MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 1130–1142 doi: 10.1111/j.1365-294X.2011.05435.x

Disentangling the effects of breakdown of self-incompatibility and transition to selfing in North American *Arabidopsis lyrata*

ANNABELLE HAUDRY*†, HONG GUANG ZHA†‡, MARC STIFT†§¶ and BARBARA K. MABLE†
**Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada, †Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, Scotland, UK, ‡Kunming Institute of Botany, Kunming, China, §Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), Vairão, Portugal, ¶Department of Biology, University of Konstanz, Konstanz, Germany*

Abstract

A breakdown of self-incompatibility (SI) followed by a shift to selfing is commonly observed in the evolution of flowering plants. Both are expected to reduce the levels of heterozygosity and genetic diversity. However, breakdown of SI should most strongly affect the region of the SI locus (*S*-locus) because of the relaxation of balancing selection that operates on a functional *S*-locus, and a potential selective sweep. In contrast, a transition to selfing should affect the whole genome. We set out to disentangle the effects of breakdown of SI and transition to selfing on the level and distribution of genetic diversity in North American populations of *Arabidopsis lyrata*. Specifically, we compared sequence diversity of loci linked and unlinked to the *S*-locus for populations ranging from complete selfing to fully outcrossing. Regardless of linkage to the *S*-locus, heterozygosity and genetic diversity increased with population outcrossing rate. High heterozygosity of self-compatible individuals in outcrossing populations suggests that SI is not the only factor preventing the evolution of self-fertilization in those populations. There was a strong loss of diversity in selfing populations, which was more pronounced at the *S*-locus. In addition, selfing populations showed an accumulation of derived mutations at the *S*-locus. Our results provide evidence that beyond the genome-wide consequences of the population bottleneck associated with the shift to selfing, the *S*-locus of *A. lyrata* shows a specific signal either reflecting the relaxation of balancing selection or positive selection.

Keywords: balancing selection, mating system evolution, self-incompatibility, *S*-locus

Received 4 August 2011; revision received 11 November 2011; accepted 3 December 2011

Introduction

The large spectrum of mating systems among flowering plants (Richards 1997; Goodwillie *et al.* 2005) has fuelled interest in its genetic and environmental causes and evolutionary consequences. A population's mating system is a major factor determining genetic diversity and genome evolution (reviewed in Charlesworth & Wright 2001). By reducing the number of independent gametes sampled at each generation, self-fertilization directly reduces effective population size (N_e), up to 50% under complete selfing (Pollak 1987). The decrease in heterozygosity with selfing also leads to a reduction in effective recombination rate (Nordborg 2000), which in turn further reduces N_e through hitchhiking effects associated with either background selection (Charlesworth *et al.* 1993) or selective sweeps (Maynard Smith & Haigh 1974). Moreover, increased population extinction/recolonization bottlenecks in selfing plants may contribute to further reductions in N_e (Ingvarsson 2002). Reduction in N_e is expected to increase effects of genetic drift, resulting in reduced genetic diversity within selfing populations. Furthermore, predominant selfing is

Correspondence: Annabelle Haudry, Fax: +1 416 978 5878;
E-mail: annabelle.haudry@gmail.com

expected to reduce gene flow between populations and lead to increased population genetic structure (Hamrick & Godt 1989). In terms of fitness, selfing could have negative effects through inbreeding depression—the reduced fitness of offspring of related, relative to unrelated, individuals (Knight 1799).

Many plants have genetic self-incompatibility (SI) systems to prevent self-fertilization. The Brassicaceae family is characterized by a sporophytic form of SI: *S*-phenotypes are determined by the diploid *S*-genotype of the parent, with complex dominance relationships among *S*-haplotypes (Hatakeyama *et al.* 1998). Pollen and stigma *S*-phenotypes are encoded by two tightly linked genes (*SRK* and *SCR*, respectively), located in a genomic region referred to as the *S*-locus (Kusaba *et al.* 2001). Because two plants with the same *S*-phenotype are incompatible, rare *S*-phenotypes are selected for, resulting in negative frequency-dependent selection. Balancing selection maintains high heterozygosity and extremely high levels of polymorphism at the *S*-locus, which can be maintained over evolutionary time, resulting in sharing of ancestral *S*-haplotypes between species and even between genera (Castric & Vekemans 2004; Igic *et al.* 2004, 2006). To maintain tight linkage of *SRK* and *SCR*, it has been suggested that recombination was drastically reduced in the genomic region surrounding the *S*-locus (Stein *et al.* 1991). Indeed, very low recombination rates and extended linkage disequilibrium were found in this region in several SI Brassicaceae (Awadalla & Charlesworth 1999; Casselman *et al.* 2000; Charlesworth *et al.* 2003; Kamau *et al.* 2007). In line with theoretical predictions for a region of restricted recombination and strong hitchhiking, balancing selection acting on the *S*-locus has led to high polymorphism in genes closely linked to it. In Brassica (*B. rapa* and *B. oleracea*) and Arabidopsis (*A. lyrata* and *A. halleri*), synonymous diversity is elevated in a region encompassing several tens of kilobases beyond the actual targets of selection (Kamau & Charlesworth 2005; Takuno *et al.* 2007; Ruggiero *et al.* 2008).

Breakdown of SI is a recurrent transition in angiosperm evolution (Bateman 1955; Weller & Sakai 1999; Igic *et al.* 2008), and multiple independent breakdowns have been observed within genera and species (e.g. Lloyd 1965; Busch 2005a). However, a breakdown of SI does not necessarily lead to a transition from outcrossing to selfing. Self-compatible (SC) plants are still capable of outcrossing, and their outcrossed progeny are expected to have a fitness advantage relative to inbred progeny that suffer from inbreeding depression (Busch2005b). Several explanations exist for how selfing could evolve despite inbreeding depression (reviewed in Jain 1976). Reproductive assurance when mating partners or pollinators are scarce (Darwin 1876; Kalisz *et al.* 2004) and the transmission advantage of selfers over outcrossers (Fisher 1941) may counterbalance any fitness disadvantages attributed to inbreeding depression (Lloyd 1979; Charlesworth & Charlesworth 1987). This trade-off between outcrossing and selfing is influenced by the magnitude of the effect of genes influencing the selfing rate (Lande & Schemske 1985) and by purging of the genetic load in the early selfing generations (Byers & Waller 1999). However, a modifier that increases the selfing rate by a sufficiently large amount will theoretically often be favoured, even in initial populations with large inbreeding depression values (Charlesworth *et al.* 1990). Ecological conditions favouring evolution to selfing are often associated with colonization of new habitats, where mate limitation is expected (Baker 1955; Stebbins 1957; Jain 1976; Pannell & Barrett 1998).

To date, studies investigating the genetic consequences of a shift in mating system have mainly compared obligately outcrossing species with closely related selfing ones (e.g. Wright *et al.* 2002; Haudry *et al.* 2008; Foxe *et al.* 2009). The best-studied example, the model plant *Arabidopsis thaliana*, is thought to have diverged from its SI ancestor 5 My ago (Koch *et al.* 2000) and evolved to self-fertilization <1 My ago, after inactivation of the *S*-locus (Tang *et al.* 2007). A mutation in *SCR* may have been involved in the evolution of selfing in European *A. thaliana* (Tsuchimatsu *et al.* 2010). *Capsella rubella* represents an even more recent speciation event associated with the shift to selfing, estimated between 20 000 and 50 000 years ago, with a dramatic reduction in diversity across the genome (Foxe *et al.* 2009; Guo *et al.* 2009).

Interspecific comparisons may confound the effects of two separate processes: initial breakdown of the SI system and subsequent transition to selfing. Evolution to selfing might not immediately occur after breakdown of SI, but selection against selfing is no longer mediated by the *S*-locus. In SC individuals, *S*-genotypes are not expected to be under frequency-dependent selection; the breakdown of SI should therefore affect evolution at the *S*-locus by relaxing balancing selection, regardless of whether the mutation causing breakdown of SI linked or unlinked to the *S*-locus. In addition, if the mutation responsible of the breakdown of SI was linked to the *S*-locus and positively selected, a selective sweep may have cleared out the diversity of the genomic region.

