Genet. Res., Camb. (2008), **90**, *pp.* 37–46. © 2008 Cambridge University Press
doi:10.1017/S0016672307008932 Printed in the United Kingdom

Hitch-hiking to a locus under balancing selection: high sequence diversity and low population subdivision at the S-locus genomic region in *Arabidopsis halleri*

MARIA VALERIA RUGGIERO, BERTRAND JACQUEMIN, VINCENT CASTRIC AND XAVIER VEKEMANS*
Université des Sciences et Technologies de Lille 1, Laboratoire de génétique et évolution des populations végétales, CNRS UMR 8016, 59655 Villeneuve d'Ascq, France

(*Received 8 September 2007 and in revised form 15 October 2007*)

Summary

Hitch-hiking to a site under balancing selection is expected to produce a local increase in nucleotide polymorphism and a decrease in population differentiation compared with the background genomic level, but empirical evidence supporting these predictions is scarce. We surveyed molecular diversity at four genes flanking the region controlling self-incompatibility (the S-locus) in samples from six populations of the herbaceous plant *Arabidopsis halleri*, and compared their polymorphism with sequences from five control genes unlinked to the S-locus. As a preliminary verification, the S-locus flanking genes were shown to co-segregate with *SRK*, the gene involved in the self-incompatibility reaction at the pistil level. In agreement with theory, our results demonstrated a significant peak of nucleotide diversity around the S-locus as well as a significant decrease in population genetic structure in the S-locus region compared with both control genes and a set of seven unlinked microsatellite markers. This is consistent with the theoretical expectation that balancing selection is increasing the effective migration rate in subdivided populations. Although only four S-locus flanking genes were investigated, our results suggest that these two signatures of the hitch-hiking effect are localized in a very narrow genomic region.

1. Introduction

Since the seminal work of Maynard Smith & Haigh (1974), the 'hitch-hiking' effect has been recognized as an important cause of genome-wide variation in neutral molecular diversity (Fay & Wu, 2000). For instance, it has been put forward as a major explanation for the observed correlation between genomic diversity and local recombination rates in *Drosophila* and humans (Begun & Aquadro, 1992; Payseur & Nachman, 2002). In the presence of linkage disequilibrium between loci, hitch-hiking can in fact change the frequency of an allele depending on selection acting at a distinct but closely linked locus. If positive selection is acting, the favoured allele increases its frequency in the population, reducing allelic diversity at the target locus. When such a selective sweep occurs in regions of low recombination, neutral variability at flanking linked genes follows the evolutionary trajectory of the selected locus. A genomic region in which neutral variation is lower than expected under the mutation–drift balance could thus represent the signature of a recent selective sweep. Most of the studies so far that suggested evidence for an hitch-hiking effect were considering a decrease in neutral molecular diversity due to purifying or positive selection at a linked locus (reviewed in Fay & Wu, 2000).

However, hitch-hiking to a locus subject to balancing selection would have the opposite effect. Since balancing selection maintains alternative alleles at the selected locus within populations for much longer evolutionary times than neutral alleles (Gillespie, 1991; Takahata, 1990), closely linked neutral variants would also be actively maintained, thereby substantially increasing local levels of neutral molecular diversity (Charlesworth, 2006; Hudson & Kaplan, 1988). The predicted increase in nucleotide diversity

* Corresponding author. Laboratoire GEPV, UMR CNRS 8016, Bat. SN2, Université des Sciences et Technologies de Lille 1, 59655 Villeneuve d'Ascq, France. Telephone: +33 3 20 43 67 53. Fax: +33 3 20 43 69 79. e-mail: xavier.vekemans@univ-lille1.fr

near the selected locus is due to the evolutionary divergence among allelic subpopulations linked to distinct alleles at that locus, because of their long coalescence times (Kamau *et al.*, 2007; Takahata & Satta, 1998). By analogy to the island model of migration, recombination allows 'migration' between allelic subpopulations, disrupting the coupling between the evolutionary fate of the target locus and its nearby genomic region. As a consequence, the diversity-enhancing effect of balancing selection is predicted to decline with chromosomal distance, at a rate dependent on local recombination rates.

One of the strongest forms of balancing selection described in natural populations is negative frequency-dependent (NFD) selection with a rare-allele advantage. NFD selection is predicted to maintain many alleles in populations for long evolutionary times, generating high allelic diversity and consequently high levels of sequence polymorphism (Takahata, 1990). A less recognized feature of systems under NFD selection is a noticeable reduction in among-population genetic differentiation at the selected locus, compared with neutral references, due to the high invasion success of immigrating rare alleles in the recipient population, causing a rapid increase in their frequency (Muirhead, 2001; Schierup *et al.*, 2000*b*). By studying hitch-hiking to a locus subject to balancing selection in a subdivided population, Schierup *et al.* (2000*a*) also showed that a 'valley' in F_{ST} values for neutral loci was expected in the genomic region surrounding the selected locus, extent of which was dependent on the local recombination rate and selection strength.

A well-known example of NFD selection is that acting on homomorphic self-incompatibility (SI) systems in hermaphroditic plants (Wright, 1939), where a mating occurs only if the allelic specificity expressed at the stigma surface is different from the one borne by the pollen grain (de Nettancourt, 2001). In Brassicaceae, pollen and stigma allelic specificities are encoded by two genes (respectively *SCR* and *SRK*) located in a region commonly referred to as the S-locus (Kusaba *et al.*, 2001). The two genes are thought to be under very strong linkage, as recombination would decouple pollen and pistil specificities, thereby leading to a breakdown of SI (Casselman *et al.*, 2000; Awadalla & Charlesworth, 1999). This genomic region is thus experiencing very low rates of recombination. Consistent with strong NFD selection, an extremely high allelic diversity, with elevated levels of divergence between allelic classes, has been observed at *SRK* in several Brassicaceae species (reviewed in Castric & Vekemans, 2004; Charlesworth *et al.*, 2005; Takebayashi *et al.*, 2003). Kamau & Charlesworth (2005) showed that a peak of synonymous nucleotide diversity was present in the genomic region surrounding the S-locus in a single population of *Arabidopsis lyrata*, consistent with a hitch-hiking scenario and confirming that recombination could be suppressed in the region. This result was confirmed by a more detailed analysis using a larger set of flanking genes and characterizing the extent of linkage disequilibrium between those and the S-locus across several populations (Kamau *et al.*, 2007). Further support for strong long-term hitch-hiking effects was also provided by the observation of high numbers of trans-specific polymorphisms at two S-locus flanking genes between the species *A. lyrata* and *A. thaliana* (Charlesworth *et al.*, 2006). In contrast, in *Nicotiana alata*, low diversity was found in a gene located just a few kilobases from the S-locus (Takebayashi *et al.*, 2003).

