JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12372

The evolution of selfing from outcrossing ancestors in Brassicaceae: what have we learned from variation at the *S*-locus?

X. VEKEMANS, C. POUX, P. M. GOUBET & V. CASTRIC

Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8198, Université Lille 1, Villeneuve d'Ascq Cedex, France

Keywords:

mating system;
phylogenetics;
self-incompatibility;
selfing;
S-locus.

Abstract

Evolutionary transitions between mating systems have occurred repetitively and independently in flowering plants. One of the most spectacular advances of the recent empirical literature in the field was the discovery of the underlying genetic machinery, which provides the opportunity to retrospectively document the scenario of the outcrossing to selfing transitions in a phylogenetic perspective. In this review, we explore the literature describing patterns of polymorphism and molecular evolution of the locus controlling self-incompatibility (*S*-locus) in selfing species of the Brassicaceae family in order to document the transition from outcrossing to selfing, a retrospective approach that we describe as the 'mating system genes approach'. The data point to strikingly contrasted scenarios of transition from outcrossing to selfing. We also perform original analyses of the fully sequenced genomes of four species showing self-compatibility, to compare the orthologous *S*-locus region with that of functional *S*-locus haplotypes. Phylogenetic analyses suggest that all species we investigated evolved independently towards loss of self-incompatibility, and in most cases almost intact sequences of either of the two *S*-locus genes suggest that these transitions occurred relatively recently. The *S*-locus region in *Aethionema arabicum*, representing the most basal lineage of Brassicaceae, showed unusual patterns so that our analysis could not determine whether self-incompatibility was lost secondarily, or evolved in the core Brassicaceae after the split with this basal lineage. Although the approach we detail can only be used when mating system genes have been identified in a clade, we suggest that its integration with phylogenetic and population genetic approaches should help determine the main routes of this predominant mating system shift in plants.

Introduction

Breeding systems show striking evolutionary flexibility in plants (Barrett, 2002), and processes such as the transition from an outcrossing to a selfing mating system have occurred repetitively and independently within plant families (e.g. more than 60 times in the Solanaceae, Igic *et al.*, 2006), within genera (e.g. four times in the genus *Amsinckia*, Schoen *et al.*, 1997) or even within species (e.g. two times in *Leavenworthia alabamica*, Busch *et al.*, 2011). Ecological and genetic factors influencing the relative evolutionary advantage of outcrossing vs. selfing mating systems have been studied extensively using theoretical modelling, and this has led to some understanding of the conditions under which either outcrossing, selfing or even mixed mating systems should be maintained (Lande & Schemske, 1985; Charlesworth *et al.*, 1990; Porcher & Lande, 2005a). Ecological scenarios such as reproductive assurance favouring shifts towards selfing in association with colonization of new habitats ('Baker's law': Baker, 1955; Pannell & Barrett, 1998), with a reduction in mate availability in small or low density populations (Busch & Schoen, 2008;

Correspondence: Xavier Vekemans, Laboratoire GEPV, UMR CNRS 8198, Université Lille 1, Bat. SN2, Cité scientifique, F-59655 Villeneuve d'Ascq, France. Tel.: +33 320 43 67 53; fax: +33 320 43 69 79; e-mail: xavier.vekemans@univ-lille1.fr

Tsuchimatsu & Shimizu, 2013), or with changes in pollinator services (Biesmeijer *et al.*, 2006) have frequently been cited in the literature (Busch & Delph, 2012). Population genetic scenarios involving a reduction in inbreeding depression due to a strong genetic bottleneck have also been proposed (Byers & Waller, 1999; Bataillon & Kirkpatrick, 2000).

In contrast to this wealth of theoretical predictions, we actually know very little about the main causes of the evolutionary shifts in mating systems that have occurred in the recent history of extant taxa, and empirical data lag far behind. To identify recurrent patterns and establishing general rules about evolutionary transitions from outcrossing to selfing taxa, detailed studies on many different clades comprising mixtures of outcrossing and selfing taxa are crucially needed (Barrett, 1995). Ideally, such retrospective approaches should integrate different sources of information, including phylogenetic approaches aiming at linking patterns of species diversification with mating system traits (e.g. Goldberg *et al.*, 2010) and population genetic approaches aiming at inferring historical demographic parameters from genetic diversity data to detect processes such as genetic bottlenecks associated with these shifts (e.g. Roselius *et al.*, 2005; Busch *et al.*, 2011).

In the present review, we highlight a complementary approach, the 'mating system genes approach', which consists in comparing closely related outcrossing and selfing taxa for allelic diversity, allelic phylogeny and patterns of molecular evolution in the genomic regions responsible for mating system evolution. This approach was made possible by the discovery of the molecular determinants of plant mating systems, in particular in several families exhibiting homomorphic self-incompatibility (SI; Takayama & Isogai, 2005; Charlesworth, 2010). Indeed, the *S*-locus exhibits a set of striking evolutionary features that we can use to learn about the scenario of the transition to selfing from outcrossing ancestors. In SI species, very high allelic diversity is typically found at the *S*-locus (Lawrence, 2000; Castric & Vekemans, 2004), and these alleles have long residence time within lineages (Schierup *et al.*, 1997), so that trans-specific or even transgeneric polymorphisms are commonly found (reviewed in Castric & Vekemans, 2004). These properties are due to the strong negative frequency-dependent selection acting on this multi-allelic system (Wright, 1939; Vekemans & Slatkin, 1994). Moreover, studies of the organization and patterns of molecular evolution in the *S*-locus genomic region of SI species have shown that different *S*-haplotypes exhibit exceptionally low levels of homology among haplotypes, high accumulation of transposable elements, and variable gene orders and sizes (Goubet *et al.*, 2012). These properties are likely caused by the suppression of recombination in the region, and by intense genetic drift due to low effective population sizes within allelic lineages (Vekemans & Slatkin, 1994).