Within populations of plants with a normally functioning SI response, *S*-locus genes are expected to exhibit an excess of heterozygosity ($F_{is} < 0$) and extreme allelic diversity, because of the prevention of fertilization by pollen with the same *S*-genotype (Strobeck 1980; Glémin *et al.* 2001). Breakdown of SI is thus expected to have a local effect: ultimately, heterozygosity and diversity should decrease towards values expected under the new

Table 1 Theoretical expectations about the impact of (A) balancing selection and (B) mating system on genetic properties. Balancing selection associated with SI acts locally around the *S*-locus, whereas a shift to selfing is expected to affect loci across the whole genome

		(A) Balancing selection impact	(B) Mating system impact
Genetic properties	Estimators	Loci linked to *S*-locus	Reference loci
Heterozygosity	H_o	$H_{o(SI)} > H_{o(SC)}$ with $H_{o(SI)}$ tending to 1 because of SI preventing homozygosity at the *S*-locus	$H_{o(OUT)} > H_{o(SELF)}$ reflecting increase of H_o associated with a shift to selfing
Nucleotide diversity (4 $N_e\mu$)	π (θ_w)	$\pi_{(SI)} > \pi_{(SC)}$ because of excess of diversity in SI resulting from balancing selection	$\pi_{(OUT)} > \pi_{(SELF)}$ because of negative correlation between selfing rate s and N_e, and, consequently 4 $N_e\mu$
Coefficient of inbreeding	F_{is}	$F_{is(SI)} < F_{is(SC)}$ $F_{is(SI)} < 0$ because of excess of heterozygosity induced by SI $F_{is(SC)}$ tends to 0*	$F_{is(OUT)} < F_{is\ (SELF)}$ $F_{is\ (OUT)}$ expected to be 0 * $F_{is\ (SELF)} > 0$ by definition, $F_{is} = s/(2 - s)$, with s the selfing rate

SI, self-incompatibility.
*Assuming Hardy–Weinberg equilibrium.

(selection-) mutation–drift equilibrium at the *S*-locus and closely linked loci (i.e. $F_{is} = 0$, in the case of random mating). SC individuals are therefore expected to show lower diversity and heterozygosity than SI individuals specifically in the *S*-locus genomic region (Table 1). If a subsequent shift to selfing occurs, the whole genome is expected to experience a strong population bottleneck (Charlesworth and Wright 2001), resulting in a genome-wide deficit of heterozygosity ($F_{is} > 0$, Wright 1921) and reduction in genetic diversity (Table 1). Systems with both variation in frequency of SC and outcrossing rate may offer a rare opportunity to estimate separately potential local effects (at the *S*-locus) of the breakdown of SI (relaxation of balancing selection and/or selective sweep) from the genome-wide effects of a shift to selfing.

Arabidopsis lyrata, a closely relative to the selfing model *A. thaliana*, is usually considered as obligately outcrossing owing to its strong SI system, particularly in European populations (Schierup *et al.* 2001). However, breakdown of SI has occurred within North American *A. lyrata* populations and some populations have shifted to high levels of selfing (Mable *et al.* 2005; Mable & Adam 2007). Previous work (Foxe *et al.* 2010) suggested that the shift to selfing occurred independently in different genetic backgrounds but whether this was attributed to independent losses of SI could not be established because the causal mutation(s) for breakdown of SI have yet to be determined. Although predominantly selfing populations showed a high proportion of SC individuals and predominantly outcrossing populations showed a high proportion of SI individuals, most populations showed a mixture of SI and SC. Whereas SC plants from selfing populations showed a substantial loss of heterozygosity compared to SI plants from the same populations, SC plants from outcrossing populations did not, suggesting that they are mainly outcrossing (Foxe *et al.* 2010).

Here, we aimed to (i) estimate separately effects of breakdown of SI and transition to selfing on genetic diversity and (ii) identify which evolutionary forces have shaped patterns of diversity at the *S*-locus during these transitions. We compared two genes flanking the *S*-locus to 18 reference loci scattered throughout the genome in populations with different phenotypes (SI vs. SC) and outcrossing rates. The genome-wide pattern of diversity is expected to reflect the impact of neutral bottlenecks (e.g. as a result of postglaciation recolonization and transition to selfing). In addition, *S*-linked genes might reflect changes in selection patterns associated with the breakdown of SI. For all loci, diversity increased with population outcrossing rate; however, the increase was more pronounced at *S*-linked loci. Moreover, selfing was associated with a shift in the derived allele frequency spectrum of B80 that could not be explained by a bottleneck alone. Together, our results highlight the specific impact of the breakdown of SI, in addition to the overall loss of diversity associated with the transition to selfing.

Material and methods

Target genes

We analysed two genes, located on either side of the *S*-locus, to evaluate diversity of this region. Both genes have been found to be tightly linked to *SRK* in *A. hal-*

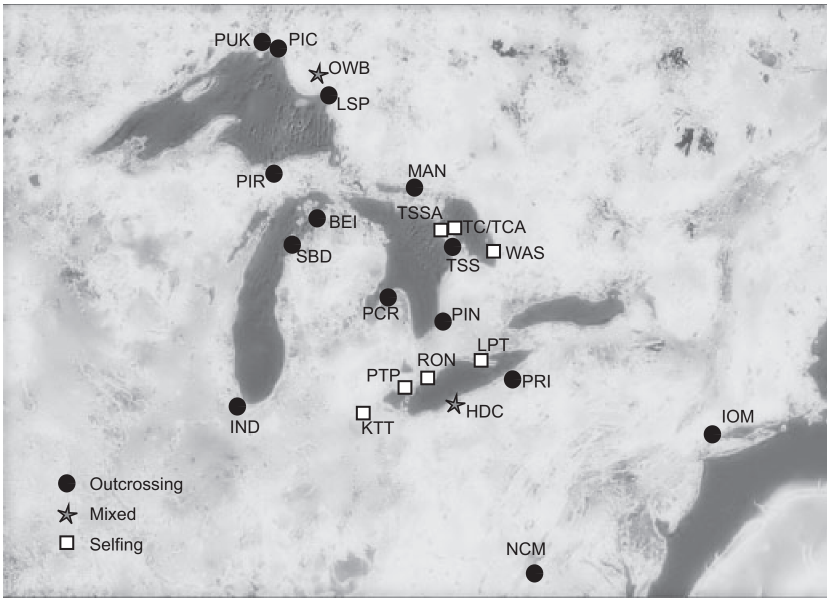

Fig. 1 Spatial distribution of 24 *Arabidopsis l. lyrata* populations sampled from the Great Lakes region of Eastern North America. Six populations were selfing (white squares), two populations intermediate (grey stars) and 14 populations were predominantly outcrossing (black circles). See Methods for more details on mating system determination and Foxe *et al.* (2010) for a complete description of localities.

leri (Ruggiero *et al.* 2008), *Capsella grandiflora* (Guo *et al.* 2009) and in European populations of *A. lyrata* (Kamau & Charlesworth 2005). *B80* and *B160* are therefore expected to evolve under balancing selection because of hitchhiking with the *S*-locus in the SI individuals. *B80* (homologous to *A. thaliana* At4g21350) flanks the *S*-locus at the *SCR* side and is located between 66 and 41 kb from *SRK* (in *A. lyrata* haplotypes *Sa* and *Sb*, respectively; Kusaba *et al.* 2001). It encodes a U-box/ARM repeat protein involved in the pathway leading to cellular death required for SI (Liu *et al.* 2007). *B160* (homologous to *A. thaliana* At4g21430) flanks the *S*-locus on the *SRK* side and is located between 31 and 53 kb from *SRK* (in *A. lyrata* haplotypes *Sa* and *Sb*, respectively; Kusaba *et al.* 2001). It encodes a transcription factor. Both genes can be genotyped using direct sequencing, unlike the highly polymorphic *SRK* and *SCR* genes, for which insertions or deletions among alleles make resolution of heterozygotes difficult.