Based on a species-wide sampling approach, we examined sequence diversity and population genetic structure in the genomic region surrounding the S-locus in *Arabidopsis halleri*, a sister species to *A. lyrata* with which it shares the same SI system (Bechsgaard *et al.*, 2006; Castric & Vekemans, 2007). We obtained nucleotide sequences for four genes in the S-locus genomic region as well as five unlinked control genes, in a sample composed of individuals from six geographically distant populations. For S-linked genes, we checked for co-segregation with the *SRK* locus by analysing a large backcross family. For each gene we estimated synonymous nucleotide diversity as well as among-population differentiation, checking for differences between the sets of S-linked and control genes. Our results are consistent with the hitch-hiking model affecting both patterns of molecular diversity and population structure.

2. Materials and methods

(i) *Sampling*

Individual DNAs were chosen from a collection of European accessions of *Arabidopsis halleri* (Pierre Saumitou-Laprade, University of Lille 1). Based on a previous study that used this collection (Pauwels *et al.*, 2005) we have chosen six populations representative of the species-wide diversity. Populations were from Germany (Harz, abbreviated as AL13, located at 51°55′ N, 10°19′ E), France (Auby, AU: 50°25′ N, 3°03′ E), Italy (St Leonhard in Passeier, I5: 46°49′ N, 11°15′ E), Poland (Katowice, PL1: 50°15′ N, 18°57′ E), Slovenia (Stojnci, SLO5: 46°22′ N, 16°00′ E) and the Czech Republic (Zaton, TC8: 48°57′ N, 13°48′ E). Five individuals from each population were randomly chosen, except for the population of Auby from which 6 individuals were taken, for a total of 31 samples.

Attempts to genotype individuals at the *SRK* gene were performed *a posteriori* using PCR primer pairs designed to amplify specifically each of the 25 S-haplotypes currently identified in *A. halleri*

https://doi.org/10.1017/S0016672307008932 Published online by Cambridge University Press

(Castric & Vekemans, 2007; primer sequences given in Supplementary Table S1). We identified two S-haplotypes at *SRK* in 11 individuals, a single S-haplotype in 14 individuals, and none in six individuals (Supplementary Table S2). Missing data is probably due to the occurrence of S-haplotypes that have not been characterized yet. Genotypes with a single S-haplotype could also be homozygotes for recessive alleles. Because of these uncertainties, and because we had only access to DNA samples, it was not possible to investigate the association between particular S-haplotypes and observed haplotypes at the S-locus flanking genes.

(ii) *Loci surveyed*

We obtained sequences for nine genes. Four of them are located in the genomic region flanking the S-locus at various distances from each side of *SRK*, according to the physical maps available for *A. thaliana* (www.arabidopsis.org) and *A. lyrata* (Kusaba *et al*., 2001). The genes flanking the S-locus are: *ARK3* (At4g21380 according to the *A. thaliana* genome annotation, 1·7 kb from *SRK*), *B120* (At4g21390, 7·2 kb) and *B160* At4g21430, 20·7 kb) on one side of *SRK* and *B80* (At4g21350, 26·6 kb) on the other side.

To test for the occurrence of hitch-hiking, these flanking genes were compared with five presumably unlinked loci, namely: *HAT4* (At4g16780), *CAULIFLOWER* (*CAUL*) (At1g26310), *CHS* (At5g13930), *scADH* (At4g05530) and *Aly9* (called *SLR1* in *A. thaliana*, At3g12000, which belongs to the same gene family as *SRK*; Charlesworth *et al*., 2003). Although *HAT4* and *scADH* are located on the same chromosome as the S-locus, they are sufficiently distant from *SRK* to be considered as genetically unlinked. For genes flanking *SRK*, all individuals sampled were sequenced, while for control genes a subset of individuals (3 per population) was used. Sequences for *B80* and *Aly9* consisted of a single exon.

In order to get a control for genome-wide population differentiation, seven microsatellite loci (AthZFPG, GC22, H117, Ice 13, MDC16, nga 112 and nga 361) were also genotyped on the whole sample, using a multiplex PCR procedure according to Llaurens *et al*. (submitted).

(iii) *PCR amplification and sequencing*

Primer sequences for amplification of *B120* and *B160* were taken from Kamau & Charlesworth (2005). For *B80*, the reverse primer was as in Kamau & Charlesworth (2005), and the forward primer was newly designed based on the *A. thaliana* sequence: *B80*-Ah-F 5′CGATCGGGTCTCTATCCAAC3′. Primers for *ARK3* were taken from Hagenblad *et al*. (2006). For *HAT4*, primer sequences were taken from Wright *et al*. (2003), for *CAUL* from Purugganan & Suddith (1998) and for *CHS* from Ramos-Onsins *et al*. (2004). For *scADH*, primers were designed from *A. thaliana* and were as follows: *scADH*_Ah_F2 5′ACATCGCCGCAATCTTGT3′ and *scADH*_Ah_R2 5′CAGAAGAACCCTTCTCTAGGTGA3′. Annealing temperatures were as follows: 54 °C for *B120* and *B160*, 57 °C for *CAUL* and *CHS*, 58 °C for *B80* and for *ARK3*, 65 °C for *scADH*, *HAT4* and *Aly9*. PCR amplification consisted of 1 min at 94 °C, 40 s at the annealing temperature and 40 s at 72 °C, for 35 cycles.

PCR fragments were cloned into the PCR 2.1 vector using the TA cloning kit (Invitrogen Life Technologies), and at least eight clones were sequenced individually using the BigDye Terminator Kit 3.1 (Applied Biosystems) and run on an ABI-3100 capillary sequencer (Applied Biosystems). The universal M13 primers were used to sequence cloned PCR products. All sequences have been deposited in GenBank, with accession numbers EU273946–EU274288.