A breakdown of SI, necessarily associated with a transition to selfing, is expected to leave signatures on patterns of diversity and patterns of molecular evolution at the *S*-locus, which could also encapsulate information on the transition process. Indeed, a loss of SI would cause a strong reduction in allelic diversity at the *S*-locus, either as a consequence of positive selection if a self-compatible (SC) mutant allele at the *S*-locus became favoured and swept through the population (Shimizu *et al.*, 2008), or due to the effect of drift, if the mutation to SC was unlinked to the *S*-locus, combined with relaxation of the frequency-dependent selection on the *S*-locus (Charlesworth & Vekemans, 2005). In this situation, alleles found in SC species should represent subsets of alleles found in SI species, and the phylogenetic positions of alleles in different SC species should indicate whether the loss of SI in several related lineages occurred independently (different alleles fixed in different SC species) or not (same allele fixed in different SC species). This approach could underestimate the number of independent transition events if two mutations for SC occurred independently on different copies of the same ancestral *S*-allele, but this is highly unlikely because of the high allelic diversity commonly found in SI lineages. Conversely, this approach could also overestimate the number of transitions if a mutation to SC that is unlinked to the *S*-locus became fixed in a clade and distinct *S*-alleles drifted to fixation in separate lineages. Hence, linkage of the SC mutation to the *S*-locus needs to be verified. Finally, in SC species, fixation of a single allele at the *S*-locus should restore strong homology in this region, thereby increasing allelic effective population size and levels of recombination (because recombination restriction in SI species seems to be due essentially to the high sequence divergence among alleles, Castric *et al.*, 2010).

In this study, we update the review by Mable (2008) of the case studies in Brassicaceae for which polymorphism at the *S*-locus has been investigated in selfing species. We integrate these results with those of other approaches with the aim of identifying recurrent patterns about evolutionary transitions from outcrossing to selfing taxa. For instance, we compare the levels of allelic diversity at the *S*-locus between SI and SC species, and we try to determine whether transitions from SI to SC occurred several times independently within a taxon, or within a clade. We also perform original analyses of the fully sequenced genomes of four species showing self-compatibility. We identify the *S*-locus region in these four species, which we compare with functional *S*-locus haplotypes.

Patterns of allelic diversity at the *S*-locus as an indicator of the timing and mutational causes of the loss of SI

Patterns of allelic diversity at the *S*-locus in species that have lost the SI response may be useful to reconstruct

the evolutionary transition process towards SC. Molecular variation at the *S*-locus has been studied in detail in a few SC species of Brassicaceae including *Arabidopsis thaliana* (e.g. Shimizu *et al.*, 2004, 2008; Bechsgaard *et al.*, 2006; Boggs *et al.*, 2009a), *Arabidopsis kamchatica* (Tsuchimatsu *et al.*, 2012) and *Capsella rubella* (Guo *et al.*, 2009), as well as in secondarily evolved SC populations within predominantly SI species such as *L. alabamica* (Busch *et al.*, 2011) and *Arabidopsis lyrata* (Mable *et al.*, 2005). We do not report results from a survey of the presence/absence of a small set of alleles at the S-locus in selfing and outcrossing populations of *Arabis alpina* (Tedder *et al.*, 2011), as ascertainment bias in this study could alter the interpretations.

Comparisons of allelic diversity at the *S*-locus among SI and SC taxa and/or populations reveal strikingly different patterns (Fig. 1). Within SI taxa or populations, very large numbers of alleles are typically found, as expected theoretically. Individual allelic frequencies vary greatly, however, as a consequence of the dominance relationships typical of sporophytic SI, with recessive alleles showing higher population frequencies than dominant alleles (e.g. Schierup *et al.*, 1997; Billiard *et al.*, 2007). In contrast, in the selfing species *C. rubella*, or in selfing populations of *L. alabamica* (races *a2* and *a4*), only a single allele is found at the *S*-locus (Fig. 1), suggesting that the transition to SC was accompanied by the fixation of a single SC mutation in each species/population (Guo *et al.*, 2009; Busch *et al.*, 2011). Similar results have been observed in the SC species *Leavenworthia torulosa*, *Leavenworthia uniflora* and *Leavenworthia exigua* (Herman *et al.*, 2012). This pattern implies either that the mutation occurred at the *S*-locus and went quickly to fixation, or that the mutation occurred elsewhere but the shift in mating system occurred a long time ago and drift erased *S*-locus polymorphism. An intermediate pattern is observed in the SC species *A. thaliana* and *A. kamchatica* where, respectively, three and five different *SRK* alleles were identified species-wide, suggesting that several independent SC mutations occurred through the transition process (Shimizu *et al.*, 2008; Boggs *et al.*, 2009a; Tsuchimatsu