Plant materials and mating system determination

Eight individuals were sampled from 24 populations of *A. lyrata* located in the Great Lakes region of eastern North America (*A. l. lyrata*; Fig. 1, Foxe *et al.* 2010 and Appendix S1, Supporting information). Divergence between North American and European populations suggests that colonization of America occurred before the end of the Wisconsin glaciations (Ross-Ibarra *et al.* 2008). Plants would have colonized the study area more recently, probably experiencing a founder effect, as suggested in Foxe *et al.* (2010). The outcrossing rate (T_m) of each population had been determined previously based on progeny arrays and ranged from 0.09 to 0.99 (Foxe *et al.* 2010). These estimated outcrossing rates reflect realized outcrossing rates, that is, selfing events that did not produce offspring (because of strong inbreeding depression) are not taken into account. Conventionally, populations producing more than 50% of their progeny through selfing are considered selfing, while the others are considered outcrossing. The distribution of the outcrossing rates estimated in our populations was mainly bimodal (Fig. S1, Supporting information), with only two populations deviating from this pattern. These two populations had intermediate outcrossing rates and an intermediate ratio of SI/SC individuals. Otherwise, highly selfing populations are nearly exclusively consisted of SC individuals, and highly outcrossing populations nearly exclusively consisted of SI individuals (Foxe *et al.* 2010). We therefore considered these two populations separately. On this basis, we defined eight selfing ($T_m < 0.50$; Table 2), 14 outcrossing ($T_m > 0.75$; Table 2) and two 'mixed' ($0.50 < T_m < 0.75$; Table 2) populations.

Amplification and sequencing

PCR amplifications were performed using a DNA Engine PTC-200 thermocycler (MJ Research, Watertown, MA) under the following conditions: 94°C for 3 min; 33 cycles of 94°C for 30 s, 53°C for 30 s and 72°C for 60 s; and a final 10-min extension at 72°C. Primers were previously described in Kamau & Charlesworth (2005). PCR products were purified using QIAquick Gel Extraction Kits (Qiagen, Inc.).

Table 2 Population genetic diversity statistics for the *S*-linked *B160* and *B80* based on samples of eight individuals from populations classified as selfing, mixed or outcrossing based on their outcrossing rates (T_m): number of polymorphic sites (*S*); number of haplotypes (*h*); observed heterozygosity (H_o); Watterson's diversity (θ_W); synonymous nucleotide diversity (π_S); coefficient of inbreeding (F_{is}); Tajima's *D* (D_{Taj})

			B160 (1061 bp alignment)							*B80* (648 bp alignment)						
Population	Mating system	T_m	*S*	*h*	H_o	θ_W	π_S	F_{is}	$D_{Taj}^{\S}$	*S*	*h*	H_o	θ_W	π_S	F_{is}	$D_{Taj}^{\S}$
PTP	Selfing	$0.09^{\dagger}$	0	1	0.00	0.000	0.000	NA	NA	0	1	0.00	0.000	0.000	NA	NA
LPT	Selfing	$0.13^{\dagger}$	7	3	0.13	0.002	0.003	0.448	−2.06*	26	2	0.13	0.012	0.013	−0.11	−2.40***
TC	Selfing	$0.18^{\dagger}$	0	1	0.00	0.000	0.000	NA	NA	27	2	0.13	0.013	0.015	−0.11	−2.41***
WAS	Selfing	$0.25^{\dagger}$	6	5	0.50	0.002	0.007	0.123	0.66	21	2	0.50	0.010	0.036	−0.33	1.32
RON	Selfing	$0.28^{\dagger}$	3	2	0.00	0.001	0.001	1.000	−0.65	0	1	0.00	0.000	0.000	NA	NA
KTT	Selfing	$0.31^{\dagger}$	0	1	0.00	0.000	0.000	NA	NA	0	1	0.00	0.000	0.000	NA	NA
TSSA	Selfing	$0.41^{\dagger}$	5	3	0.25	0.001	0.002	−0.103	−1.93*	40	5	0.50	0.019	0.075	0.33	1.54
TCA	Selfing	$0.48^{\dagger}$	0	1	0.00	0.000	0.000	NA	NA	0	1	0.00	0.000	0.000	NA	NA
OWB	Mixed	$0.64^{\dagger}$	0	1	0.00	0.000	0.000	NA	NA	30	3	0.63	0.014	0.054	−0.16	2.30*
HDC	Mixed	$0.65^{\dagger}$	6	5	0.13	0.002	0.006	0.781	0.52	34	6	0.38	0.016	0.029	0.23	−1.44
NCM	Outcrossing	NA‡	1	2	0.38	0.000	0.003	0.238	1.47	34	5	0.63	0.016	0.053	0.04	0.77
PIC	Outcrossing	$0.77^{\dagger}$	2	3	0.63	0.001	0.000	−0.356	−0.08	28	7	0.75	0.013	0.055	−0.07	1.3654
MAN	Outcrossing	$0.83^{\dagger}$	8	3	0.25	0.002	0.001	−0.103	−1.46	37	6	0.75	0.017	0.057	−0.17	0.46
PIN	Outcrossing	$0.84^{\dagger}$	2	2	0.13	0.001	0.000	−0.067	−1.50	41	9	1.00	0.019	0.063	−0.17	0.82
PIR	Outcrossing	$0.88^{\dagger}$	9	7	0.75	0.003	0.002	−0.011	−0.80	56	8	0.75	0.026	0.066	0.01	−0.74
PRI	Outcrossing	$0.89^{\dagger}$	7	3	0.25	0.002	0.004	0.385	−0.80	21	4	0.88	0.010	0.023	−0.38	−0.83
TSS	Outcrossing	$0.91^{\dagger}$	6	3	0.25	0.002	0.002	−0.103	−2.00*	51	11	0.88	0.024	0.078	−0.01	0.50
IOM	Outcrossing	$0.94^{\dagger}$	8	3	0.14	0.002	0.010	0.736	1.98	46	8	0.63	0.021	0.077	0.21	0.76
SBD	Outcrossing	$0.94^{\dagger}$	9	4	0.63	0.003	0.005	−0.290	−0.52	54	9	0.88	0.025	0.085	−0.05	0.50
LSP	Outcrossing	$0.96^{\dagger}$	2	2	0.38	0.001	0.002	−0.231	0.20	33	4	0.63	0.015	0.038	−0.23	−0.63
PUK	Outcrossing	$0.96^{\dagger}$	9	5	0.75	0.003	0.007	−0.011	1.22	46	6	0.88	0.021	0.083	−0.09	1.24
BEI	Outcrossing	$0.98^{\dagger}$	7	4	0.75	0.002	0.003	−0.333	−0.80	53	7	0.63	0.025	0.047	−0.11	−1.50
PCR	Outcrossing	$0.98^{\dagger}$	2	2	0.13	0.001	0.001	−0.067	−1.50	40	6	0.75	0.019	0.054	−0.05	−0.02
IND	Outcrossing	$0.99^{\dagger}$	1	2	0.88	0.000	0.000	−0.778	1.47	44	13	0.88	0.020	0.081	−0.01	1.19

NA indicates when data were not available.

†Based on multilocus estimates obtained from microsatellite variation in progeny arrays, from Y. Willi (IOM population, personal communication) and from Foxe *et al.* (2010) for the other populations.

‡Not enough seeds were available to establish outcrossing rates; only one individual was partially self-compatible (SC); population was therefore assumed to be mostly outcrossing (Foxe *et al.* 2010).

