soil mixed with 15–20% sand. To keep the seeds humid throughout the germination process, the pots were covered with thin perforated white plastic until the seedlings had reached a size of about 1–2 cm. Plants were grown in a greenhouse where temperature was kept constant (22 °C day, 10 °C night) under a 12 h/day light regime. When plants reached a size of 5 cm, fresh leaf material was collected

from individual plants from each reciprocal cross and stored at –70 °C.

### Genotyping of SI alleles

DNA was extracted from leaf tissue homogenized with liquid nitrogen, using a CTAB based method (see Junghans & Metzloff, 1990). The S-haplotypes of both parents and the progeny of each cross were genotyped by determining their SRK variant using the polymerase chain reaction (PCR)-based approach developed by Schierup *et al.* (2001) and Mable *et al.* (2003). PCR was performed with specific primers anchored in conserved regions of the S-domain of the SRK gene. These regions are not completely conserved among the different *A. lyrata* SRK alleles, and we used different sets of primer, recognizing different conserved regions. The forward primers 13seq1F, 13seq2F and the reverse primer SLGR were used (for PCR conditions and primer sequences, consult Schierup *et al.*, 2001). The amplified DNA from the S-locus was digested with two restriction enzymes (Alu I and Msp I) and run on a 3% agarose gel, which resulted in different Cleaved Amplified Polymorphic Segments (CAPS) patterns corresponding to different S-alleles. The identification of S-alleles from the CAPS pattern was then confirmed by using primers designed to specifically recognize the alleles identified (for specific primer sequences, see Charlesworth *et al.*, 2003). Through this method it was possible to distinguish all genotypes in nine of the 11 families (Table 1). In the two remaining families (family 21 and 32), one of the four parental S-alleles could not be amplified. In Family 21 only a single S-allele,  $S_1$ , was amplified. From one of the parents however,  $S_1$  was only found in some of the offspring of that cross, therefore the parental plant was not homozygous  $S_1S_1$ , and an unknown allele  $S_X$  must be present in the parent. However, transmission of the alleles  $S_1$  and  $S_6$  carried by the other parental plant could still be investigated in this family. In Family 32 only allele  $S_{27}$  was amplified from one of the parental plants, but not all progeny received this allele, and the parent was therefore assumed to have a further unknown allele  $S_Y$  (note that  $S_X$  and  $S_Y$  may be the same allele).

**Table 1** S-allele genotype of parents used to obtain each family.

Family	Parental genotypes
6	$S_1S_{11}^1 \times S_6S_{27}^2$
7	$S_1S_{11}^1 \times S_6S_9^3$
8	$S_1S_1 \times S_{11}S_{15}$
10	$S_1S_6 \times S_1S_{25}$
12	$S_1S_{33}^4 \times S_9S_{14}^a$
13	$S_1S_{33}^4 \times S_9S_{14}^a$
14	$S_1S_{33}^4 \times S_1S_{25}$
21	$S_1S_6 \times S_1S_X^b$
23	$S_6S_{27}^2 \times S_6S_9^3$
31	$S_1S_1 \times S_9S_{12}$
32	$S_1S_{14} \times S_{27}S_Y^b$

<sup>a</sup>Related parents (siblings).

<sup>b</sup> $S_X, S_Y$  allele not amplified by polymerase chain reaction.

\*1, \*2, \*3 and \*4 are same parents involved in different crossings.

**Table 2** Grouping of S alleles by dominance hierarchy (Mable *et al.*, 2003, L. Nielsen and M.H. Schierup, unpublished results).

Dominance classes	Alleles
Most recessive	$S_1$
Intermediate	$S_6, S_{14}, S_{25}$
Dominant	$S_9, S_{11}, S_{15}, S_{12}$

### Data analysis

Patterns of segregation of S-alleles within families were analysed using a maximum-likelihood approach and a series of nested models. These models, detailed below, were used to compare the data with segregations expected under several biological scenarios.

The structure of the data for a single family is presented as follows (see Table 3 for actual data).