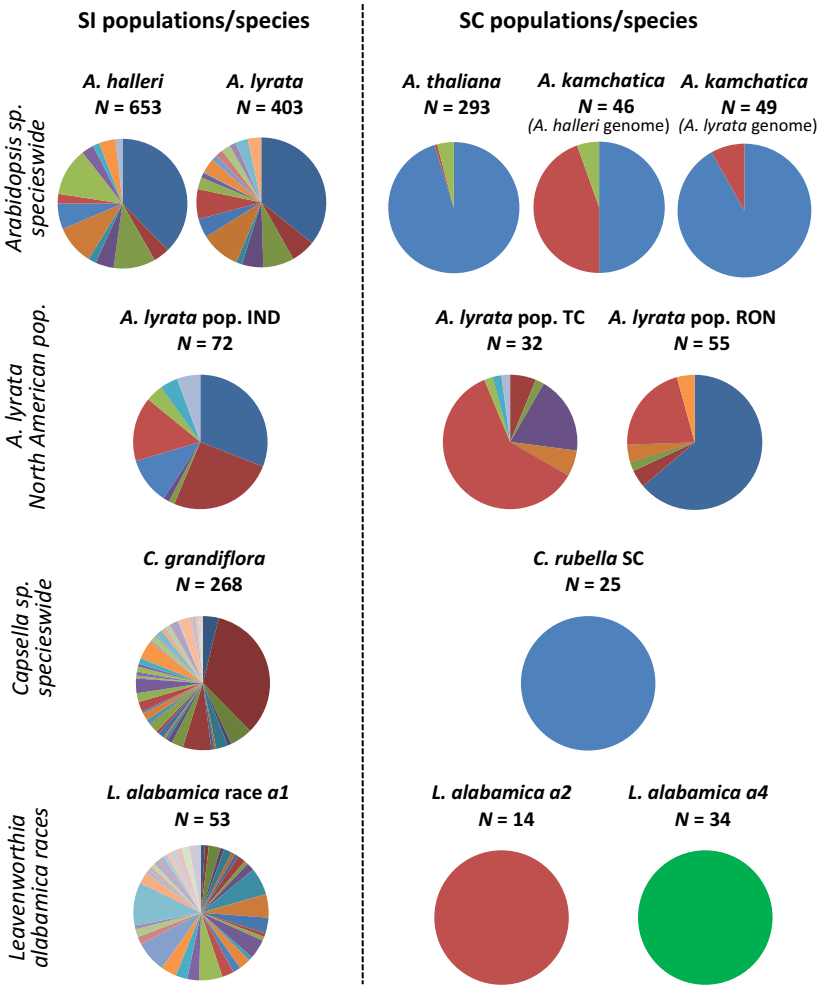

Fig. 1 Distributions of allelic frequencies at the *S*-locus in self-incompatible (SI, *Arabidopsis lyrata*, *Arabidopsis halleri*, *A. lyrata* US population IND, *Capsella grandiflora*, *Leavenworthia alabamica* race *a1*) and self-compatible (SC) taxa or populations (*Arabidopsis thaliana*, *Arabidopsis kamchatica*, *A. lyrata* US populations TC and RON, *Capsella rubella*, *L. alabamica* SC races *a2* and *a4*). *n*: numbers of individuals genotyped in each sample. For the allotetraploid species *A. kamchatica*, allelic frequencies are reported separately for the *A. halleri* and the *A. lyrata* parental genomes. References: *A. lyrata* SI populations, Schierup *et al.* (2008); *A. lyrata* US populations, Mable *et al.* (2005); *A. halleri*, Leducq *et al.* (2011); *A. thaliana*, Shimizu *et al.* (2008); *A. kamchatica*, Tsuchimatsu *et al.* (2012); *C. grandiflora*, J. Bechsgaard & M. Schierup, unpublished; *C. rubella*, Guo *et al.* (2009); *L. alabamica* SI race a1, Joly and Schoen (2011); *L. alabamica* SC races a2 and a4, Busch *et al.* (2011).

et al., 2012), although a scenario with an unlinked modifier causing SC and drift erasing *S*-locus polymorphism secondarily can also be suggested. A way to disentangle these scenarios is to determine whether the SC mutation is linked to the *S*-locus or not. This has been assessed either by mapping the SC mutation in a cross between an SC species and a closely related SI lineage in *C. rubella* (Nasrallah *et al.*, 2007; Slotte *et al.*, 2012), or by restoring the SI response by introducing functional *SCR* and *SRK* alleles from a closely related SI species into the SC species genomic background through transformation experiments in *A. thaliana* (Nasrallah *et al.*, 2002, 2004; Boggs *et al.*, 2009b). Finally, in some North American populations of *A. lyrata* (e.g. populations TC and RON, Fig. 1), evolution towards SC does not seem to have caused a substantial reduction in allelic diversity at the *S*-locus when compared to the SI population IND (Mable *et al.*, 2005).

The causes for these strikingly different patterns of allelic diversity in SC taxa have received little attention. Castric *et al.* (2014) discussed the difference between the *A. thaliana* (i.e. co-occurrence of three independent SC lineages) and the *C. rubella* patterns (i.e. a single SC lineage is fixed). By extending our survey to other data sets, we now highlight four different patterns and discuss their possible causes in relation to the timing of the transition with respect to population history and to the mutational cause of the loss of SI. By doing so, we aim at illustrating how the mating system genes approach, focusing on the patterns of allelic diversity and phylogeny at the *S*-locus, may complement other approaches focusing on functional analyses of the SC phenotype, and on phylogeographic or population genomic analyses of the effects of the transition on non-*S*-linked genomic regions, in order to reconstruct the evolutionary scenario of the transition event.

Independent mutations to SC at the *S*-locus within a species: the *Arabidopsis thaliana* example

In *A. thaliana*, several lines of evidence suggest that the transition to SC occurred very recently (e.g. Shimizu *et al.*, 2011; but see Tang *et al.*, 2007; Indriolo *et al.*, 2012). In particular, the lack of detectable acceleration of the accumulation of nonsynonymous mutations along the three segregating lineages of *SRK* in *A. thaliana* suggests that pseudogenization is very recent (Bechsgaard *et al.*, 2006). Indeed, the split between the *A. thaliana* lineage and its sister SI taxa (*A. lyrata* and *Arabidopsis halleri*) occurred 5.1–13 million years (Myrs) ago (3.1–17.9 Myrs ago considering confidence intervals; Koch *et al.*, 2001; Beilstein *et al.*, 2010), whereas the rate of nonsynonymous to synonymous mutations in the remnants of *SRK* points to a loss-of-function of this gene more recent than 0.413 Myrs ago taking 5 Myrs ago as divergence age between *A. thaliana* and its relatives (Bechsgaard *et al.*, 2006; see Shimizu *et al.*, 2011; for details). Hence, the transition to selfing and the speciation event leading to the *A. thaliana* lineage seem clearly decoupled. Furthermore, the restoration of an apparently full SI phenotype in several accessions of *A. thaliana* in transformation experiments with functional copies of *SCR* and *SRK* from *A. lyrata* (Nasrallah *et al.*, 2002, 2004; Boggs *et al.*, 2009b) demonstrated that the cause of the SC phenotype must be associated with mutations in the pollen/pistil recognition genes rather than to unlinked modifier genes (but see Indriolo *et al.*, 2012). The putative identity of the causal mutations in the three *S*-haplotypes of *A. thaliana* has been investigated thoroughly (Boggs *et al.*, 2009b; Tsuchimatsu *et al.*, 2010; Dwyer *et al.*, 2013).

A striking feature of the *A. thaliana S*-locus is the segregation of three rather than a single SC allele (Fig. 1), as would be expected under a simple selective sweep scenario. The three co-occurring SC haplotypes have slightly different geographical distributions, with one haplotype restricted to offshore western African islands, and contrasting frequencies observed between Europe and Asia for the two others (Shimizu *et al.*, 2008). Thus, a scenario of geographically distinct and independent origins of SC lineages within *A. thaliana* has been proposed, assuming some degree of population subdivision (Shimizu *et al.*, 2008; Boggs *et al.*, 2009a). Such evolutionary scenario, however, raises the question of why SC evolved contemporaneously in different locations in *A. thaliana*, a rather unlikely scenario. We suggest three independent but mutually nonexclusive explanations:

Firstly, the *A. thaliana* lineage, sometime after the split from the *A. lyrata*/*A. halleri* lineage, has experienced major genetic changes in addition to the SI to SC transition. These changes include chromosomal rearrangements leading to a shift from 8 to 5 chromosomes as well as an ongoing process of genome compaction driven by natural selection. This involves a reduction in both the genic repertoire and the number of transposable elements as well as the shortening of intronic and intergenic sequences (Hu *et al.*, 2011; de la Chaux *et al.*, 2012).

Secondly, the transition to selfing is associated with a major change in life form from perennial to annual habit, possibly through loss of the age-dependent vernalization response allowing the species to flower at a younger age (Bergonzi *et al.*, 2013; Zhou *et al.*, 2013). Although the relative timing of these major changes is still unknown, it could be argued that either of these, or both, could have promoted a transition towards selfing. For instance, a recent paper reported that a gene involved in the downstream signalling cascade of SI in *A. lyrata* was actually deleted in *A. thaliana* (*ARC1*, Indriolo *et al.*, 2012). If *ARC1* belonged to a deletion that contributed to decrease the genome size of *A. thaliana*, selection for a smaller genome may have favoured its fixation, hence increasing the rate of selfing as a side

effect. Such a transient period of mixed mating could have facilitated the transition to selfing at the species level (discussed in Castric *et al.*, 2014).