§Significance: *$P < 0.5$; ***$P < 0.1$; ***$P < 0.001$

Direct sequencing was performed on both strands using ABI PRISM 3730 sequencers (the Sequencing Service at the University of Dundee and the Genepool at the University of Edinburgh) using the same primers as for the PCR amplifications. The sequences were verified manually and corrected using Sequencher 4.8 (Gene Codes, Ann Arbor, MI). Ambiguous genotypes were confirmed by cloning, using TA cloning kits for sequencing (Invitrogen Ltd., Paisley, UK). Details and the strategy used to resolve the phase for heterozygotes are outlined in the Appendix S1 (Supporting information). Sequences were aligned and further adjusted manually using Bioedit (Hall 1999). Intron–exon boundaries were determined based on alignment with the annotated sequences of *A. halleri* orthologs, obtained from GenBank (EU274019 and EU274126, Ruggiero *et al.* 2008).

Diversity and heterozygosity

To estimate population diversity statistics, sequences were analysed using custom C++ programs based on Bio++ libraries (Dutheil *et al.* 2006). For *B80* and *B160*, we calculated number of haplotypes, synonymous and nonsynonymous nucleotide diversity (π_S and π_N; Tajima 1983) and Watterson's statistics θ_w (Watterson 1975), both estimated per site, Tajima's *D* (Tajima 1989). Significance of within-population mean Tajima's *D* was determined using a two-tailed test, assuming that this statistic follows a beta distribution (Tajima 1989) as implemented in DnaSP (Librado & Rozas 2009). Observed and expected heterozygosities (H_o and H_e, respectively) in each population were estimated based on haplotypes, using an R script. An individual was heterozygous if it carried two different haplotypes at the considered locus;

H_o was computed as the average number of heterozygotes in the population. H_o and H_e were used to calculate the inbreeding coefficient, $F_{is} = 1 - H_o/H_e$. Scripts are available upon request to AH.

*Changes in selection patterns at the S-*locus

To investigate whether changes in the selective pressure acting at the *S*-locus could be detected in selfing populations compared to outcrossing ones, we studied derived allele frequency spectra (DAFS). Because neutral markers suggested strong structure among our sampled populations (Foxe *et al.* 2010), DAFS were analysed separately for each population. Polymorphic SNP frequency was computed per chromosome and therefore can vary from 1 to 15 in samples of eight diploid individuals per population. To detect potential changes in selective patterns at the *S*-linked *B80* gene, DAFS expected under neutrality (defined using the model developed in Kern & Haussler 2010) were compared to observed DAFS for each population using a Kolmogorov–Smirnov test. We also compared the number of derived SNPs observed in selfing and outcrossing populations for three classes of frequency (1–5, 6–10, and 11–15 derived SNPs, respectively). Because of their intermediate status, the populations with a mixed mating system were excluded from this analysis.

Trans-specific polymorphisms could make inferences on the ancestral state of mutations spurious. Shared polymorphisms at *B80* have been described between *A. thaliana* and *A. lyrata* (Charlesworth *et al.* 2006), so we used the most distantly related species for which flanking gene sequence data were available (*C. grandiflora*) to infer ancestral states (accession numbers FJ650027–FJ650040, Guo *et al.* 2009). Sites that were polymorphic within *C. grandiflora* were not considered, as they are uninformative with respect to the ancestral state.

Balancing selection vs. mating system

To test whether breakdown of SI affected heterozygosity at the *S*-locus independently of the effects of the shift to selfing, we compared H_o both at *S*-linked and unlinked genes of: (i) SI individuals from outcrossing populations; (ii) SC individuals from outcrossing populations; and (iii) SC individuals from selfing populations. Comparing (i) and (ii) shows the effect of the breakdown of SI within outcrossing populations. The comparison between (ii) and (iii) indicates the effect of the shift to selfing on H_o. Differences in H_o were tested by Wilcoxon rank sum tests in R (R Development Core Team 2010).

To investigate whether we could detect a further reduction in diversity beyond the direct reduction expected as a result of increased homozygosity at the *S*-locus as a consequence of selfing alone, we calculated corrected values of the genetic diversity, as proposed in Nordborg (2000): $\pi_{Scorr} = \pi_{Sobs}*(1 + F)$, where π_S is the nucleotide diversity (Tajima 1983) estimated on silent sites and F the inbreeding coefficient defined as $F = s/(2 - s)$. As an estimation of the selfing rate, we used $1 - T_m$. We then tested for an effect of the outcrossing rate on π_{Scorr} by performing a linear regression.

Because no outcrossing population was composed only of SC individuals, it was impossible to test the theoretical predictions associated with relaxation of balancing selection listed in Table 1 *per se*. As an approximation, we compared heterozygosity and diversity in selfing populations (composed of SC individuals) compared to outcrossing ones (essentially composed of SI individuals). As North American populations of *A. lyrata* showed typical departures from a standard neutral model expected after a recent bottleneck (Foxe *et al.* 2010), we cannot model neutral expectations easily. As the situation is comparable to domesticated plants that experienced a severe population bottleneck <10 000 years ago, we employed a similar approach (Wright *et al.* 2005; Haudry *et al.* 2007). We estimated the relative reduction in diversity for reference genes (expected to be neutral) and for *S*-linked genes to detect the specific impact of selection. Here, whereas reference loci are expected to only be affected by the shift to selfing, *S*-loci should be affected both by shift to selfing and breakdown of SI.

To separately test for the effects of the outcrossing rate (T_m) and the gene class (linked or unlinked to the *S*-locus) on the diversity variables (H_o, π_S and θ_W), we used a generalized linear model (GLM). Two-way nonparametric analysis of variance Scheirer-Ray-Hare rank tests were also performed to ensure that violations of assumptions of the GLM did not bias conclusions. All statistical analyses were performed using the R software (R Development Core Team 2010) and the additional package lmtest. The loss of heterozygosity (L_{Ho}) experienced in selfing populations was estimated using the formula $L_{Ho} = 1 - H_{o(in)}/H_{o(out)}$, with $H_{o(in)}$ and $H_{o(out)}$ representing H_o estimated in selfing and outcrossing populations, respectively. Similarly, we estimated loss of $\theta_W(L_\theta)$ and $\pi_S(L_\pi)$. We used Kruskal and Wallis's tests to assess whether L values were different between selfing and outcrossing populations, both for *S*-linked and unlinked loci.

Results

*Diversity and heterozygosity at the S-*locus

As expected for *S*-linked genes, levels of diversity were extremely high for *B80* (Table 2). Heterozygosity was about three times higher at *B80* compared to reference loci. In contrast, at *B160*, levels of nucleotide diversity were low even in outcrossing populations ($\pi_S = 0.003$) and no significant reduction in selfing compared to outcrossing populations ($\pi_S = 0.001$). This suggests that *B160* is not linked to the *S*-locus in *A. lyrata* and essentially behaves as any random locus. Accordingly, at *B160*, only H_o increased significantly with increasing outcrossing rate T_m (Fig. S3, Supporting information), whereas at *B80* significant linear regressions indicated that H_o, π_S and θ_W increased with T_m (Figs 2 and S3, Supporting information). When we corrected the nucleotide diversity for selfing rate, T_m still explained a significant proportion of the variation in π_{Scorr} ($R^2 = 32.1\%$, $P = 0.0029$). Thus, direct reduction in N_e because of selfing rate is not sufficient to explain differences in nucleotide diversity observed between selfing and outcrossing populations for *B80*.