The parental cross	$S_iS_j \times S_kS_l$			
The genotypic classes	$S_iS_k$	$S_iS_l$	$S_jS_k$	$S_jS_l$
Observed number of individuals	$n_1$	$n_2$	$n_3$	$n_4$

**Table 3** Observed number of individuals within each genotypic class, within each reciprocal cross. For each reciprocal cross, each genotype was used both as female (♀) or male (♂) parent.

Family	Progeny			
	S <sub>1</sub> S <sub>6</sub>	S <sub>1</sub> S <sub>27</sub>	S <sub>11</sub> S <sub>6</sub>	S <sub>11</sub> S <sub>27</sub>
6				
S <sub>1</sub> S <sub>11</sub> ♀ × S <sub>6</sub> S <sub>27</sub> ♂	14	12	1	3
S <sub>6</sub> S <sub>27</sub> ♀ × S <sub>1</sub> S <sub>11</sub> ♂	8	7	6	0
Total	22	19	7	3
	S <sub>1</sub> S <sub>6</sub>	S <sub>1</sub> S <sub>9</sub>	S <sub>6</sub> S <sub>11</sub>	S <sub>9</sub> S <sub>11</sub>
7				
S <sub>1</sub> S <sub>11</sub> ♀ × S <sub>6</sub> S <sub>9</sub> ♂	1	4	3	1
S <sub>6</sub> S <sub>9</sub> ♀ × S <sub>1</sub> S <sub>11</sub> ♂	12	8	2	5
Total	13	12	5	6
	S <sub>1</sub> S <sub>11</sub>	S <sub>1</sub> S <sub>15</sub>		
8				
S <sub>1</sub> S <sub>1</sub> ♀ × S <sub>11</sub> S <sub>15</sub> ♂	9	13		
S <sub>11</sub> S <sub>15</sub> ♀ × S <sub>1</sub> S <sub>1</sub> ♂	15	12		
Total	24	25		
	S <sub>1</sub> S <sub>1</sub>	S <sub>1</sub> S <sub>25</sub>	S <sub>6</sub> S <sub>1</sub>	S <sub>6</sub> S <sub>25</sub>
10				
S <sub>1</sub> S <sub>6</sub> ♀ × S <sub>1</sub> S <sub>25</sub> ♂	4	11	1	1
S <sub>1</sub> S <sub>25</sub> ♀ × S <sub>1</sub> S <sub>6</sub> ♂	9	2	3	1
Total	13	13	4	2
	S <sub>1</sub> S <sub>9</sub>	S <sub>1</sub> S <sub>14</sub>	S <sub>33</sub> S <sub>9</sub>	S <sub>33</sub> S <sub>14</sub>
12				
S <sub>1</sub> S <sub>33</sub> ♀ × S <sub>9</sub> S <sub>14</sub> ♂	12	7	10	4
S <sub>9</sub> S <sub>14</sub> ♀ × S <sub>1</sub> S <sub>33</sub> ♂	11	5	8	6
Total	23	12	18	10
	S <sub>1</sub> S <sub>9</sub>	S <sub>1</sub> S <sub>14</sub>	S <sub>33</sub> S <sub>9</sub>	S <sub>33</sub> S <sub>14</sub>
13				
S <sub>1</sub> S <sub>33</sub> ♀ × S <sub>9</sub> S <sub>14</sub> ♂	16	4	12	7
S <sub>9</sub> S <sub>14</sub> ♀ × S <sub>1</sub> S <sub>33</sub> ♂	4	5	10	8
Total	20	9	22	15
	S <sub>1</sub> S <sub>1</sub>	S <sub>1</sub> S <sub>25</sub>	S <sub>1</sub> S <sub>33</sub>	S <sub>25</sub> S <sub>33</sub>
14				
S <sub>1</sub> S <sub>33</sub> ♀ × S <sub>1</sub> S <sub>25</sub> ♂	6	5	6	5
S <sub>1</sub> S <sub>25</sub> ♀ × S <sub>1</sub> S <sub>33</sub> ♂	2	5	1	5
Total	8	10	7	10
	S <sub>1</sub> S <sub>1</sub> /S <sub>1</sub> S <sub>m</sub>	S <sub>1</sub> S <sub>6</sub> /S <sub>6</sub> S <sub>m</sub>		
21				
S <sub>1</sub> S <sub>6</sub> ♀ × S <sub>1</sub> S <sub>m</sub> ♂	17	7		
S <sub>1</sub> S <sub>m</sub> ♀ × S <sub>1</sub> S <sub>6</sub> ♂	27	20		
Total	44	27		
	S <sub>6</sub> S <sub>6</sub>	S <sub>6</sub> S <sub>9</sub>	S <sub>27</sub> S <sub>6</sub>	S <sub>27</sub> S <sub>9</sub>
23				
S <sub>6</sub> S <sub>27</sub> ♀ × S <sub>6</sub> S <sub>9</sub> ♂	8	12	10	6
S <sub>6</sub> S <sub>9</sub> ♀ × S <sub>6</sub> S <sub>27</sub> ♂	8	10	11	10
Total	16	22	21	16