(iv) *Co-segregation analysis*

Although a physical map of the S-locus in *A. halleri* is not available, it is available for two haplotypes in its sister species *A. lyrata*. Kusaba *et al*. (2001) showed high levels of macrosynteny in the S-locus region between *A. lyrata* and *A. thaliana*, suggesting that the relative position of the genes is well conserved in the genus, although intergenic distances are variable. In order to verify co-segregation of flanking genes with *SRK* in *A. halleri*, we genotyped individuals from a large first-generation backcross family available in the laboratory (Willems *et al*., 2007). In detail, the F1 of an interspecific cross *A. halleri* × *A. lyrata* was backcrossed against a different *A. lyrata* individual and 331 progeny were genotyped for *B80*, *B120*, *B160*, *ARK3* and *HAT4*. This backcross family had been used previously to obtain a saturated genetic map in which three markers were located at less than 10 cM from *SRK* (Willems *et al*., 2007), allowing us to estimate recombination rates in the S-locus genomic region.

Nucleotide sequences for each flanking gene and for *HAT4* were obtained for each parent, and the web-based software WebCutter 2.0 (http://rna.lundberg.gu.se/cutter2/) was then used to generate a restriction map for each allele. Restriction enzymes were chosen so as to distinguish parental alleles in the progeny based on expected banding patterns. All restriction reactions were carried out for at least 4 h at 37° C, using 1 u enzyme/100 ng DNA. Fragments were separated through electrophoresis on 2% agarose gels and visualized using ethidium bromide under ultraviolet light. A single restriction enzyme per gene was

sufficient to distinguish the *A. halleri* allele from that of *A. lyrata*.

(v) *Data analyses*

Sequences were aligned using ClustalW software (Thompson *et al.*, 1994), implemented in BioEdit v.7.0.4.1 (Hall, 1999). BioEdit was also used to manually adjust alignments when necessary. Coding regions were identified according to the *A. thaliana* genome annotation provided by the TAIR database (www.arabidopsis.org). In the case of homozygous individuals, two copies of the allele were included in the alignment. For some heterozygous individuals it was not possible to obtain the second allele, and only one was included in the data set.

Total, synonymous, non-synonymous and silent nucleotide diversity values (π; Nei & Gojobori, 1986; Nei, 1987) were calculated for each gene using DNAsp v.4.10 (Rozas *et al.*, 2003). The average within-population diversity was also calculated, weighted by the sample sizes (Wright *et al.*, 2003).

Under the neutral model, within-species polymorphism should be proportional to between-species divergence (Hudson *et al.*, 1987). In order to test for deviation from neutral expectations in flanking genes, a multilocus HKA test (Hudson *et al.*, 1987) was used, testing the significance of the ratio of polymorphism within *A. halleri* to divergence between *A. halleri* and *A. thaliana* on all loci. Sequences for *A. thaliana* for each gene were retrieved from the TAIR database and used to estimate average nucleotide divergence between *A. halleri* and *A. thaliana*. To test for significance, 5000 coalescent simulations were run using the HKA program (http://lifesci.rutgers.edu/wheylab/HeylabSoftware.htm). We also used the maximum-likelihood multilocus HKA framework developed by Wright & Charlesworth (2004) to test for an overall difference in polymorphism between the set of four S-locus flanking genes against the control genes. Specifically, we used the MLHKA program distributed by Stephen Wright to compare a model with free mutation at each locus and no selection against a model with free mutation and selection on the four S-locus flanking genes. Since these two models are nested, we used a log-likelihood ratio test with 4 degrees of freedom to compare their likelihood. Chain length was set to 100 000. Deviation from neutrality was also tested using Tajima's *D* for each gene (Tajima, 1989), for which an excess or a lack of intermediate frequency polymorphisms suggests the presence, respectively, of balancing selection (positive values of *D*) or purifying selection (negative values of *D*). All these analyses were conducted with DNAsp, unless specified otherwise.

The genetic differentiation between the six sampled populations was assessed with the statistic F_{ST} computed according to Hudson *et al.* (1992*a*), as implemented in DNAsp. Significance of population differentiation was assessed through the Kst* test (Hudson *et al.*, 1992*b*) using 1000 permutations. We tested for differences between mean values of F_{ST} for S-linked versus control genes using the Mann–Whitney non-parametric test.

Table 1. *Length of the alignments and number of sequences obtained from each population for each gene*

	Length (bp)	AU	AL13	I	PL1	SLO5	TC8	Tot
S-locus flanking								
B160	1154	10	10	8	10	8	10	56
B120	998	12	10	8	10	7	10	57
ARK3	942	9	8	7	7	4	8	43
B80	843	12	10	9	10	7	10	58
Control genes								
HAT4	911	6	6	6	6	6	4	34
scADH	1477	6	6	6	6	4	6	34
CAUL	911	6	6	6	6	6	6	36
Aly9	497	6	6	2	6	4	6	30
CHS	866	4	6	4	4	4	6	28

Sample size per population is 5 individuals for S-locus flanking genes (6 in the population from Auby, AU) and 3 individuals for control genes. For detailed information on the number of sequences obtained for each individual, see Supplementary Table S3.

3. Results

Due to differences in sample sizes among genes and populations, and to technical problems with the cloning procedure, we obtained two to 12 sequences for each gene in each population, for a total ranging from 28 to 58 (Table 1 and Supplementary Table S3). Few homozygous individuals were found (8 in *Aly9*, 6 in *CAUL*, 5 in *CHS*, 2 in *HAT4*, 2 in *scADH* and 9 in *B160*). The length of the aligned sequences varied from 497 bp for *Aly9* to 1477 bp for *scADH*. For *ARK3*, two different groups of sequences were found, differing by a long indel of 346 bp in the intron as was previously reported in *A. lyrata* (Hagenblad *et al.*, 2006). Two smaller indels were also found within the coding sequence: an in-frame 3 bp indel, and an 8 bp-long indel introducing a premature stop codon. For the latter, the associated sequence was thus considered as a pseudogene (it occurred in a single individual for which three distinct *ARK3* sequences were obtained) and discarded from further analyses. In a few other cases, three haplotypes were found per individual. The presence of pseudogenes, characterized by indels in the coding parts of the gene and the existence of duplicated copies in *ARK3* have also been reported in *A. lyrata* (Hagenblad *et al.*, 2006), although it is not clear whether the duplicated copies

https://doi.org/10.1017/S0016672307008932 Published online by Cambridge University Press