Breakdown of self-incompatibility alone does not affect heterozygosity

At the individual level, we compared H_o for the *S*-linked *B80* and reference loci between (i) SI individuals from outcrossing populations ($n = 111$); (ii) SC individuals from outcrossing populations ($n = 6$); and (iii) SC individuals from selfing populations ($n = 55$). At *B80*, in outcrossing populations, SI individuals did not exhibit higher H_o than SC individuals (0.72 and 0.83, respectively, (i) vs. (ii): $W = 370.5$, $P = 0.55$). In SC individuals from selfing populations, H_o was significantly reduced to 0.15 [(iii) vs. (i): $W = 51.5$, $P = 1.14 \times 10^{-4}$; (iii) vs. (ii): $W = 1296.5$, $P = 3.24 \times 10^{-12}$]. At the reference loci, patterns were similar. In outcrossing populations, SI individuals did not exhibit higher H_o than SC individuals (0.21 and 0.23, respectively; (i) vs. (ii): $W = 378.5$, $P = 0.58$). In SC individuals from selfing populations, H_o was significantly reduced to 0.05 compared to individuals from outcrossing populations [(iii) vs. (i): $W = 26.5$, $P = 0.0052$; (iii) vs. (ii): and W = 365, $P = 2.0 \times 10^{-11}$].

Changes in selection patterns at the S-locus

As *B160* appeared to be unlinked to the *S*-locus, we deemed it uninformative to detect potential changes in selection at the *S*-locus after the breakdown of SI and transition to selfing. Moreover, *B160* showed only 15 derived polymorphic sites within the pool of all populations, and only three of these were polymorphic in selfing populations. At *B80*, we found 60 derived alleles segregating within the pool of all populations. The DAFS observed at *B80* showed a tendency for an excess of derived alleles at intermediate and high frequency (Fig. S2, Supporting information). However, only four out of 14 outcrossing populations and all four selfing populations with polymorphism showed a significantly

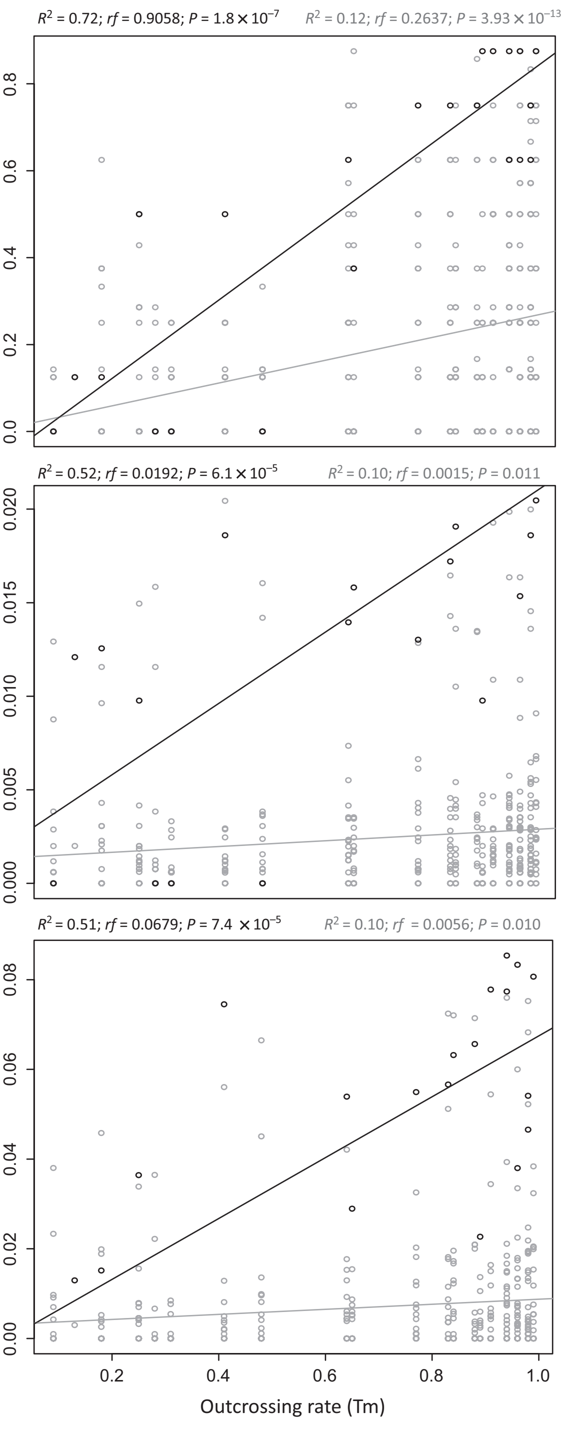

Fig. 2 Linear regressions between genetic diversity statistics (H_o, θ_W and π_S) and outcrossing rate T_m for *B80* (black) and reference loci (grey). Solid lines represent significant linear regressions ($P < 0.05$). R^2, regression coefficient (*rf*) and *P*-value (*P*) are given on the top of each graph (in black for *B80* and grey for the reference loci).

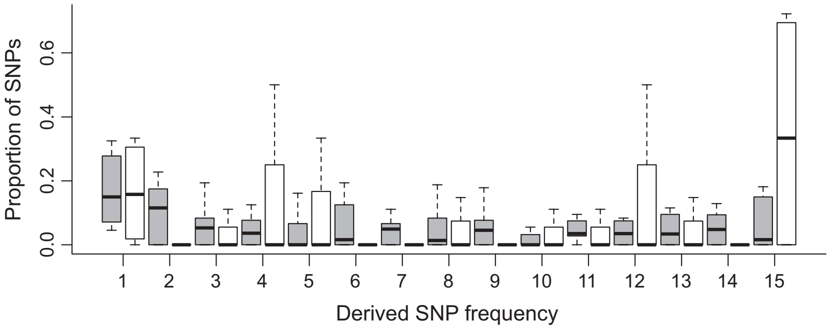

Fig. 3 Derived allele frequency spectra at *B80* for outcrossing (grey bars) and selfing (white bars) populations.

different DAFS from neutral expectations ($P < 0.05$, Fig. S2, Supporting information). The other four selfing populations were not polymorphic so the DAFS could not be inspected. The comparison of DAFS compiled per mating system (Fig. 3) showed that the selfing and the outcrossing DAFS differed significantly ($D = 0.6$, $P = 0.009$). We found that the repartition of the derived SNPs in low-, intermediate- and high-frequency classes differed between selfing and outcrossing populations ($\chi^2 = 14.88$; df = 2; $P = 6 \times 10^{-4}$).

Effects of S-locus linkage and mating system on diversity

H_o, π_S and θ_W increased significantly with T_m both for *B80* and reference genes, but the relative loss was stronger and diversity was generally higher at *B80* compared to the rest of the genome (Fig. 2; Tables 3 and S1, Supporting information). The loss of H_o associated with selfing was very large regardless of gene class: 70–79% (Table 3). The relative loss of nucleotide diversity (π_S and θ_W) was most pronounced for *B80* ($L_\pi = 72\%$ and $L_\theta = 68\%$), more moderate ($L_\pi = 38\%$ and $L_\theta = 33\%$) for the unlinked loci and nonsignificant for *B160*. Contrary to what was found for *B80* and *B160*, F_{is} for the unlinked loci increased significantly in selfing populations (Table 3). GLM showed significant effects of the outcrossing rate (T_m), the gene class (linked—*B80*; unlinked to the *S*-locus—18 reference loci) and their interaction on the diversity variables H_o, π_S and θ_W (Fig. 2 and Table 3). Slopes of linear regression of each gene set were found to be significantly different, supporting evidence for interaction between T_m with the gene class (H_o: $P = 6.41 \times 10^{-6}$, π_S: $P = 2.54 \times 10^{-10}$, θ_W: $P = 9.83 \times 10^{-11}$). Alternative nonparametric Scheirer-Ray-Hare rank tests gave similar results (data not shown). Coalescent simulations were used to ensure that any greater reduction in diversity for *S*-linked genes was not merely due to the extremely high diversity in populations with a functional SI system (Appendix S1, Supporting information).