**Table 3** Continued.

Family	Progeny			
	S <sub>1</sub> S <sub>9</sub>	S <sub>1</sub> S <sub>12</sub>		
31				
S <sub>1</sub> S <sub>1</sub> ♀ × S <sub>9</sub> S <sub>12</sub> ♂	10	11		
S <sub>9</sub> S <sub>12</sub> ♀ × S <sub>1</sub> S <sub>1</sub> ♂	9	12		
Total	19	23		
	S <sub>1</sub> S <sub>27</sub>	S <sub>1</sub> S <sub>x</sub>	S <sub>14</sub> S <sub>27</sub>	S <sub>14</sub> S <sub>x</sub>
32				
S <sub>1</sub> S <sub>14</sub> ♀ × S <sub>27</sub> S <sub>x</sub> ♂	5	4	6	7
S <sub>27</sub> S <sub>x</sub> ♀ × S <sub>1</sub> S <sub>14</sub> ♂	3	14	2	4
Total	8	18	8	11

The total number of individuals is  $N = n1 + n2 + n3 + n4$ .

Under each segregation model listed below, the proportion of individuals expected in each genotypic category is collected in a vector  $\mathbf{p} = (p_{ik}p_{ij}p_{jk}p_{ji})^T$ , where  $\mathbf{v}^T$  denotes the transpose of vector  $\mathbf{v}$ . The  $\mathbf{p}$ 's are specified by the model and must satisfy the condition  $p_{ik} + p_{ji} + p_{jk} + p_{ij} = 1$ . Given a model specifying  $\mathbf{p}$ , the likelihood of the data obtained from a cross,  $\mathbf{n} = (n1n2n3n4)^T$ , is multinomial:  $L(\mathbf{n}|M) \sim \text{Multinomial}(N, \mathbf{p})$ , where  $M$  stands for the segregation model.

All the models used in this study describe the transmission of the S-alleles from the parental plants to the progeny surviving to the seedling stage, and any deviations observed may therefore be due to selection at the gametophytic or early sporophytic stage.

A complete model ( $M_{\text{zygotic}}$ ), ignoring information brought by separating reciprocals in a cross, has three free parameters,  $p1$ ,  $p2$  and  $p3$ , and models  $\mathbf{p}$  as  $[p1, p2, p3, 1 - (p1 + p2 + p3)]^T$ . This 'saturated model' merely designed to fit the data, is not built under any explicit selection model but rather encompasses the type of selection that may happen at the gametophytic or/and zygotic stage.

A reduced model ( $M_{\text{gametic}}$ ), nested within  $M_{\text{zygotic}}$ , assumes that the expected frequencies of each diploid genotype depend only on the segregation patterns in each of the parent involved in the cross.  $M_{\text{gametic}}$  thus has two free parameters,  $\alpha$  and  $\beta$ , that determine the probability of transmission of each S allele after gametic selection as:

$$P(\text{transmission of } S_1) = T_1 = 1/2 + \alpha$$

$$P(\text{transmission of } S_2) = T_2 = 1/2 - \alpha$$

$$P(\text{transmission of } S_3) = T_3 = 1/2 + \beta$$

$$P(\text{transmission of } S_4) = T_4 = 1/2 - \beta$$

$$p_{ij} = T_i \times T_j.$$

The  $M_{\text{reciprocal}}$  model not only assumes gametic selection between the two alleles carried by each parent, but it

also incorporates information from reciprocal crosses by allowing for different selection pressures on the S-alleles in the stigma and pollen of a given parental genotype.  $M_{\text{reciprocal}}$  is formally identical to  $M_{\text{gametic}}$  and has two free parameters for each reciprocal cross, and a total of four free parameters.