Thirdly, demographic processes affecting the species as a whole such as response to glaciation cycles could have modified the selfing vs. outcrossing balance through several mechanisms including reproductive assurance or founder effects that may have allowed a purge of inbreeding depression and/or a decrease in the number of *S*-alleles (Brennan *et al.*, 2006; Pujol *et al.*, 2009), both factors that are expected to facilitate a transition towards SC (Charlesworth & Charlesworth, 1979; Porcher & Lande, 2005b). Hence, post-glacial colonization can modify important parameters affecting the probability of the transition towards selfing. Phylogeographic structure is strongly pronounced in *A. thaliana*, suggesting that Pleistocene glacial dynamics with recolonization from several independent glacial refugia had a major importance in the recent history of the species (Beck *et al.*, 2008; François *et al.*, 2008).

Fixation of a single SC mutation at the *S*-locus and genomic bottleneck: the *Capsella rubella and Leavenworthia alabamica* examples

In *C. rubella*, the split from the closely related outcrosser *Capsella grandiflora* is very recent (50–100 000 years ago) and is associated with both a strong genome-wide genetic bottleneck and the loss of SI (Foxe *et al.*, 2009; Guo *et al.*, 2009; St Onge *et al.*, 2011; Brandvain *et al.*, 2013). Moreover, in contrast to its sister SI species *C. grandiflora*, which has a very restricted distribution in the northwest of Greece, the highly selfing species *C. rubella* has spread nowadays throughout much of Southern and Western Europe. This contrasts with the *A. thaliana* case where the transition seems to have occurred independently at several locations within a previously large distribution. The strong genetic bottleneck in *C. rubella* has likely been caused by a founding effect at speciation in relation to migration into glacial refuges (Guo *et al.*, 2009; St Onge *et al.*, 2011). In such case, the association between genetic bottleneck and mating system shift could be caused by reproductive assurance favouring an SC mutation in the colonization process (Baker, 1955), or alternatively could be due to a reduction in inbreeding depression (Bataillon & Kirkpatrick, 2000) altering the conditions for spread of the SC mutation.

Genetic mapping of the SC phenotype in *C. rubella* by crosses with *C. grandiflora* showed that the mutation to SC is dominant and is indeed linked to the *S*-locus (Nasrallah *et al.*, 2007; Slotte *et al.*, 2012). Nasrallah *et al.* (2007) performed controlled pollinations among F2 progenies from a cross between a SI individual from *C. grandiflora* and a SC individual from *C. rubella* and suggested that the causal SC mutation occurred in the gene controlling pollen specificity (*SCR*). However, upon re-analysis of their crossing data, we propose the alternative interpretation that one cannot determine precisely whether the mutation occurred in *SRK* or *SCR*, because they could also be explained by effects of allelic dominance, assuming that the SC mutation did not alter the dominance relationship between alleles in pollen. Guo *et al.* (2009) further suggested that the *SCR* gene had lost its second exon, but again our current analysis invalidates this suggestion, such that we argue that the causal mutation has not been identified yet. Beside genetic changes at the *S*-locus, floral and other reproductive characters have also been subject to major adaptive changes (converging to the 'selfing syndrome' detected in other selfing species) during the transition to high selfing (Sicard *et al.*, 2011; Slotte *et al.*, 2012).

A similar association between transition from SI to SC, strong genome-wide genetic bottleneck and evolution of the selfing syndrome has been observed in the selfing race *a4* of *L. alabamica* (Busch *et al.*, 2011). The split between race *a4* (SC) and race *a1* (SI populations) of *L. alabamica* is estimated to have occurred about 150 000 years ago; however, it is not known whether the SC transition occurred at the time of the split, or later. Evolution of selfing has been interpreted as a consequence of reproductive assurance within these peripheral populations: physiological constraints associated with drought in the summer could have generated a pressure to shorten the life cycle and thus leading to a shift towards early-flowering before the emergence of pollinators (Busch & Urban, 2011).

Selfing also evolved independently in the *a2* race in *L. alabamica*, and the SC alleles fixed in both races are distinct (Busch *et al.*, 2011). In contrast to the *a4* race, the evolution of selfing in the *a2* race has not been associated with a strong genome-wide bottleneck nor a selfing syndrome (Busch *et al.*, 2011). Hence, the transition from SI to SC in the history of the *a2* race could have occurred through a local selective sweep of a SC mutation in altered environmental conditions. Hence, this example is closer to the history of the transition to SC in *A. thaliana*, but thanks to the endemic nature of the *a2* race, a single SC allele has been fixed in contrast to *A. thaliana*, which probably showed population subdivision at the time of the transition.

In the two SC races of *L. alabamica*, molecular analyses of the *S*-locus region together with controlled pollination assays showed that the pollen gene (*SCRL*) produced a nonfunctional protein, whereas the pistil gene appeared to be functional (Busch *et al.*, 2011; Chantha *et al.*, 2013). Moreover, analyses of F2 families of crosses between SI and SC individuals of the two races showed cosegregation of SC phenotype with *S*-alleles (Busch *et al.*, 2011). Hence, SC mutations in *L. alabamica* are found at the *S*-locus in both SC races, although we cannot exclude the occurrence of additional mutations at a modifier locus like *ARC1* missing in individuals from the *a4* SC race (Indriolo *et al.*,

2012). However, because the SI system in *Leavenworthia* appears to be nonhomologous to that of other Brassicaceae (Chantha *et al.*, 2013), we cannot know whether *ARC1* is also involved in SI function in this species.