Discussion

We showed that heterozygosity and genetic diversity of both *S*-linked and reference genes are reduced in populations that have undergone a transition to selfing. However, the relative reduction in diversity was more

Table 3 Comparison between heterozygosity, nucleotide diversity and the inbreeding coefficient estimated in outcrossing (OUT) and selfing (SELF) populations at the *S*-linked *B80* and reference loci

		Balancing selection impact + mating system impact			Mating system impact		
		B80 (*S*-linked)			Reference loci (unlinked to the *S*-locus)†		
Genetic properties	Estimators	OUT	SELF	$L^‡$	OUT	SELF	$L^‡$
Heterozygosity	H_o	0.74	0.16	79%***	0.23	0.07	70%***
Nucleotide diversity (4 N_e μ)	π_s	0.061	0.017	72%***	0.008	0.005	38%***
	θ_w	0.019	0.007	63%***	0.003	0.002	33%***
Coefficient of inbreeding	F_{is}	−0.08	−0.05	−21% ns	0.26	0.45	−72%*

†Data estimated based on 18 unlinked loci (Foxe *et al*. 2010).

‡Estimation of the relative reduction in the parameters in selfing compared to outcrossing populations (see Methods for details).

pronounced at the *S*-linked gene *B80*. This illustrates that changes in the selection regime operating at the *S*-locus—relaxation of balancing selection and potentially positive selection—have left a detectable signature beyond the genome-wide effect associated with the shift to selfing (Table 1). Two scenarios may explain such a signature: (i) a delay between the breakdown of SI (affecting the *S*-locus region) and the following transition to selfing (affecting the whole genome); (ii) a selective sweep of a mutation linked to the *S*-locus beneficial in a selfing background.

B80 is, but B160 is not affected by balancing selection

Balancing selection acting at a particular locus affects the surrounding genome region in maintaining high polymorphism (Charlesworth 2006). In SI species of both *Brassica* and *Arabidopsis*, diversity at neutral sites is elevated in a region encompassing several tens of kilobases beyond the actual targets of selection because of reduced recombination at the *S*-locus (Kamau & Charlesworth 2005; Takuno *et al.* 2007; Ruggiero *et al.* 2008). Because *B160* is physically linked to the *S*-locus (Kamau & Charlesworth 2005), it was expected to evolve under balancing selection because of hitchhiking with the *S*-locus. Unexpectedly, genetic diversity at *B160* was very low in every populations of *A. lyrata*. Silent nucleotide diversity observed at *B160* in outcrossing populations of *A. l. lyrata* ($\pi_S = 0.003$) and *A. l. petraea* ($\pi_S = 0.011$; Kamau & Charlesworth 2005 and Kamau *et al.* 2007) was approximately half that estimated across 77 randomly chosen nuclear loci ($\pi_S = 0.006$ and 0.021 for *A. l. lyrata* and German *A. l. petraea*, respectively; Ross-Ibarra *et al.* 2008). This suggests that *B160* is not genetically linked to the *S*-locus and therefore balancing selection has not increased its diversity. In contrast, *B80* followed the expected pattern for genes evolving under balancing selection, with an extremely high level of nucleotide diversity in outcrossing populations ($\pi_S = 0.06$ and 0.09 for *A. l. lyrata* and *A. l. petraea*, respectively). We consequently consider *B80* to better indicate the evolutionary forces acting at the *S*-locus and focused on its analysis to investigate the genetic impact associated with the breakdown of SI.

Major impact of the transition to selfing

Reduction in both heterozygosity and genetic diversity are expected at the *S*-locus in association with (i) breakdown of SI and (ii) a shift to selfing (Table 1). Our results showed that mating system strongly affected heterozygosity of the sampled populations at the *S*-linked *B80*. H_o tended to 0 in selfing populations, whereas it approached 1 in outcrossing populations. In sporophytic SI typical of the Brassicaceae there is a dominance hierarchy between *S*-haplotypes, and the *S*-locus can be homozygous for the most recessive haplotypes. Still, SI is expected to result in an excess of heterozygotes at the *S*-locus compared to Hardy–Weinberg expectations, because the more dominant *S*-haplotypes cannot normally occur in homozygous. Also in agreement with expectations, selfing populations showed lower genetic diversity compared to outcrossing populations (Table 3). We found that population outcrossing rate was a good predictor of heterozygosity and genetic diversity observed within *A. l. lyrata*.

Changes in selection patterns at the S-locus

Inbreeding coefficients can be affected by both mating system and selection (Wright 1939). In SI populations, negative values of F_{is} are expected at the *S*-locus and at closely linked loci (Ohta & Kimura 1970). In agreement with these expectations, we found slightly negative F_{is} in outcrossing populations at *B80* (−0.08 on average). In contrast, a deficit of heterozygotes is theoretically predicted for selfing populations (Wright 1921). Although F_{is} could not be estimated because of a lack of diversity in four of the selfing populations, intriguingly, *B80* showed no such heterozygous deficit in the four remaining populations. Because negative frequency-dependent selection has likely shaped diversity at the *S*-locus, extremely high levels of allelic polymorphism are expected in this genomic region (Emerson 1939; Wright 1939). A large number of alleles are thus expected to segregate at low/intermediate frequencies in a population. Our results may suggest that, despite high selfing rates, occasional outcrossing events are sufficient to restore observed heterozygosity in accordance with Hardy–Weinberg equilibrium. Alternatively, time since the transition to selfing might have been insufficient to erode the excess of heterozygosity associated with functional SI systems in the outcrossing ancestral population to the deficit typically associated with selfing. Alternatively, *B80* may have a function other than its predicted role as a modifier of SI and still be under some form of balancing selection (Liu *et al.* 2007).

Derived allele frequency spectra reflect the evolution of new mutations arising in populations and can help to indicate the selection pressures acting on them (Nielsen 2005); they can be used to infer complex demographic models (Pool *et al.* 2010). At equilibrium under neutrality, DAFS in a large population is expected to have a large majority of segregating derived alleles at low frequencies, and only a few at high frequencies. Deviations from such an allele frequency distribution (Tajima 1989) can provide insights into population histories of selection and/or demographic changes (Bustamante *et al.* 2001; Caicedo *et al.* 2007; Keinan *et al.* 2007). Balancing selection increases the number of derived variants segregating at intermediate frequencies, and positive selection gives an excess of high-frequency-derived mutations (Nielsen 2005).

We observed an excess of high-frequency-derived mutations for both outcrossing and selfing populations at *B80* (Fig. 3). Misidentification of ancestral states owing to the fixation of the derived allele in the outgroup or occurrence of multiple mutations since divergence could have caused this pattern. The increased frequency of derived mutations at *S*-linked loci in selfing populations could be due to positive selection at a linked mutation during shift to selfing and/or greater drift induced by selfing. That is, it may reflect the action of drift instead of balancing selection acting on the *S*-locus after breakdown of the SI system. This should lead to an increased frequency and fixation of the derived haplotypes that were at the highest frequency before the breakdown. Alternatively, it may reflect the shift to selfing, which is expected to cause a strong reduction in effective population size. This should affect not only the *S*-locus, but the entire genome. A recent founder effect associated with a shift to selfing is expected to reduce the proportion of low frequency variants and produce positive Tajima's *D* values as a starting state. Diversity then could recover through mutation, producing low frequency variants and negative Tajima's *D*. Alternatively, gene flow could affect Tajima's *D* in either direction, depending on how often it occurs. Selfing populations of *A. l. lyrata* did not show a signature of a recent bottleneck and were characterized by zero or negative values of Tajima's *D* at *B80* (Table 2), reflecting a relative over-representation of rare alleles segregating. This could be explained by strong population expansion owing to the success of the selfing strategy after the initial bottleneck associated with selfing. Alternatively, over-abundance of high-frequency-derived alleles is often attributed to recurrent positive selection (Nielsen 2005; Wu *et al.* 2010), which could have led to the rise of new mutations in the populations. Although the relatively high levels of nucleotide diversity at *B80* compared to reference in the selfing populations appear to argue against a selective sweep scenario, we cannot exclude this scenario with our data.