The fourth model ( $M_{\text{parental}}$ ) assumes that the S-alleles in the reciprocals are exposed to equal selection pressures, which is assumed to be gametic. Therefore, the model only describes the transmission of the S-alleles from one of the parental genotypes, and has only one free parameter. The fifth model ( $M_{\text{mendelian}}$ ) assumes equal transmission of the two S-alleles from a parental plant. This model has zero free parameters.

For each family, the likelihood of the data under each model was maximized and pairs of nested models were compared. Within each family, we first tested  $M_{\text{zygotic}}$  against  $M_{\text{gametic}}$  using the likelihood ratio test ( $-2\Delta L$ ). Under the null hypothesis that  $M_{\text{zygotic}}$  provides no better fit of the data compared with  $M_{\text{gametic}}$ ,  $-2\Delta L$  is asymptotically distributed as a  $\chi^2$  distribution with 1 d.f. (see for instance Lynch & Walsh, 1998). Secondly we tested whether the transmission of the S-alleles was homogeneous in the reciprocal crosses, by comparing the likelihood of  $M_{\text{reciprocal}}$  against  $M_{\text{gametic}}$ . Under the null hypothesis, the  $-2\Delta L$  is distributed as a  $\chi^2$  with 2 d.f. Finally we tested whether the two S-alleles within each parent of a family were transmitted with equal probability (1/2) by comparing the likelihood associated with  $M_{\text{parental}}$  and  $M_{\text{mendelian}}$ . All calculations were carried out using Mathematica (version 4.1) (Wolfram Research, Champaign, IL, USA) and likelihood were maximized over the space of parameters using the function Find-Minimum.

## Results

The S-locus genotypes of the parents and the progeny in each of the 11 families are shown in Table 3. The numbers of progeny in the families ranged from 32 in Family 10 to 75 in Family 23. In total, 17 different parents and 565 progeny were genotyped at the S-locus. No progeny carried alleles not present in the parents, so we infer that contamination during pollination was virtually absent. Inspection of the number of individuals observed in each genotype class (Table 3) immediately suggests that some S-alleles have distorted segregation in several families. For example, in family 6, the genotypic classes are represented by 43, 37, 14 and 6% of the progeny, far from the expected 25% for each class.

In eight families, we could distinguish all four S-alleles in the parental genotypes, and we tested whether a saturated 'zygotic selection' model fitted the data better than a model based only on gametic selection (Table 4, 'Zygotic vs. gametic'). None of the eight families provided evidence for the presence of additional selection beyond gametic selection. If zygotic selection was the only

**Table 4** Likelihood ratio tests ( $-2\Delta L$ ) and associated  $P$ -value for the comparison of three nested models. Within a family, we first compared the fit of a saturated model with a model based only on gametic selection (zygotic- vs. gametic), secondly we tested whether gametic selection was identical in pollen and stigma (equal reciprocals?).

Family	Zygotic- vs. gametic		Equal reciprocals	
	$-2\Delta L$	$P$ -value [ $\chi^2(1)$ ]	$-2\Delta L$	$P$ -value [ $\chi^2(2)$ ]
6	0.902	0.342	3.209	0.201
7	0.131	0.717	1.196	0.550
8	–	–	1.045	0.307
10	0.555	0.456	9.799	<b>0.007</b>
12	0.014	0.906	0.191	0.909
13	0.639	0.424	4.842	0.089
14	0.038	0.845	3.490	0.175
21	–	–	1.233	0.267
23	1.616	0.204	0.675	0.713
31	–	–	0.096	0.757
32	0.613	0.434	9.110	<b>0.011</b>

Values in bold indicate significance at the 5% level.

selective force, the sample size should be sufficient to detect it. For example, with a sample size of 60, segregation of  $S_1S_3 = 11$ ,  $S_1S_4 = 19$ ,  $S_2S_3 = 19$  and  $S_2S_4 = 11$  would be significant at the 5% level with the test used.