Table 2. *Results of the co-segregation analysis: multilocus haplotypes observed in the backcross progeny are listed, together with the number of individuals sharing them*

Distance from *SRK* (kb)	*TSB2* 2199·3	*At4-TC1* 1116·7	*B160* 20·7	*B120* 7·2	*ARK3* 1·8	*SRK*	*B80* 26·6	*HAT4* 1933·1	*FCA* 2169·1	*N* ind
	l	l	l	l	l	l	l	l	l	160
	l	l	l	l	l	l	l	l	**h**	1
	l	l	l	l	l	l	l	**h**	**h**	13
	l	l	l	l	l	l	**h**	**h**	**h**	1
	l	l	**h**	**h**	**h**	**h**	**h**	**h**	**h**	8
	l	**h**	**h**	**h**	**h**	**h**	**h**	**h**	**h**	5
	h	l	l	l	l	l	l	**h**	**h**	1
	h	l	l	l	l	l	l	l	l	7
	h	**h**	l	l	l	l	l	l	l	5
	h	**h**	**h**	**h**	**h**	**h**	**h**	l	l	12
	h	**h**	**h**	**h**	**h**	**h**	**h**	**h**	l	1
	h	**h**	**h**	**h**	**h**	**h**	**h**	**h**	**h**	116
N recombinants with *SRK*	26	13	0	0	0	–	1	27	29	331
Expected *N* recombinants[a]	29·12	14·79	0·27	0·10	0·02	–	0·35	25·59	28·72	
Genetic distances (cM) from *SRK*	7·85	3·93	0·0	0·0	0·0	–	0·30	8·16	8·76	
Recombination rate in the interval (cM/Mb)		3·63	3·52	0	0	0	11·36	4·12	2·56	

For each locus, the distance in kilobases from SRK (according to the *A. thaliana* physical map), the number of recombinant individuals and the genetic distances are shown. A total of 331 individuals have been genotyped.
l, allele from *A. lyrata*; *h*, allele from *A. halleri*.
[a] Expected number of recombination events with *SRK* occurring under the hypothesis of a 4 cM/Mb recombination rate in the region.

do co-segregate. For the purpose of conservativeness, we eliminated from our data set all putative pseudogenes and individuals with duplicated copies of *ARK3*. Numerous indels of up to 91 bp were also observed in the introns of *scADH*.

(i) *Co-segregation analysis*

Table 2 shows the results of the co-segregation analysis in a first-generation backcross progeny of *A. halleri* and *A. lyrata* between *SRK* and each of the *B80*, *ARK3*, *B120*, *B160* and *HAT4* genes together with three previously studied marker loci of the genomic region surrounding the S-locus (*TSB2*, *At4-TC1* and *FCA*; Willems *et al.*, 2007). This region spans about 2·2 Mb on each side of *SRK*, as inferred from *A. thaliana*'s *Col-0* full genomic sequence. In a total progeny size of 331 individuals, 55 individuals (16%) showed a recombination event within the 4·4 Mb region surveyed. Among the flanking genes, no recombinants were detected between *SRK* and each of the *ARK3*, *B120* and *B160* genes, whereas one recombination event was observed with *B80*. For *HAT4*, about 2 Mb distant from *SRK* (according to the *A. thaliana* genome), 25 recombinants were detected. Our results confirm thus that the four putative S-locus flanking genes studied here (*B80*, *ARK3*, *B120* and *B160*) are closely linked to *SRK*, while *HAT4* cannot be considered as S-linked. The region between markers TSB2 and FCA spanned over 7·85 cM on one side of *SRK* and 8·76 cM on the other side, which amounts to an overall recombination rate of 3·8 cM/Mb, a value very close to the average genomic rate in non-centromeric regions reported by Kawabe *et al.* (2006) for *A. lyrata* (4 cM/Mb). Additionally, within the 2 Mb region lying between *SRK* and marker *TSB2*, we found very similar estimates of recombination rate per physical distance in the 1 Mb segment adjacent to *SRK* (3·52 cM/Mb in the region between the S-locus and the marker *At4-TC1*) and the next 1 Mb segment (3·63 cM/Mb in the region between *At4-TC1* and *TSB2*). These results are only compatible with a narrow region of suppressed recombination around the S-locus.

(ii) *Patterns of polymorphism in S-linked and control genes*

Patterns of nucleotide polymorphism in S-linked and control genes are shown in Table 3. S-locus flanking genes showed on average nucleotide diversities about twice as high as those in control genes, both at the overall and at the within-population levels, and for total nucleotide diversity as well as for synonymous or silent diversity. The two genes directly flanking the S-locus on each side, *B80* and *ARK3*, are those that show the highest level of polymorphism (about 4 times higher than the average of control genes). It should be

https://doi.org/10.1017/S0016672307008932 Published online by Cambridge University Press

Table 3. *Analysis of nucleotide polymorphism, population differentiation in* A. halleri *and divergence versus* A. thaliana *at four S-locus flanking genes and at five control genes. Values of Tajima's D statistic are also shown* (*all are non-significant*)