Multiple independent speciation events in the allopolyploid *Arabidopsis kamchatica*

The tetraploid selfing species *A. kamchatica* is widely distributed in East Asia and North America and is known to have an allopolyploid origin, with *A. lyrata* and *A. halleri* as its two SI parental species (Shimizu-Inatsugi *et al.*, 2009). Speciation may have taken place about 250 000 years ago and was apparently associated with a large genome-wide genetic bottleneck (Tsuchimatsu *et al.*, 2012). It is not known whether the loss of SI occurred at speciation time, but the strong bottleneck and the putative reproductive isolation expected for newly evolved allopolyploids suggest that reproductive assurance may have promoted rapid transition to SC in the new species. Patterns of allelic diversity at the *S*-locus in *A. kamchatica*, however, show the co-occurrence of five different *S*-haplotypes segregating at two distinct *S*-loci in this allotetraploid genome (Fig. 1), one locus inherited from *A. halleri* with three co-occurring *S*-haplotypes and the other from *A. lyrata* with two *S*-haplotypes (Tsuchimatsu *et al.*, 2012). This probably results from several independent allopolyploidization events in coherence with data from other loci showing that multiple diploid individuals from both parental species contributed to the origin of *A. kamchatica* (Shimizu-Inatsugi *et al.*, 2009), a pattern that seems to be common in allopolyploids (Soltis *et al.*, 2003). Hence, despite a strong genetic bottleneck and probable association between speciation time and mating system shift, as found in *C. rubella*, the selfing *A. kamchatica* exhibits substantially more allelic diversity than *C. rubella* at the *S*-locus because of the two parental genomes and the multiple origins of the species.

SC mutations potentially unlinked to the *S*-locus, the example of North American populations of *Arabidopsis lyrata*

In SC populations from North American *A. lyrata*, patterns of allelic diversity at the *S*-locus are strikingly different from those in other SC populations or taxa (Fig. 1). Indeed, allelic diversity in populations with a high proportion of SC individuals was found to be almost as high as in geographically nearby SI populations, and not a single *S*-allele with high frequency was found to be strictly associated to the SC phenotype (Mable *et al.*, 2005). The authors suggested that SC in these North American populations could be due to mutations in an unlinked modifier rather than to mutations at the pollen/pistil recognition genes themselves, but this has not been confirmed yet. Indriolo *et al.* (2012) recently found that the *ARC1* gene was present in the *A. lyrata* genome, which was sequenced from an individual from a North American population showing SC individuals. As it cannot be excluded that these populations are segregating for the presence/absence of this modifier gene, we cannot conclude yet whether this modifier gene is involved in the transition to SC in North American *A. lyrata* populations. Moreover, using nuclear and chloroplast nucleotide sequences and microsatellite markers from a range of outcrossing and selfing populations, Foxe *et al.* (2010) suggested that selfing mutations evolved multiple times or that they secondarily spread to multiple genetic backgrounds.

In contrast to the apparently high allelic diversity observed at the *S*-locus in North American selfing populations, Haudry *et al.* (2012) found a significantly lower nucleotide diversity in a gene (*B80*, a *U-Box* gene) flanking the *S*-locus in selfing as compared to outcrossing populations. They found that this reduction is stronger than that expected from a two-fold reduction in effective population size due to selfing. Previous analyses of molecular diversity at *B80* in SI populations of European *A. lyrata* (Kamau & Charlesworth, 2005; Kamau *et al.*, 2007) and *A. halleri* (Ruggiero *et al.*, 2008; Roux *et al.*, 2013) have found that this gene exhibits very high levels of diversity as a consequence of indirect effects of balancing selection acting on the highly linked *S*-locus. Moreover, Roux *et al.* (2013) showed that this linkage effect increases diversity through both an increase in the conservation of ancestral polymorphisms and an increase in the maintenance of recently arisen mutations at *B80*. The relaxation of balancing selection at the *S*-locus associated with the loss of SI is thus expected to reduce these two components of diversity at *B80* beyond the sole reduction in effective population size due to selfing, such that the observation of Haudry *et al.* (2012) cannot be used to determine whether the transition to SC is associated with mutations in the *S*-locus. Altogether, although molecular variation at the *S*-locus in the North American SC populations of *A. lyrata* needs more detailed investigations, it seems to be the only known case of breakdown of SI in Brassicaceae not involving a large decrease in allelic diversity at the *S*-locus.

Identification and characterization of the *S*-locus region in SC species with a full genome sequenced

In Brassicaceae SC species, the *S*-locus region has been studied in details only in *A. thaliana*, where the full sequence of the *S*-locus has been characterized in each of the three different haplogroups as well as in a recombinant haplotype (Kusaba *et al.*, 2001; Shimizu *et al.*, 2004; Tang *et al.*, 2007; Tsuchimatsu *et al.*, 2010; Dwyer *et al.*, 2013). Here, we further analysed full genome sequences from *C. rubella* (Slotte *et al.*, 2013) and

four additional SC species of Brassicaceae (*Eutrema salsugineum*, Yang *et al.*, 2013; *Schrenkiella parvula*, Dassanayake *et al.*, 2011; *Sisymbrium irio* and *Aethionema arabicum*, Haudry *et al.*, 2013), in order to identify the *S*-locus region, to search for functional or nonfunctional copies of *SRK* and *SCR*, and to characterize patterns of molecular evolution in this region as compared to the *S*-locus region in SI species (methods in Appendix S1). To avoid confusion, please note that *E. salsugineum* is sometimes referred to as *Thellungiella salsuginea*, or erroneously as *Thellungiella halophila*, and *S. parvula* was previously, and erroneously, called *Thellungiella parvula* (Koch & German, 2013). In SC species, processes such as accumulation of transposable elements should be stopped, and this could give indication on the time since the loss of SI. Moreover, examination of sequences of the pollen/pistil recognition genes, and of their patterns of expression, as compared to homologous functional alleles in closely related SI species, can also give indication on the functional cause of the breakdown of SI (e.g. Kondo *et al.*, 2002; Tsuchimatsu *et al.*, 2010; Tsuchimatsu & Shimizu, 2013), and on its timing (e.g. Bechsgaard *et al.*, 2006). Note, however, that these results should be interpreted with caution, as multiple mutations would accumulate at *S*-locus genes after the loss of SI, making it difficult to infer the initial causal mutation except in some exceptional cases (e.g. Tsuchimatsu *et al.*, 2012; Dwyer *et al.*, 2013). In all SC species investigated here (but with some ambiguity for *A. arabicum*), we found remnants of the *S*-locus in the same genomic location as that of the *A. thaliana S*-locus, that is, a region within block U of the linkage group 7 of the ancestral Brassicaceae karyotype (Schranz *et al.*, 2006) that is flanked by *B80* (*At4g21350*) and the *SRK* paralog *ARK3* (*Aly8*, *At4g21380*). As shown in Fig. 2 and Table 1, these *S*-locus regions in SC species share many properties with those of SI species (Guo *et al.*, 2011; Goubet *et al.*, 2012), such as the presence of full or partial sequences of *SCR* and *SRK*, large intergenic regions, presence of many transposable elements and high synteny in the flanking regions. In *A. arabicum*, however, many of these features are missing, leaving an ambiguity regarding the derived vs. ancestral nature of SC in this species. Because our results depend on the quality of the published genome sequences which vary considerably among the five published genomes, and have not been checked by subsequent sequencing investigations, we will adopt a conservative attitude and restrict our conclusions to observations of similarities rather than dissimilarities between *S*-locus sequences in SI and SC species.