As gametic selection seems to be acting on the S-locus, we then tested whether different selection pressures were operating in the female (stigma) and male gametes (pollen). The results of this test, performed within each family, are shown in Table 4 (see section Equal reciprocals). The test is significant for Families 10 and 32 only. This is because of a better transmission of  $S_{25}$  than  $S_1$  in the male gametes whereas the opposite pattern is observed in female gametes in Family 10, and an excess of  $S_1S_x$  genotypes in the second reciprocal in Family 32. The power of the test is limited, with a sample size of 30 in each reciprocal cross, significance at the 5% level is obtained if segregation is equal for both parents in one family and 9 vs. 21 for each of the two parents in the reciprocal. Thus, more cases of differential selection pressures in pollen vs. stigma may be detected if the sample size in a family was increased. Nevertheless, very similar segregation in female and male gametes appears to be the rule and different selection pressures in the male and female side probably do not play a large role in the patterns of segregation at the S-locus in these families.

Given the few instances of clearly different selection pressures in male and female gametes (except in families 10 and 32) we pooled the results obtained from reciprocal crosses within each family. We then tested whether the transmission of each parental pair of S-alleles to the progeny deviated significantly from a 1 : 1 expectation. The results of these tests, performed separately on each parent pair of S-alleles within each family, are shown in Table 5. In six of 19 tests significant deviations from

**Table 5** Test of the segregation of paired S alleles against the null expectation of equal probability of transmission. Bold figures indicate rejection of the null hypothesis at the 5% level. See methods for details on each model. Maximum likelihood estimates (MLE) and approximate 95% confidence interval (95% CI) for the of segregation advantage ( $\alpha$ ) of S alleles. The model assumes that when matching  $S_x$  vs.  $S_y$ ,  $S_x$  (respectively  $S_y$ ) is transmitted with probability  $1/2 + \alpha$  (respectively  $1/2 - \alpha$ ). Approximate confidence intervals were built (drop of 2 Logs in Likelihood relative to the MLE).

Family	Allele pair	$-2\Delta L$	$P$ -value [ $\chi^2(1)$ ]	Segregation advantage
6	$S_1$ vs. $S_{11}$	20.219	<b>6.90E-06</b>	0.304 [0.18, 0.40]
	$S_6$ vs. $S_{27}$	0.964	0.326	0.069 [-0.07, 0.12]
7	$S_1$ vs. $S_{11}$	5.591	<b>0.018</b>	0.194 [0.03, 0.33]
	$S_6$ vs. $S_9$	0	1	0 [-0.16, 0.16]
8	$S_{11}$ vs. $S_{15}$	0.020	0.886	-0.010 [-0.15, 0.13]
10	$S_1$ vs. $S_6$	13.476	<b>0.0002</b>	0.313 [0.19, 0.41]
	$S_1$ vs. $S_{25}$	0.125	0.724	0.031 [-0.14, 0.2]
12	$S_1$ vs. $S_{33}$	0.779	0.377	0.056 [-0.06, 0.17]
	$S_9$ vs. $S_{14}$	5.820	<b>0.016</b>	0.151 [0.02, 0.27]
13	$S_1$ vs. $S_{33}$	0.972	0.324	-0.061 [-0.17, 0.05]
	$S_9$ vs. $S_{14}$	4.972	<b>0.026</b>	0.136 [0.01, 0.25]
14	$S_1$ vs. $S_{33}$	0.029	0.866	0.014 [-0.15, 0.18]
	$S_1$ vs. $S_{25}$	0.717	0.397	-0.071 [-0.31, 0.19]
21	$S_1$ vs. $S_6$	4.110	<b>0.043</b>	0.120 [0.002, 0.23]
23	$S_6$ vs. $S_{27}$	0.013	0.908	0.007 [-0.1, 0.11]
	$S_6$ vs. $S_9$	0.013	0.908	-0.007 [-0.11, 0.1]
31	$S_9$ vs. $S_{12}$	0.382	0.537	-0.048 [-0.2, 0.1]
32	$S_1$ vs. $S_{14}$	1.093	0.276	0.078 [-0.07, 0.22]
	$S_{27}$ vs. $S_x$	3.810	0.051	-0.144 [-0.28, 0.01]