		N	Length (bp)[a]	θ_w[b]	π_{tot}[c]	π_{within}[d]	π_s[e]	π_{sil}[f]	π_n/π_s[g]	Tajima's *D*	F_{ST}[h]	K_{tot}[i]
S-locus flanking genes	*B160*	56	1146	0·0129	0·0117	0·0119	0·0291	0·0183	0·2144	−0·3252	−0·0155 ns	0·0519
	B120	57	912	0·0152	0·0154	0·0141	0·0351	0·0313	0·0635	0·0432	0·1042**	0·0509
	ARK3	43	522	0·0363	0·0363	0·0334	0·077	0·0599	0·0604	0·0001	0·0756*	0·0853
	B80	58	843	0·0277	0·029	0·0277	0·0862	0·0862	0·1074	0·1751	0·0522**	0·0543
	Average			0·0230	0·0231	0·0218	0·0569	0·0489	0·0984	−0·0267	0·0541	0·0606
	(SD)			(0·0110)	(0·0115)	(0·0104)	(0·0289)	(0·0303)	(0·1019)	(0·2124)	(0·0511)	(0·0165)
Control genes	HAT4	34	884	0·0055	0·0038	0·0036	0·0071	0·0062	0·1831	−1·0581	0·0908*	0·0630
	CAUL	36	824	0·0114	0·012	0·0016	0·0242	0·0142	0·2880	0·1896	0·2484***	0·0980
	ScADH	34	1196	0·0133	0·0146	0·0129	0·0288	0·0239	0·0372	0·3716	0·1338**	0·0932
	Aly9	30	497	0·0031	0·0026	0·0023	0·0082	0·0082	0·1350	−0·415	0·182**	0·0766
	CHS	28	860	0·0081	0·0101	0·0062	0·0399	0·0297	0·0271	0·8973	0·3771**	0·0572
	Average			0·0083	0·0086	0·0053	0·0216	0·0164	0·1065	−0·0029	0·2064	0·0776
	(SD)			(0·0042)	(0·0052)	(0·0046)	(0·0140)	(0·0101)	(0·1865)	0·7535	(0·1120)	(0·0179)

[a] Sequence length excluding all gaps from the alignment.
[b] Overall Watterson's θ statistic computed over all sites.
[c] Overall nucleotide diversity computed over all sites.
[d] Average within-population nucleotide diversity computed over all sites.
[e] Overall synonymous nucleotide diversity.
[f] Overall silent diversity (synonymous sites + introns).
[g] Ratio of overall non-synonymous over synonymous nucleotide diversity.
[h] Estimates of FST and results from the Kst* test: ns, not significant; *$P<0·5$; **$P<0·1$; ***$P<0·001$.
[i] Number of substitutions per total site versus *A. thaliana*.

noted, however, that the polymorphism previously reported at *SRK* in *A. halleri* is still an order of magnitude higher than at these two polymorphic genes ($\pi_{tot}=0·388$ and $\pi_s=0·819$; Castric & Vekemans, 2007). The average ratio of π_n (non-synonymous sites only) over π_s (synonymous sites) was about 0·1 in both types of genes, suggesting no apparent overall difference in the level of adaptive constraints between them.

In the HKA test we compared nucleotide polymorphism within *A. halleri* with divergence from *A. thaliana* for the four S-locus flanking genes, and for all control genes. The multilocus HKA test showed a significant departure from neutral expectation ($\chi^2=22·57$, $P=0·004$), indicating that some loci differed in their relative patterns of polymorphism and divergence. In Fig. 1, the relative contribution of polymorphism (filled diamonds) and divergence (open squares) to the overall χ^2 test statistics at each locus is shown. S-linked genes showed a clear excess of intraspecific polymorphism with respect to neutral expectations and a deficit in interspecific divergence, while the opposite situation was found in control genes. One control gene, *Aly9*, showed a substantial excess of divergence relative to polymorphism, suggesting a signature of positive directional selection that could bias our results. We performed the multilocus HKA test by excluding this gene and still found significant departure from neutral expectation

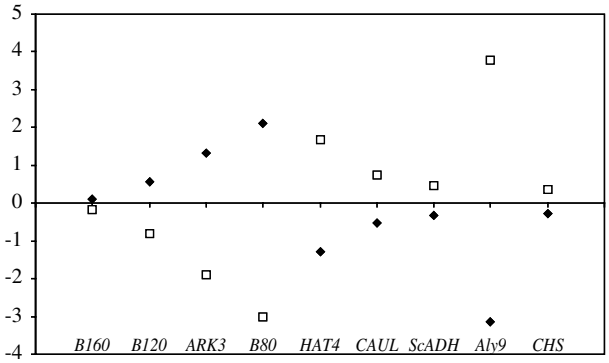

Fig. 1. Results of the multilocus HKA test. Filled diamonds represent deviation from the neutral expectation for polymorphism and open squares represent divergence. Points above the line indicate deviation toward an excess, points below the line indicate deviation towards a deficit. The test rejects neutrality ($P=0·004$).

($\chi^2=18·50$, $P=0·010$). Comparison of a free-mutation model allowing for selection at the four S-locus flanking genes, with a strictly neutral model, under a maximum likelihood framework, gives log likelihood-ratio statistics of 27·0 and 20·8, for runs with or without *Aly9*, respectively, which are highly significant ($P<0·001$) against the χ^2 distribution with 4 degrees of freedom. The estimates of the selection parameter for *B160*, *B120*, *ARK3* and *B80* are, respectively, 2·48, 3·20, 4·50 and 5·48. This parameter

https://doi.org/10.1017/S0016672307008932 Published online by Cambridge University Press

corresponds here to the relative increase in polymorphism due to genetic hitch-hiking, taking into account differences across loci in patterns of divergence between *A. halleri* and *A. thaliana*. Hence, these results provide evidence for a higher overall level of polymorphism at the four S-locus flanking genes compared with control loci. Tajima's *D* values, however, did not detect any deviations from neutrality for any of the genes, either S-linked or unlinked, and showed no differences between mean values for S-linked and control genes (Mann–Whitney test, $P = 0{\cdot}46$).

(iii) *Population genetic differentiation*

Results for population subdivision show a clear difference between S-linked and control genes, with the expected valley in F_{ST} values around the S-locus (Table 3). We obtained consistent estimates of F_{ST} for the seven unlinked microsatellite loci ($F_{ST} = 0{\cdot}204$) and for the control genes ($F_{ST} = 0{\cdot}206$; range 0·09–0·38). Such close agreement between these two types of loci is typically not found because of large differences in mutation rates (Charlesworth, 1998), but is consistent with the observation of moderate polymorphism at the studied microsatellite loci (heterozygosity values ranging from 0·28 to 0·64, data not shown). This also suggests that the small sample size used for control genes has not produced a biased estimate of F_{ST}. In contrast, the average F_{ST} for S-locus flanking genes was about 4 times lower ($F_{ST} = 0{\cdot}054$; range $-0{\cdot}02$ to 0·10). Still, most of the loci, except *B160*, showed significant population differentiation with the Kst* test (Table 3). Overall, control loci showed significantly higher levels of population differentiation than S-locus flanking genes (Mann–Whitney test, $P = 0{\cdot}027$). Hence, these results thus clearly demonstrate the expected valley in F_{ST} values around the S-locus.