Occurrence of *SCR* and *SRK*-related sequences at the *S*-locus in SC species

Apparently full sequences of *SCR*, that is, including two exons and 6–8 cysteine residues, were detected in *E. salsugineum*, *C. rubella* and *A. arabicum* (Table 1). However, no *SCR* was detected in *S. irio*, and only the second exon of *SCR* was detected in *S. parvula*. Our results on *C. rubella* are somewhat different from those of Guo *et al.* (2009), who detected only a single copy for which only the first exon was reported, and therefore suggested that the loss of SI was due to a deletion of the second exon of *SCR*. Indeed, based on a more complete database of *SCR* sequences (Goubet *et al.*, 2012), we predicted two copies of *SCR*, each possessing two exons (see Supplementary Material and methods). Functional analyses of these two *SCR* copies, in comparison with the related functional *SCR* copy present in *C. grandiflora*, will now be necessary to further understand what caused a breakdown of SI in *C. rubella*. Apparently full sequences of *SRK*, that is, with at least seven exons, were found in *S. irio*, *E. salsugineum*, *S. parvula* and *C. rubella*. In these species, *S*-locus resequencing as well as RNAseq and quantitative PCR data from pistils would now be useful to check for expression of *SRK* and to obtain its actual coding sequence.

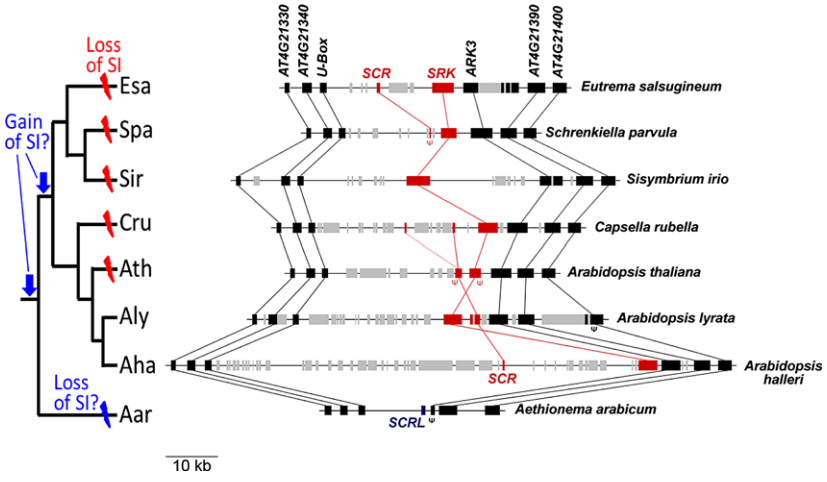

Fig. 2 Synteny in the *S*-locus region among SC and SI species of Brassicaceae for which a full genome sequence is available. The left-hand panel indicates the phylogenetic relationships among the species, together with putative placement of events of gain and loss of SI (see text). The SI genes *SCR* and *SRK* are shown in red; flanking genes in black; predicted transposable elements in grey; a putative non-SI related *SCR-like* gene in *Aethionema arabicum* is shown in blue.

Table 1 Comparison of *S*-locus properties in several SC and SI species of Brassicaceae (see Fig. 2).

Plant genome (SI/SC status)	*S*-locus size (kb)	*SCR* features	*SRK* features	Transposable elements (TE) predictions	Reference for *S*-locus identification	Putative cause of SI loss
Eutrema salsugineum (SC)	26.9 kb	2 exons, 91 aa. residues	8 predicted exons, 834 aa. residues	TE present (17.9% of the *S*-locus), one TE in intron 1 of *ARK3*	This paper (genome paper: Yang *et al.*, 2013)	*SRK*: intron/exon boundary mutation
Schrenkiella parvula (SC)	24.7 kb	Only the second exon was detected	7 exons, 779 aa. residues (deletion in the first exon)	TE present (9.5% of the *S*-locus)	This paper (genome paper: Dassanayake *et al.*, 2011)	*SCR*: loss of exon 1 *SRK*: deletion in exon 1
Sisymbrium irio (SC)	46.4 kb	No *SCR* detected	7 exons, 858 aa. residues	TE present (10.3% of the *S*-locus)	This paper (genome paper: Haudry *et al.*, 2013)	*SCR*: loss of the gene
Capsella rubella (SC)	37.8 kb	2 copies of *SCR*, both with 2 exons, 71 and 75 aa. residues	8 predicted exons, 806 aa. residues	TE present (31.4% of the *S*-locus)	Guo *et al.* (2009) and this paper	*SCR*: unknown after identification of exon 2 *SRK*: intact haplotypes can be found
Arabidopsis thaliana Col-0 (SC)	32 kb	2 exons but second exon interrupted (inversion)	7 exons present but premature stop codon	TE present (35.2% of the *S*-locus)	Tsuchimatsu *et al.* (2010)	*SCR*: inversion *SRK*: intact haplotypes can be found
Arabidopsis lyrata S-allele 13 (SI)	37.3 kb	2 exons, 86 aa. residues	7 exons, 852 aa. residues	TE present (33.7% of the *S*-locus), one TE in the intron of *SCR*	Guo *et al.* (2011) (genome paper: Hu *et al.*, 2011)	Functional
Arabidopsis halleri S-allele 28 (SI)	88.2 kb	2 exons, 81 aa. residues	7 exons, 839 aa. residues	TE present (34.8% of the *S*-locus)	Goubet *et al.* (2012)	Functional
Aethionema arabicum (SC)	13 kb (*ARK3* partially deleted)	*SCR-like* gene, 85 aa. residues	Absent	Absence of predicted TE in the *S*-locus	This paper (genome paper: Haudry *et al.*, 2013)	*SRK*: absence of the gene

Finally, no *SRK* sequence was detected in *A. arabicum*, but again this should be confirmed by resequencing. Overall, sequences of either or both of *SCR* and *SRK* were found in all SC species investigated here. Comparisons with functional sequences of both genes from closely related SI species could be used to infer the time of the SC transitions, as performed in *A. thaliana* (Bechsgaard *et al.*, 2006).