equal transmission were found at the 5% level. Note that one would expect on average about one test among the 19 performed to be significant at the 5% level just by chance and that some tests exhibiting  $P$ -values on the order of a few per cent (such as in Family 13 and 21) may very well be false positive. At any rate, bearing in mind the fact that multiple test corrections would inflate slightly the  $P$ -values we calculated here, this was a surprisingly high fraction given that transmission probabilities need to differ by at least 30% ( $\alpha = 0.3$ ,  $P < 0.05$ ) to be statistically significant with 10 individuals and 13.8% ( $\alpha = 0.138$ ,  $P < 0.05$ ) to be significant with 50 individuals with our design. It suggests that, beyond the frequency-dependent selection due to the SI system, some S-haplotypes are under strong selection.

Estimates of the transmission ratios obtained for a given pair of S-alleles (Table 5) are consistent among families (for parental origin, see Table 3).  $S_6$  and  $S_9$ , originating from same parental plant have equal segregation patterns in two families (7 and 23). The same is true for  $S_6$  and  $S_{27}$  (Family 6 and 23), and for  $S_1$  and  $S_{33}$  (Family 12, 13 and 14).  $S_1$  and  $S_{25}$  have equal segregation patterns in Families 10 and 14 where independent parental plants were used,  $S_1$  segregates consistently better than  $S_{11}$  in two families obtained from same  $S_1S_{11}$  parental plant, in Families 6 and 7,  $S_9$  segregates better than  $S_{14}$  in two families from two sister parental plants,

Family 12 and 13, and  $S_1$  segregates better than  $S_6$  in two families originating from independent parental plants, Family 10 and 21.

The strength of the selection detected depends on the S-allele combination. In some pairs the effect is very strong. The maximum likelihood estimates of the segregation advantage,  $\alpha$ , illustrate the strength of the selection detected here (Table 5). In Family 10, e.g. the transmission of  $S_1$  vs.  $S_6$  has an estimated  $\alpha$  of 0.313, which means that  $S_1$  is transmitted to 81.3% of the seedlings and  $S_6$  only to 18.7% of the seedlings. This is more than a fourfold transmission advantage. Compared with the strength of frequency-dependent selection this may be a quite extreme selective advantage. For example, in a population of 200 individuals with 20 S-alleles in equal frequency, the selective advantage of a completely new S-allele entering the population, because of frequency-dependent selection, is approximately 10% (Schierup *et al.*, 1997). Dominance and unequal S-allele frequencies may alter this prediction, but this illustrates the potential impact of such extra selection effect on the dynamics of S-alleles.

## Discussion

The crossing data suggest that segregation differences are often found when comparing different S-alleles in *A. lyrata*. This segregation distortion appears to operate mainly at the gametic level and has about the same strength in both male and female parent. This finding is surprising, given that many more genes are expressed in the male than in the female gametophyte (Honys & Twell, 2003). Furthermore, certain alleles appear to enjoy an advantage consistently in different families. The segregation distortions observed here are strong enough to affect the equilibrium frequencies of S-alleles in natural populations. An ongoing survey of S-allele frequencies in several Icelandic populations shows that frequencies of  $S_1$  range from 25 to 40%, although there are in total 17 different S-alleles (M. H. Schierup and J. Bechsgaard, unpublished results). This frequency is higher than expected under most models of dominance (Schierup *et al.*, 1997) but is consistent with the finding in this study that  $S_1$  enjoys a transmission advantage against some of the other alleles.

The absence of a detectable fitness effect acting on zygotes suggests that recessive deleterious, linked variation may play a limited role in the segregation distortion observed. Attempts were made to investigate the effect of recessive, deleterious effects using selfed progenies, where homozygote classes can be compared with the heterozygote class, but inbreeding depression was too severe to allow sufficient seed set. Much larger sample size are required, however, to rule out that zygotic selection plays a role. The present results only suggest that gametic selection alone can explain the results. This leaves one with dominant, or partially dominant linked variation or a pleiotropic effect on segregation for some of the S-locus