4. Discussion

In the presence of balancing selection, genetic hitch-hiking is expected to cause a higher polymorphism in the flanking region compared with unlinked regions, but the size of this affected region can be small, depending on the local recombination rate (Wiuf *et al.*, 2004, Takahata & Satta, 1998). In this paper, we confirmed in *A. halleri* the observation of Kamau & Charlesworth (2005) that hitch-hiking in the region of the S-locus in *A. lyrata* produced a local increase in polymorphism, and showed that it is also causing a local decrease in population genetic structure, compared with control loci unlinked to the S-locus. Such a localized effect on population structure was predicted by theoretical models (Schierup *et al.*, 2000*a*) and is due to the fact that strong NFD selection acting on the S-locus within subpopulations counteracts allele frequency divergence due to random genetic drift and causes higher effective migration rate for S-alleles than for neutral alleles (Schierup *et al.*, 2000*b*, Muirhead, 2001).

(i) *How large is the genomic region linked to the S-locus?*

Co-segregation analyses based on 331 offspring from a first-generation backcross between *A. halleri* and *A. lyrata* showed that the four putative S-locus flanking genes investigated here are indeed closely linked to *SRK*. Moreover our data provided some estimates of recombination rate in the S-locus region. Kawabe *et al.* (2006) analysed recombination data in a mapping population of 99 individuals from a F2 intraspecific cross in *A. lyrata* subsp. *petraea*. They found evidence for a significantly lower recombination rate in a region spanning over about 600 kb from each side of the S-locus, compared with more distant flanking regions. In contrast, our results do not show indication of a reduced recombination rate in a 4 Mb region centred on the S-locus compared with the average genomic rate in non-centromeric regions reported by Kawabe *et al.* (2006) for *A. lyrata* (4 cM/Mb). We have no definite explanation for the discrepancy between our results and those of Kawabe *et al.* (2006), but note several differences between the two studies. First, our mapping population consisted of an interspecific first-generation backcross, which could generate some bias. However, any bias expected would be of the opposite direction, i.e. generating lower rather than higher recombination rates, compared with intraspecific crosses (Williams *et al.*, 1995). Moreover, Willems *et al.* (2007) showed for the same mapping family that recombination between *A. halleri* and *A. lyrata* genomes was as efficient in the interspecific hybrid as in the *A. lyrata* intraspecific crosses, based on the studies by Kuittinen *et al.* (2004) and Yogeeswaran *et al.* (2005). Secondly, the number of markers in the study of Kawabe *et al.* (2006) was higher than in ours, with notably many markers in the 600 kb region on both sides of the S-locus, whereas we had no marker in the interval between 25 kb and 1 Mb. Thus, due to the paucity of markers, we could have missed the region of reduced recombination. However, the high overall rate of recombination that we observed would still not be consistent with an extended region of suppressed recombination around the S-locus. Hence, a more detailed study with large mapping populations and large numbers of markers would be necessary to better characterize the size of the genomic region in close linkage to the S-locus. Uncertainty in the estimates of recombination rate also arises because the physical sizes of the S-locus regions in *A. halleri* and *A. lyrata* are not known

https://doi.org/10.1017/S0016672307008932 Published online by Cambridge University Press

precisely. The physical map of *A. thaliana* was used for calculations in this study and that of Kawabe *et al.* (2006), which could lead to overestimates of recombination rates because the genome of *A. thaliana* is about 40% smaller than that of its sister species (Johnstone *et al.*, 2005). Notably the size of the S-locus itself has been found to be larger in two S-haplotypes of *A. lyrata* compared with that of the *A. thaliana* Col-0 accession (Kusaba *et al.*, 2001), and to vary extensively between the two S-haplotypes. This could lead to differences among S-haplotypes in the size of the region with reduced recombination, as suggested by observed variation in the strength of linkage between *ARK3* and *SRK* (Hagenblad *et al.*, 2006) or between *B80* and *SRK* (Kamau *et al.*, 2007) across S-haplotypes. As discussed in Kawabe *et al.* (2006), better knowledge of the size of the region in linkage disequilibrium with the S-locus is of critical importance to evaluate the evolutionary significance of S-haplotype-specific genetic load, a feature that would be generated in non-recombining regions surrounding the S-locus because of high heterozygosity and long divergence times of S-haplotypes (Uyenoyama, 1997). Experimental observations consistent with the existence of a linked genetic load in the S-locus region have been reported by Stone (2004) in *Solanum carolinense*, and by Bechsgaard *et al.* (2004) in *A. lyrata*. By fitting data on diversity at flanking neutral sites within and between different functional S-allele classes to a model of hitch-hiking under balancing selection, Kamau *et al.* (2007) obtained indirect estimates of the recombination rate in the S-locus region. Their results pointed out that recombination is highly reduced in only a very narrow region flanking the S-locus, so that the number of highly linked genes is expected to be low. We suggest that this conclusion could still be compatible with the empirical observations of linked genetic load if this were to be caused by few deleterious mutations with large effect. The approach of Kamau *et al.* (2007) could not be used in this study because we could not resolve the phases between *SRK* and the flanking genes.