Size of the *S*-locus and density of transposable elements

When defining the *S*-locus as the genomic region ranging from the start codon of the *U-box* flanking gene to the stop codon of the *ARK3* flanking gene (see Fig. 5 in Goubet *et al.*, 2012), the size of the *S*-locus in SC species varied from 13 kb (in *A. arabicum*) to 46.4 kb (in *S. irio*). In SI species, Goubet *et al.* (2012) reported a range of variation between 31 and 110 kb. Distributions of the *S*-locus size in SI and SC species are shown on Fig. 3. The two distributions look different, but with an overlap in the 30- to 50-kb interval. As discussed by Goubet *et al.* (2012), the large size of the *S*-locus was due in part to an accumulation of transposable elements (TE), probably in relation to the absence of recombination in this region. We applied tools relying on TE annotations from the *A. thaliana* genome and found that the density in predicted TEs in the *S*-locus region (expressed as the proportion of the *S*-locus

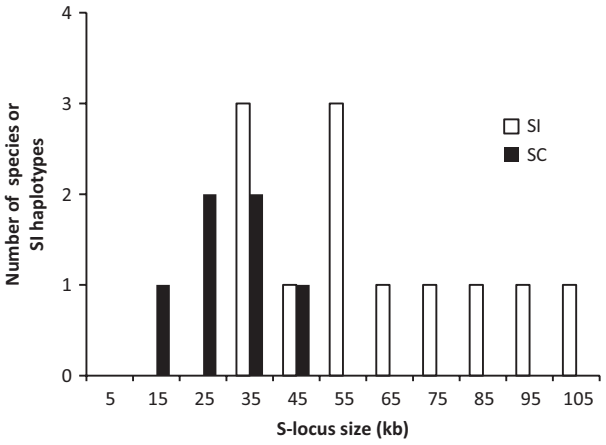

Fig. 3 Distribution of *S*-locus sizes in *S*-haplotypes from SI (in white) and SC (in black) species. Data from SI species taken from Goubet *et al.* (2012).

sequence that matches TE references) of SC species ranged from zero (in *A. arabicum*) to 35.2% (in *A. thaliana*), whereas in SI haplotypes, it was found to range from 10% to 50% (Goubet *et al.*, 2012). Among the six SC species, *S*-locus size and TE density are not significantly correlated ($r = 0.465$). We suggest that TE elimination could proceed more efficiently in the *S*-locus of SC species, due to the recovery of the process of recombination in the region after the likely fixation of a single *S*-haplotype within each species (with the notable exception of *A. thaliana*). Fixation of a single *S*-haplotype should restore a high degree of homology in the *S*-locus region, and recombination among homologous copies of *S*-haplotypes, such as in recessive homozygous individuals in SI species, has been reported previously (Castric *et al.*, 2010).

Genomic location of the *S*-locus in Brassicaceae

The sporophytic SI system found in Brassicaceae is unique, as it has not been reported in the related families Capparaceae and Cleomaceae, nor elsewhere in the eudicot clade. By showing homology of the *SCR* and *SRK* genes involved in pollen–pistil SI recognition in species from two widely divergent lineages in Brassicaceae (*A. lyrata* belonging to lineage I, *Brassica rapa* belonging to lineage II), Kusaba *et al.* (2001) showed that the SI was ancestral in the core Brassicaceae, although the genomic locations of the *S*-locus in the two genera were not shared. Chantha *et al.* (2013) showed that the genomic location of the *S*-locus in *Sisymbrium*, a genus belonging to Brassicaceae lineage II together with *Brassica*, was shared with that of *Arabidopsis* and *Capsella* genera belonging to lineage I, suggesting that the ancestral *S*-locus location was that in *Arabidopsis*. Our results confirm these conclusions as additional taxa from lineage II (*Schrenkiella* and *Eutrema*) also share the *Arabidopsis S*-locus location. Hence, the Brassica *S*-locus has a derived location, putatively in association with the whole-genome triplication that occurred in the Brassiceae tribe (Lysak *et al.*, 2005; Wang *et al.*, 2011). In addition, the *S*-locus in *Leavenworthia* was found at a unique genomic location, different from that in *Arabidopsis* and *Brassica*, and was shown to be derived secondarily in association with the independent recruitment of nonhomologous pollen–pistil recognition genes (Chantha *et al.*, 2013). Our results showing strong homology between several SI and SC species for *SCR* and *SRK* sequences suggest that the *Leavenworthia* case is an exception rather than the rule within Brassicaceae.

Is SC ancestral or derived in *Aethionema arabicum*?

Aethionema arabicum belongs to the early branching sister group to the remainder of the core Brassicaceae (Couvreur *et al.*, 2010) with which it shares a whole-genome duplication event (Haudry *et al.*, 2013). Because species from this lineage do not exhibit self-incompatibility, the SI system present among members of the different lineages of the core Brassicaceae could have appeared either after the split from the *Aethionema* lineage or before the split, in which case it would have secondarily lost its function (Fig. 2). Unfortunately, our analysis of the *S*-locus region in *A. arabicum* does not allow discriminating conclusively between these two scenarios. Among the different SC species analysed here, the *A. arabicum* *S*-locus region shows four unique features (Table 1, Fig. 2): (i) it has by far the smallest *S*-locus region (13 kb instead of 24.7 kb for the second smallest *S*-locus found in *S. parvula*); (ii) *SRK* is completely absent from this genomic region, whereas almost full-length *SRK* sequences were found in all other SC species; (iii) no TEs are detected in this genomic region; and (iv) *ARK3* is partially truncated in the flanking region. According to the genome comparisons made by Haudry *et al.* (2013), these features could not be explained by a smaller genome size or lower TE content in *A. arabicum* as compared to other SC species. Hence, these observations do not point to a signature of recent loss of SI as in other SC species, or in contrast to the other cases, they could suggest that SI was lost a very long time ago. Under the latter scenario, *SRK* could have been lost by a large deletion, which would also explain why only a fragment of *ARK3* is found, and also why the size of the *S*-locus region is so small. However, we found a gene apparently homologous to *SCR* in this region, with no evidence of loss-of-function, which would be incompatible with a scenario of very ancient loss of SI. We suggest that this gene, which we name *SCR-like*, could represent the ancestor of *SCR*, with a function in *A. arabicum* unrelated to SI. Under that scenario, SI would have evolved after the split between the core Brassicaceae and the *Aethionema* lineage (Fig. 2).

Transitions to selfing: a phylogenetic perspective

Remarkable recent developments of phylogenetic comparative methods have allowed a better large-scale understanding of both the directionality and rates of transition between SI and SC and their possible causes (Igic *et al.*, 2004, 2006, 2008; Ferrer & Good-Avila, 2007; Goldberg *et al.*, 2010). For instance, in the Solanaceae family, Goldberg *et al.* (2010) estimated that the rate of transition to selfing was as high as 0.55 transitions per lineage per million years, and Igic *et al.* (2008) showed that the rate of SI loss was 70 times higher than the rate of gain, representing up to 60 losses in this family alone (Igic *et al.*, 2006). Furthermore, Goldberg *et al.* (2010) demonstrated that SI lineages show a substantially higher rate of diversification than SC lineages, which explains the maintenance of the SI system although it has repeatedly been lost. Even though SC lineages display both higher speciation and extinction rates than

SI lineages, their net diversification rate is negative confirming the paradigm describing SC lineages as evolutionary dead-ends (Stebbins, 1957; Igic & Busch, 2013).