Disentangling breakdown of self-incompatibility and shift to selfing

Selfing populations of *A. l. lyrata* experienced a breakdown of the SI system and then a transition to selfing. Both processes are expected to lead to a reduction in genetic diversity; the former should locally affect the *S*-locus whereas the latter should impact the entire genome (Table 1). Our results suggest that the breakdown of SI in SC individuals from outcrossing populations has not (yet) had a significant impact on observed heterozygosity in the genomic region around the *S*-locus in *A. l. lyrata*. Although only six individuals were available to test, the high levels of heterozygosity suggests that they arose from outcrossing events and are not reproductively isolated from the SI individuals in the population. If the breakdown of SI acted as a major modifier of the selfing rate, it could theoretically spread despite the existence of inbreeding depression (Lande & Schemske 1985; Charlesworth *et al.* 1990). Our results and the presence of SC individuals throughout the entire geographic region sampled but a shift to selfing in only a few (Mable *et al.* 2005) may suggest a stepwise evolution of mating system involving a breakdown of SI (affecting the *S*-locus region) followed by a later shift to selfing (affecting the whole genome).

Although transition to selfing had a major impact, it is not sufficient to fully explain the reduction in genetic diversity observed at *B80* in selfing populations. When levels of π_S were corrected for differences in N_e because of selfing, T_m was still a good indicator of diversity π_{Scorr} at *B80*. Thus, a further reduction in N_e affected the selfing populations beyond strictly neutral predictions, in contrast with what was found for unlinked loci in the same populations (Foxe *et al.* 2010) or in *Eichhornia paniculata* (Ness *et al.* 2010). As unlinked loci did not show any evidence of a demographic bottleneck associated with selfing in these species, the signal detected for *B80* in this study likely reflects the local impact of the breakdown of SI on the *S*-locus. Selfing populations exhibited a more drastic relative loss of diversity at *B80* ($L = 72\%$) than at unlinked loci ($L = 38\%$) and *B160* ($L = 33\%$; Table 3). Consistently, GLM analyses showed that reductions in H_o, π_S and θ_W with increasing selfing rate were of greater magnitude for *B80* than for reference loci. Altogether, our results demonstrate that the *S*-locus, or at least the *S*-linked *B80*, experienced a stronger bottleneck than the rest of the genome during the transition to selfing in *A. l. lyrata*.

Balancing selection at the *S*-locus has been shown to maintain a high level of diversity in linked genes (Charlesworth 2006). After breakdown of SI and relaxation of balancing selection, diversity of *S*-linked genes is expected to decrease because of the fixation of variants by drift. The stronger bottleneck observed at *B80* compared to the rest of the genome could simply be the signature of the vanishing of excessively high diversity of the locus, because of the relaxation of balancing selection. Another possible explanation might be that positive selection (a selective sweep) drove the fixation of advantageous alleles associated with SC. This could have involved the allele directly responsible for breakdown of SI or a linked allele that conferred an advantage in a selfing background. There are many steps in the SI pathway where a mutation could result in breakdown of SI (Mable 2008). However, recent studies have suggested that mutations at the *S*-locus played a role in the breakdown of SI in *A. thaliana* (Boggs *et al.* 2009). An inversion in the male component *SCR* has recently been suggested to have disrupted SI in *A. thaliana* (Tsuchimatsu *et al.* 2010), and *B80* itself has previously been proposed to be a modifier responsible for cryptic variation of SI in selfing *A. thaliana* (Liu *et al.* 2007) but whether similar factors operate in *A. lyrata* remains to be tested.

Our study is the first to our knowledge that explicitly considered separately the breakdown of SI and the transition to selfing. This is a critical first step to understand the dynamics of the evolution of selfing and towards estimating and comparing the exact timing of these two events for each of the selfing populations. We conclude here that the genetic diversity of the *S*-linked gene *B80* was reduced partly owing to the relaxation of balancing selection and/or a selective sweep (related to breakdown of SI) and partly owing to reductions in effective population size (related to transition to selfing).

Acknowledgements

We thank Peter Hoebe, Andy Tedder and Aileen Adam for assistance with DNA extractions and Yvonne Willi and David Remington for providing seeds, Paul Foxe for sharing sequence data of reference loci, and Julien Dutheil and Hawthorne Beyer for advice on programming. We greatly thank Deborah Charlesworth, Alan Moses, Mathieu Siol, Stephen Wright and two anonymous reviewers for discussions and constructive comments on the manuscript. This work was supported by UK research grants to AH (Biotechnology and Biological Sciences Research Council international exchange travel grant ISIS 2078) and BKM (BBSRC/European Research Association Plant Genomics BB/L024734/1, and Natural Environment Research Council, NE/D013461/1), and a fellowship from the Portuguese Fundação de Ciência e a Tecnologia, co-financed by the Social Fund of the European Union through Programa Operacional Potencial Humano/Quadro de Referência Estratégico Nacional (SFRH/BPD/64400/2009) to MS.

References

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

Baker HG (1955) Self-compatibility and establishment after 'long-distance' dispersal. *Evolution*, **9**, 347–349.

Bateman AJ (1955) Self-incompatibility systems in angiosperms. III. Cruciferae. *Heredity*, **9**, 5368.

Boggs NA, Nasrallah JB, Nasrallah ME (2009) Independent *S*-locus mutations caused self-fertility in *Arabidopsis thaliana*. *PLoS Genetics*, **5**, e1000426.

Busch JW (2005a) The evolution of self-compatibility in geographically peripheral populations of *Leavenworthiaalabamica* (Brassicaceae). *American Journal of Botany*, **92**, 1503–1512.

Busch JW (2005b) Inbreeding depression in self-incompatible and self-compatible populations of *Leavenworthiaalabamica*. *Heredity*, **94**, 159–165.

Bustamante CD, Wakeley J, Sawyer S, Hartl DL (2001) Directional selection and the site-frequency spectrum. *Genetics*, **159**, 1779–1788.

Byers DL, Waller DM (1999) Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. *Annual Review of Ecology, Evolution, and Systematics*, **30**, 479–513.

Caicedo AL, Williamson SH, Hernandez RD *et al.* (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics*, **3**, e163.

Casselman AL, Vrebalov J, Conner JA *et al.* (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.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, **2**, e64.

Charlesworth D, Charlesworth B (1987) Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology, Evolution, and Systematics*, **18**, 237–268.

Charlesworth D, Wright SI (2001) Breeding systems and genome evolution. *Current Opinion in Genetics & Development*, **11**, 685–690.

Charlesworth D, Morgan MT, Charlesworth B (1990) Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution*, **44**, 1469–1489.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289–1303.

Charlesworth D, Mable BK, Schierup MH, Bartolome C, Awadalla P (2003) Diversity and linkage of genes in the self-incompatibility gene family in *Arabidopsis lyrata*. *Genetics*, **164**, 1519–1535.

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

Darwin C (1876) *The Effects of Cross and Self-Fertilisation in the Vegetable Kingdom*. John Murray, London, UK. Adamant Media Corporation.

Dutheil J, Gaillard S, Bazin E *et al.* (2006) Bio++: a set of C++ libraries for sequence analysis, phylogenetics, molecular evolution and population genetics. *BMC Bioinformatics*, **7**, 188.

Emerson S (1939) A preliminary survey of the *Oenothera-organensis* population. *Genetics*, **24**, 524–537.

Fisher RA (1941) Average excess and average effect of a gene substitution. *Annals of Eugenics*, **11**, 53–63.

Foxe JP, Slotte T, Stahl EA, Neuffer B, Hurka H, Wright SI (2009) Recent speciation associated with the evolution of selfing in *Capsella*. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 5241–5245.

Foxe JP, Stift M, Tedder A, Haudry A, Wright SI, Mable BK (2010) Reconstructing origins of loss of self-incompatibility and selfing in North American *Arabidopsis lyrata*: a population genetic context. *Evolution*, **64**, 3495–3510.

Glémin S, Bataillon T, Ronfort J, Mignot A, Olivieri I (2001) Inbreeding depression in small populations of self-incompatible plants. *Genetics*, **159**, 1217–1229.