(ii) *Patterns of polymorphism and deviations from neutrality*

Patterns of total, synonymous and silent nucleotide diversity observed in *A. halleri* in the present study were all consistent with the predicted hitch-hiking effect. Results from HKA tests indicated significant heterogeneity among genes in relative patterns of polymorphism and divergence, and more specifically showed that average polymorphism at the four S-locus flanking genes was higher than at the control genes. Moreover, the two genes closely flanking *SRK* on each side, i.e. *ARK3* and *B80*, showed the highest levels of polymorphism, suggesting that the hitch-hiking effect is highly localized. These results are in close agreement with those reported in *A. lyrata*, which also showed a strong but localized signature of the hitch-hiking effect in the S-locus region (Kamau & Charlesworth, 2005; Kamau *et al.*, 2007). It is remarkable that a localized peak in nucleotide polymorphism has also been found in the S-locus region in *A. thaliana* (Shimizu *et al.*, 2004), and was interpreted as a transient signature of past hitch-hiking effects that occurred before the breakdown of self-incompatibility along the *A. thaliana* lineage (Charlesworth & Vekemans, 2005). Our results are also consistent with observations by Castric & Vekemans (2007) of a strong hitch-hiking effect on neutral polymorphism within the *SRK* gene itself, with exons 2 to 6 of this gene exhibiting extremely high polymorphism ($\pi_S = 0{\cdot}3$–$0{\cdot}7$) although they are not involved in the allelic-specificity determination. In contrast to the nucleotide polymorphism/divergence data, the Tajima's *D*-tests did not suggest any deviation from neutrality. Non-selective effects such as demographic events or population structure, however, have been demonstrated to affect this estimator (e.g. Schierup *et al.*, 2000*a*), reducing its reliability in detecting selection in flanking regions. An alternative explanation for the increased diversity at the S-locus flanking genes would be a direct effect of balancing selection on those genes. Surprisingly, it has been shown recently that one of the studied genes, *B80*, is a modifier of the expression of self-incompatibility through its role in regulation of endogenous *SRK* transcript levels in the stigmas (Liu *et al.*, 2007). However, phenotypic effects were shown to depend on variation occurring in the 5′ promoter region of the gene and not within the coding region surveyed in the present study.

(iii) *Patterns of population structure*

Schierup *et al.* (2000*b*) have shown theoretically that the effective migration rates are higher at loci under strong balancing selection than at neutral loci. This is intuitively sound, due to the underlying NFD selection, which prevents allele frequency divergence due to drift and gives an advantage to the migrant allele not previously present in the recipient population. Few empirical studies have compared population genetic structure between the S-locus and unlinked control genes or markers, but the results seem to confirm theoretical expectations. In *Brassica insularis*, Glémin *et al.* (2005) found a significantly lower population structure at the S-locus compared with microsatellite markers. In *A. lyrata*, Charlesworth *et al.* (2003) did not detect significant genetic differentiation among five populations at the *SRK* gene, whereas significant genetic structure was found at four of six

https://doi.org/10.1017/S0016672307008932 Published online by Cambridge University Press

control genes (see also Wright *et al*., 2003 for F_{ST} at control loci). In *A. halleri*, F_{ST} at *SRK* has been found to be threefold lower than at unlinked microsatellite markers (Castric *et al*., in prep.). In a separate theoretical paper, Schierup *et al*. (2000*a*) predicted that F_{ST} values for neutral sites flanking a locus subject to balancing selection should increase with increasing recombination rate from the selected locus. We found significantly lower F_{ST} values for the S-locus flanking genes compared with control loci, confirming the theoretical expectations of a reduced population subdivision in genomic regions subject to NFD selection. A similar trend has recently been reported in *A. lyrata*, where high diversity at two S-locus flanking genes was found to be caused by sequence differences among allelic classes at the S locus, rather than among populations (Kamau *et al*., 2007).

Altogether, our results suggest that strong balancing selection can be considered as a candidate process causing heterogeneity in polymorphism and population genetic structure across genomes, but its effects would only be detected in high-resolution genomic scans because of its rather local influence.

We thank A.-C. Holl and A. Courseaux for technical assistance, P. Saumitou-Laprade for DNA samples from *A. halleri*, G. Willems for kindly providing the genotypic data for markers TSB2, At4-TC1 and FCA, and E. Kamau and D. Charlesworth for useful discussions. We also thank the two anonymous reviewers and the associate editors for numerous comments on the manuscript. M.V.R. was supported by a postdoctoral grant from CNRS-Environmental Sciences and Sustainable Development department. This study was funded by an ATIP grant from CNRS-Life Science Department, a FEDER fund from the EU, an ARCIR grant from the Région Nord-Pas de Calais, and grant ANR-06-BLAN-0128-01 from the French National Research Agency.

References

Awadalla, P. & Charlesworth, D. (1999). Recombination and selection at *Brassica* self-incompatibility loci. *Genetics* **152**, 413–425.

Bechsgaard, J., Bataillon, T. & Schierup, M. H. (2004). Uneven segregation of sporophytic self-incompatibility alleles in *Arabidopsis lyrata*. *Journal of Evolutionary Biology* **17**, 554–561.

Bechsgaard, J. S., Castric, V., Charlesworth, D., Vekemans, X. & Schierup, M. H. (2006). The transition to self-compatibility in *Arabidopsis thaliana* and evolution within S-haplotypes over 10 Myr. *Molecular Biology and Evolution* **23**, 1741–1750.

Begun, D. J. & Aquadro, C. F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**, 519–520.

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.

Castric, V. & Vekemans, X. (2004). Plant self-incompatibility in natural populations, a critical assessment of recent theoretical and empirical advances. *Molecular Ecology* **13**, 2873–2889.

Castric, V. & Vekemans, X. (2007). Evolution under strong balancing selection, how many codons determine specificity at the female self-incompatibility gene *SRK* in Brassicaceae? *BMC Evolutionary Biology* **7**, 132.

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* **15**, 538–543.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics* **2**, 379–384.

Charlesworth, D. & Vekemans, X. (2005). How and when did *A. thaliana* become highly self-fertilising? *BioEssays* **27**, 472–476.

Charlesworth, D., Mable, B. K., Schierup, M. H., Bartolomé, C. & Awadalla, P. (2003). Diversity and linkage of genes in the self-incompatibility gene family in *Arabidopsis lyrata*. *Genetics* **164**, 1519–1535.

Charlesworth, D., Vekemans, X., Castric, V. & Glémin, S. (2005). Plant self-incompatibility systems: a molecular evolutionary perspective. *New Phytologist* **168**, 61–69.

Charlesworth, D., Kamau, E., Hagenblad, J. & Tang, C. (2006). *Trans*-specificity at loci near the self-incompatibility loci in *Arabidopsis*. *Genetics* **172**, 2699–2704.

de Nettancourt, D. (2001). *Incompatibility and Incongruity in Wild and Cultivated Plants*. Berlin: Springer.

Fay, J. C., & Wu, C.-I. (2000). Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413.

Gillespie, J. H. (1991). *The Causes of Molecular Evolution*. Oxford: Oxford University Press.

Glémin, S., Gaude, T., Guillemin, M.-L., Lourmas, M., Olivieri, I. & Mignot, A. (2005). Balancing selection in the wild: testing population genetics theory of self-incompatibility in the rare species *Brassica insularis*. *Genetics* **171**, 279–289.