specificities. The present study cannot distinguish among these possibilities. The possibility of linked variation is made more probable by the fact that the S-locus resides in a region with no or very little recombination (Casselman *et al.*, 2000) and a high density of genes (Boyes *et al.*, 1997; McCubbin & Kao, 1999; Suzuki *et al.*, 1999). The question remains whether fitness differences between S-haplotypes have been created by the accumulation of deleterious variation in the haplotypes exposed at the gametic stage, leading to a segregation disadvantage, or by the introduction of segregation distorters in the S-haplotypes. Accumulation of dominant deleterious variation is only expected if the effective population size of a given specificity is very small, as mutations will be selected against within the allelic class. Under this hypothesis, we predict that the more dominant S-alleles should accumulate more dominant deleterious alleles because of greater genetic drift within their allelic class. The accumulation can continue until the segregation disadvantage offsets the negative frequency dependent selection at the S-locus. Segregation distorters, on the contrary, are always expected to invade and fix within the S-haplotype where they arise. As recessive alleles are more numerous in the population, they are more likely to be hit by a segregation distorter mutation, but we do not expect the same difference between recessive and dominant classes as for the accumulation of deleterious variation. A key to distinguishing mechanisms may therefore lie in comparing the segregation effects with the dominance level of alleles. In the present study, the most recessive allele enjoys the greatest segregation advantage, as predicted. However, in family 13, the more dominant S<sub>9</sub> allele segregates better than the less dominant S<sub>14</sub> (Table 5), so no clear effect of dominance level could be demonstrated. A more detailed determination of dominance levels together with investigation of the segregation in further families is under way to resolve this question. It would then also be possible to investigate if zygotes carrying phylogenetically more closely related alleles show a disadvantage (as they would share more deleterious variation), or whether alleles arising from a more recessive ancestral allele have less genetic load.

Finally, one should be careful in equating segregation distortion with life time fitness, as pleiotropic effects of certain S-haplotypes on fitness may occur at several life stages. The progeny used in this study were followed until flowering and measured for morphological traits (growth rate) and day to first flowering. No significant differences were detected among the four genotype classes in each family (results not shown), although inbreeding depression was found to be very high (approximately 65%, J. Bechsgaard, unpublished results).

## Acknowledgments

We thank Deborah Charlesworth and Sylvain Glémin for comments and suggestions. The study was supported by

grants from the Carlsberg Foundation and from the Danish Natural Sciences Research Council grant no. 00001262 to M.H.S and a travel grant by the French Embassy in Denmark to T.B. and M.H.S.

## References

- Boyes, M., Nasrallah, E., Vrebalov, J. & Nasrallah, J.B. 1997. Self-incompatibility (S) haplotypes in Brassica contain divergent and rearranged sequences of ancient origin. *Plant Cell* **9**: 237–247.
- Campbell, J.M. & Lawrence, M.J. 1981. The population genetics of the self-incompatibility polymorphism in *Papaver rhoas*. II. The number and frequency of S-alleles in a natural population (R106). *Heredity* **46**: 181–190.
- Casselman, A.L., Vrebalov, J., Conner, J., Singhal, A., Giovannoni, J., Nasrallah, M.E. & Nasrallah, J.B. 2000. Determining the physical limits of the Brassica S locus by recombinational analysis. *Plant Cell* **12**: 23–34.
- Charlesworth, D., Bartolomé, C., Schierup, M.H. & Mable, B.K. 2003. Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. *Molecular Biology and Evolution* **20**: 1741–1753.
- Crosby, J.L. 1949. Selection of an unfavourable gene-complex. *Evolution* **3**: 212–230.
- Dixit, R., Nasrallah, M.E. & Nasrallah, J.B. 2000. Post-transcriptional maturation of the S receptor kinase of Brassica correlates with co-expression of the S-locus glycoprotein in the stigmas of two Brassica strains and in transgenic tobacco plants. *Plant Physiol.* **124**: 297–311.
- Glémin, S., Bataillon, T., Ronfort, J., Mignot, A. & Olivieri, I. 2001. Inbreeding depression in small populations of self-incompatible plants. *Genetics* **159**: 1217–1229.
- Hony, D. & Twell, D. 2003. Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol.* **132**: 640–652.
- Junghans, H. & Metzloff, M. 1990. A simple and rapid method for the preparation of total plant DNA. *Biotechniques* **8**: 176.
- Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J.B. & Nasrallah, M.E. 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.
- Lawrence, M.J. 2000. Population genetics of the homomorphic self-incompatibility polymorphisms in flowering plants. *Annals of botany* **85** (supplement A): 221–226.
- Lawrence, M.J. & Franklin-Tong, V.E. 1994. The population-genetics of the self-incompatibility polymorphism in *Papaver rhoas*. IX. Evidence of an extra effect of selection acting on the S-locus. *Heredity* **72**: 353–364.
- Lynch, M. & Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., Sunderland.
- Mable, B.K., Schierup, M.H. & Charlesworth, D. 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.
- McCubbin, A.G. & Kao, T.H. 1999. The emerging complexity of self-incompatibility (S-) loci. *Sex Plant Reprod.* **12**: 1–5.
- Nasrallah, J.B. 2000. Cell–cell signaling in the self-incompatibility response. *Curr. Opin. Plant Biol.* **3**: 368–373.
- Nasrallah, M.E., Liu, P. & Nasrallah, J.B. 2002. Generation of self-incompatible *Arabidopsis thaliana* by transfer of two S locus genes from *A. lyrata*. *Science* **297**: 247–249.