Goodwillie C, Kalisz S, Eckert CG (2005) The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 47–79.

Guo YL, Bechsgaard JS, Slotte T *et al*. (2009) Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 5246–5251.

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

Hamrick JL, Godt MJ (1989) Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding and Germplasm Resources* (eds Brown AHD, Clegg MT, Kahler AL and Weir BS), pp. 43–63. Sinauer, Sunderland, Massachusetts.

Hatakeyama K, Takasaki T, Watanabe M, Hinata K (1998) Molecular characterization of S locus genes, SLG and SRK, in a pollen-recessive self-incompatibility haplotype of *Brassica rapa* L. *Genetics*, **149**, 1587–1597.

Haudry A, Cenci A, Ravel C *et al*. (2007) Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Molecular Biology and Evolution*, **24**, 1506–1517.

Haudry A, Cenci A, Guilhaumon C *et al*. (2008) Mating system and recombination affect molecular evolution in four Triticeae species. *Genetics Research*, **90**, 97–109.

Igic B, Bohs L, Kohn JR (2004) Historical inferences from the self-incompatibility locus. *New Phytologist*, **161**, 97–105.

Igic B, Bohs L, Kohn JR (2006) Ancient polymorphism reveals unidirectional breeding system shifts. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 1359–1363.

Igic B, Lande R, Kohn JR (2008) Loss of self-incompatibility and its evolutionary consequences. *International Journal of Plant Sciences*, **169**, 93–104.

Ingvarsson PK (2002) A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. *Evolution*, **56**, 2368–2373.

Jain SK (1976) The evolution of inbreeding in plants. *Annual Review of Ecology, Evolution, and Systematics*, **7**, 469–495.

Kalisz S, Vogler DW, Hanley KM (2004) Context-dependent autonomous self-fertilization yields reproductive assurance and mixed mating. *Nature*, **430**, 884–887.

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.

Keinan A, Mullikin JC, Patterson N, Reich D (2007) Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature Genetics*, **39**, 1251–1255.

Kern AD, Haussler D (2010) A population genetic hidden Markov model for detecting genomic regions under selection. *Molecular Biology and Evolution*, **27**, 1673–1685.

Knight TA (1799) An account of some experiments on the fecundation of vegetables. In a letter from Thomas Andrew Knight, Esq. to the Right Hon. Sir Joseph Banks, K. B. P. R. S. *Philosophical Transactions of the Royal Society of London*, **89**, 195–204.

Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Molecular Biology and Evolution*, **17**, 1483–1498.

Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, Nasrallah ME (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.

Lande R, Schemske DW (1985) The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution*, **39**, 24–40.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

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

Lloyd DG (1965) Evolution of self-compatibility and racial differentiation in *Leavenworthia* (Cruciferae). *Contributions from the Gray Herbarium of Harvard University*, **195**, 3–134.

Lloyd DG (1979) Some reproductive factors affecting the selection of self fertilization in plants. *The American Naturalist*, **113**, 67–79.

Mable BK (2008) Genetic causes and consequences of the breakdown of self-incompatibility: case studies in the Brassicaceae. *Genetics Research*, **90**, 47–60.

Mable BK, Adam A (2007) Patterns of genetic diversity in outcrossing and selfing populations of *Arabidopsis lyrata*. *Molecular Ecology*, **16**, 3565–3580.

Mable BK, Robertson AV, Dart S, Berardo CD, Witham L (2005) Breakdown of self-incompatibility in the perennial *Arabidopsis lyrata* (Brassicaceae) and its genetic consequences. *Evolution*, **59**, 1437–1448.

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetics Research*, **23**, 23–55.

Ness RW, Wright SI, Barrett SCH (2010) Mating-system variation, demographic history and patterns of nucleotide diversity in the Tristylous plant Eichhornia paniculata. *Genetics*, **184**, 381–392.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Nordborg M (2000) Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics*, **154**, 923–929.

Ohta T, Kimura M (1970) Development of associative overdominance through linkage disequilibrium in finite populations. *Genetics Research*, **16**, 165–177.

Pannell JR, Barrett SCH (1998) Baker's law revisited: reproductive assurance in a metapopulation. *Evolution*, **52**, 657–668.

Pollak E (1987) On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics*, **117**, 353–360.

Pool JE, Hellmann I, Jensen JD, Nielsen R (2010) Population genetic inference from genomic sequence variation. *Genome Research*, **20**, 291–300.

R Development Core Team (2010) R: A language and environment for statistical computing. Available from http://www.r-project.org.

Richards AJ (1997) *Plant Breeding Systems*, 2nd edn. Hall and Chapman, London, UK. ed: Garland Science.

Ross-Ibarra J, Wright SI, Foxe JP *et al.* (2008) Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS ONE*, **3**, e2411.

Ruggiero MV, Jacquemin B, Castric V, Vekemans X (2008) Hitch-hiking to a locus under balancing selection: high sequence diversity and low population subdivision at the *S*-locus genomic region in *Arabidopsis halleri*. *Genetics Research*, **90**, 37–46.

Schierup MH, Mable BK, 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.

Stebbins GL (1957) Self fertilization and population variability in the higher plants. *The American Naturalist*, **91**, 337–354.

Stein JC, Howlett B, Boyes DC, Nasrallah ME, Nasrallah JB (1991) Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 8816–8820.

Strobeck C (1980) Heterozygosity of a neutral locus linked to a self-incompatibility locus or a balanced lethal. *Evolution*, **34**, 779–788.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

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

Takuno S, Fujimoto R, Sugimura T *et al.* (2007) Effects of recombination on hitchhiking diversity in the Brassica self-incompatibility locus complex. *Genetics*, **177**, 949–958.

Tang C, Toomajian C, Sherman-Broyles S *et al.* (2007) The evolution of selfing in Arabidopsis thaliana. *Science*, **317**, 1070–1072.

Tsuchimatsu T, Suwabe K, Shimizu-Inatsugi R *et al.* (2010) Evolution of self-compatibility in Arabidopsis by a mutation in the male specificity gene. *Nature*, **464**, 16–21.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical and Applied Genetics*, **7**, 188–193.

Weller SG, Sakai AK (1999) Using phylogenetic approaches for the analysis of plant breeding system evolution. *Annual Review of Ecology, Evolution, and Systematics*, **30**, 167–199.

Wright S (1921) Systems of mating. I. The biometric relations between parent and offspring. *Genetics*, **6**, 111–123.

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

Wright SI, Lauga B, Charlesworth D (2002) Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Molecular Biology and Evolution*, **19**, 1407–1420.

Wright SI, Bi IV, Schroeder SG *et al.* (2005) The effects of artificial selection on the maize genome. *Science*, **308**, 1310–1314.

Wu DD, Jin W, Hao XD, Tang NLS, Zhang YP (2010) Evidence for positive selection on the osteogenin (BMP3) gene in human populations. *PLoS ONE*, **5**, e10959.

A.H. is interested in comparative and population genomics, particularly to study how neutral forces and natural selection are interacting to drive plant genome evolution and diversity patterns. H.G.Z. is a research assistant specialized in molecular biology and plant ecophysiology. M.S. is interested in polyploidization, hybridization and the evolution of mating systems. B.K.M. is interested in the role of self-incompatibility and genome duplication in regulating interactions among organisms (such as pathogen response systems).

Data accessibility

DNA sequences: GenBank accessions JQ030112–JQ030877 (see Data S1 for details).

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1. Detailed methods and sampling information.

Data S1. GenBank accessions numbers.

Fig. S1 Distribution of estimated outcrossing rate.

Fig. S2 DAFS observed at *B80* and expected one under neutrality for each population.

Fig. S3 Linear regressions between genetic diversity statistics (H_o, θ_W and π_S) and outcrossing rate T_m for *B160*.

Table S1 Population diversity statistics for *B80*, *B160* and reference loci in each population.