Hagenblad, J., Bechsgaard, J. & Charlesworth, D. (2006). Linkage disequilibrium between incompatibility locus region genes in the plant *Arabidopsis lyrata*. *Genetics* **173**, 1057–1073.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98.

Hudson, R. R. & Kaplan, N. L. (1988). The coalescent process in models with selection and recombination. *Genetics* **120**, 831–840.

Hudson, R. R., Kreitman, M. & Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Hudson, R. R., Slatkin, M. & Maddison, W. P. (1992*a*). Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589.

Hudson, R. R., Boos, D. D. & Kaplan, N. L. (1992*b*). A statistical test for detecting population subdivision. *Molecular Biology and Evolution* **9**, 138–151.

Johnstone, J. S., Pepper, A. E., Hall, A. E., Chen, Z. J., Hodnett, G., Drabeck, J., Lopez, R. & Price, H. J. (2005). Evolution of genome size in Brassicaceae. *Annals of Botany* **95**, 229–235.

Kamau, E. & Charlesworth, D. (2005). Balancing selection and low recombination affect diversity near the self-incompatibility loci of the plant *Arabidopsis lyrata*. *Current Biology* **15**, 1773–1778.

Kamau, E., Charlesworth, B. & Charlesworth, D. (2007). Linkage disequilibrium and recombination rate estimates in the self-incompatibility region of *Arabidopsis lyrata*. *Genetics* **176**, 2357–2369.

Kawabe, B., Hansson, A., Forrest, A., Hagenblad, J. & Charlesworth, D. (2006). Comparative gene mapping in *Arabidopsis lyrata* chromosomes 6 and 7 and *A. thaliana* chromosome IV: evolutionary history, rearrangements and local recombination rates. *Genetical Research* **88**, 45–56.

Kuittinen, H., de Haan, A. A., Vogl, C., Oikarinen, S., Leppala, J., Koch, M., Mitchell-Olds, T. & Langley, H. & Savolainen, O. (2004). Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. *Genetics* **168**, 1575–1584.

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.

Liu, P., Sherman-Broyles, S., Nasrallah, M. E. & Nasrallah, J. B. (2007). A cryptic modifier causing transient self-incompatibility in *Arabidopsis thaliana*. *Current Biology* **17**, 734–740.

Maynard Smith, J. & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* **23**, 23–35.

Muirhead, C. A. (2001). Consequences of population structure on genes under balancing selection. *Evolution* **55**, 1532–1541.

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418–426.

Pauwels, M., Saumitou-Laprade, P., Holl, A. C., Petit, D. & Bonnin, I. (2005). Multiple origin of metallicolous populations of the pseudometallophyte *Arabidopsis halleri* (Brassicaceae) in central Europe, the cpDNA testimony. *Molecular Ecology* **14**, 4403–4414.

Payseur, B. A. & Nachman, M. W. (2002). Natural selection at linked sites in humans. *Gene* **300**, 31–42.

Purugganan, M. D. & Suddith, J. I. (1998). Molecular population genetics of the *Arabidopsis CAULIFLOWER* regulatory gene, non-neutral evolution and naturally occurring variation in floral homeotic function. *Proceedings of the National Academy of Sciences of the USA* **95**, 8130–8134.

Ramos-Onsins, S. E., Stranger, B. E., Mitchell-Olds, T. & Aguadé, M. (2004). Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. *Genetics* **166**, 373–388.

Rozas, J., Sánchez-Delbarrio, J. C, Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.

Schierup, M. H., Charlesworth, D. & Vekemans, X. (2000*a*). The effect of hitch-hiking on genes linked to a balanced polymorphism in a subdivided population. *Genetical Research* **76**, 63–73.

Schierup, M. H., Vekemans, X. & Charlesworth, D. (2000*b*). The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genetical Research* **76**, 51–62.

Shimizu, K. K., Cork, J. M., Caicedo, A. L., Mays, C. A., Moore, R. C., Olsen, K. M., Ruzsa, S., Coop, G., Bustamante, C. D., Awadalla, P. & Purugganan, M. D. (2004). Darwinian selection on a selfing locus. *Science* **306**, 2081–2084.

Stone, J. L. (2004). Sheltered load associated with S-alleles in *Solanum carolinense*. *Heredity* **92**, 335–342.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Takahata, N. (1990). A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proceedings of the National Academy of Sciences of the USA* **87**, 2419–2423.

Takahata, N. & Satta, Y. (1998). Footprints of intragenic recombination at HLA loci. *Immunogenetics* **47**, 430–441.

Takebayashi, N., Brewer, P. B., Newbigin, E. & Uyenoyama, M. K. (2003). Patterns of variation within self-incompatibility loci. *Molecular Biology and Evolution* **20**, 1778–1794.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W, improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Uyenoyama, M. K. (1997). Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics* **147**, 1389–1400.

Willems, G., Dräger, D. B., Courbot, M., Gode, C., Verbruggen, N. & Saumitou-Laprade, P. (2007). The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri* (Brassicaceae): an analysis of quantitative trait loci. *Genetics* **176**, 659–674.

Williams, C. G., Goodman, M. M. & Stuber, C. W. (1995). Comparative recombination distances among *Zea mays* L. inbreds, wide crosses and interspecific hybrids. *Genetics* **141**, 1573–1581.

Wiuf, C. K., Zhao Innan, H. & Nordborg, M. (2004). The probability and chromosomal extent of trans-specific polymorphism. *Genetics* **168**, 2363–2372.

Wright, S. (1939). The distribution of self-sterility alleles in populations. *Genetics* **24**, 538–552.

Wright, S. I. & Charlesworth, B. (2004). The HKA test revisited, a maximum-likelihood-ratio test of the standard neutral model. *Genetics* **168**, 1071–1076.

Wright, S. I., Lauga, B. & Charlesworth, D. (2003). Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Molecular Ecology* **12**, 1247–1263.

Yogeeswaran, K., Frary, A., York, T. L., Amenta, A., Lesser, A. H., Nasrallah, J. B., Tanksley, S. D. & Nasrallah, M. E. (2005). Comparative genome analyses of *Arabidopsis* spp.: inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*. *Genome Research* **15**, 505–515.