- de Nettancourt, D. 1977. *Incompatibility in Angiosperms. Monographs on Theoretical and Applied Genetics* 3. Springer-Verlag, Berlin.
- Sato, K., Nishio, T., Kimura, R., Kusaba, M., Suzuki, T., Hatakeyama, K., Ockendon, D.J. & Satta, Y. 2002. Coevolution of the S-locus genes SRK, SLG and SP11/SCR in *Brassica oleracea* and *B. rapa*. *Genetics* **162**: 931–940.
- Schierup, M.H. 1998. The number of self-incompatibility alleles in a finite, subdivided population. *Genetics* **149**: 1153–1162.
- Schierup, M.H., Vekemans, X. & Christiansen, F.B. 1997. Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. *Genetics* **147**: 835–846.
- Schierup, M.H., Vekemans, X. & Christiansen, F.B. 1998. Allelic genealogies in sporophytic self-incompatibility systems in plants. *Genetics* **150**: 1187–1198.
- Schierup, M.H., Mable, B.K., Awadalla, P. & Charlesworth, D. 2001. Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*. *Genetics* **158**: 387–399.
- Schopfer, C.R., Nasrallah, M.E. & Nasrallah, J.B. 1999. The male determinant of self-incompatibility in Brassica. *Science* **286**: 1697–1700.
- Suzuki, G., Kai, N., Hirose, T., Fukui, K., Nishio, T., Takayama, S., Isogai, A., Watanabe, M. & Hinata, K. 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.
- Takasaki, T., Hatakeyama, K., Suzuki, G., Watanabe, M., Isogai, A. & Hinata, K. 2000. The S receptor kinase determines self-incompatibility in Brassica stigma. *Nature* **403**: 913–916.
- Takayama, S., Shiba, H., Iwano, M., Shimosato, H. & Che, F.S. 2000. The pollen determinant of self-incompatibility in *Brassica campestris*. *P.N.A.S.* **97**: 1920–1925.
- Takayama, S., Shimosato, H., Shiba, H., Funato, M., Che, F.S., Watanabe, M., Iwano, M. & Isogai, A. 2001. Direct ligand-receptor complex interaction controls Brassica self-incompatibility. *Nature* **413**: 534–538.
- Uyenoyama, M.K. 2000. Evolutionary dynamics of self-incompatibility alleles in Brassica. *Genetics* **156**: 351–359.
- Uyenoyama, M.K. 2003. Genealogy-dependent variation in viability among self-incompatibility genotypes. *Theor Popul Biol.* **63**: 281–293.
- Vekemans, X. & Slatkin, M. 1994. Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics* **137**: 1157–1165.
- Vekemans, X., Schierup, M.H. & Christiansen, F.B. 1998. Mate availability and fecundity selection in multiallelic self-incompatibility systems in plants. *Evolution* **52**: 19–29.
- Wright, S. 1939. The distribution of self-sterility alleles in populations. *Genetics* **24**: 538–552.

Received 12 August 2003; revised 12 December 2003; accepted 16 December 2003