Again, the 'mating system genes approach' provides additional insight into the phylogenetic context of the transitions. For instance, in the genus *Solanum* section *Lycopersicon* (the tomato lineage), *S*-alleles from the two SC species *Solanum parviflorum* and *Solanum chmielewskii* were found to belong to lineages that are distinct from those of the *S*-alleles of the SC taxon *Solanum hirsutum f. glabratum*, suggesting independent losses of SI in these two groups (Kondo *et al.*, 2002). In order to evaluate the number of independent transitions from SI to SC that the taxa used in the present study represent, we then conducted a phylogenetic reconstruction based on *SRK* sequences. Sequences for *Leavenworthia* are not included here because the genes involved in pollen and pistil recognition are not homologous to those in the other species (Chantha *et al.*, 2013). The *SRK* gene was not found in *A. arabicum* and therefore not added to the analyses. For each of the alleles detected in SC species (*A. thaliana*, *A. kamchatica*, *C. rubella*, *E. salsugineum*, *S. parvula* and *S. irio*), we could identify a parental or related allele in a closely related SI species (*A. lyrata*, *A. halleri*, *C. grandiflora*; Fig. 4). These results demonstrate that the transitions to SC occurred independently in these species. Similar observations of independent transitions to SC have been reported for each of the *a2* and *a4* SC races of *L. alabamica*, and for each of the congeneric selfing species *L. torulosa*, *L. exigua* and *L. uniflora* (Busch *et al.*, 2011; Herman *et al.*, 2012). Note, however, that for the species in which a single individual genome is available and polymorphism at the *S*-locus has not been properly investigated, it is currently impossible to determine whether several S-alleles segregate (as in *A. thaliana*), such that the number of transitions may actually be even higher. Altogether, although a detailed phylogenetic analysis of the SI/SC trait remains to be carried out in the Brassicaceae (Goldberg *et al.*, 2010), our results suggest several independent transitions from SI to SC within Brassicaceae, as was reported for Solanaceae (Goldberg *et al.*, 2010).

Conclusions

Our analysis of the literature on *S*-locus polymorphism in selfing species from the Brassicaceae revealed strikingly different patterns, which were discussed in relation to the timing and geographical localization of the mating system shift, to its co-occurrence with unique

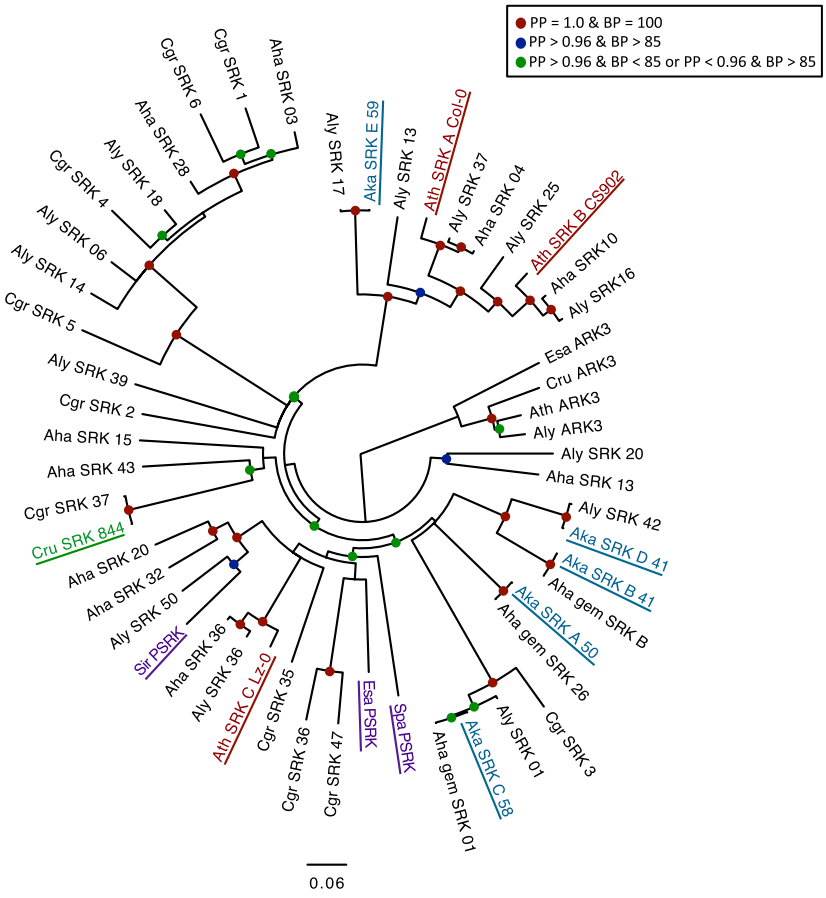

Fig. 4 Phylogenetic relationships among *SRK* sequences from SI taxa (*Arabidopsis lyrata-Aly*, *Arabidopsis halleri-Aha*, *Capsella grandiflora-Cgr*, in black), from SC taxa with species-wide samples (*Arabidopsis thaliana-Ath* in red, *Arabidopsis kamchatica-Aka* in blue, *Capsella rubella-Cru* in green) and from SC taxa with a single genome available (*Sisymbrium irio-Sir*, *Eutrema salsugineum-Esa*, *Shrenkiella parvula-Spa*, in purple). Sequences from *A. thaliana* belong to three clusters, called haplogroups A, B and C, respectively. Sequences from *C. rubella* belong to a single cluster closely related to allele 37 from *C. grandiflora*. Sequences from the allotetraploid *A. kamchatica* belong to five clusters, called haplogroups A–E, with haplogroups A, B and C originating from the *A. halleri* parental species, and haplogroups D and E originating from *A. lyrata*. ARK3 paralog sequences are used as outgroups. Node supports are estimated both by bootstrap (BP, bootstrap percentages) and posterior probabilities (PP).

or multiple speciation events and/or genome-wide bottleneck effects, and to whole-genome duplication events. Phylogenetic analyses of the *S*-haplotypes found in these species, as well as in additional species that we investigated based on full-sequenced genomes, suggested that the transitions to SC evolved independently in all cases. Moreover, sequence data from the *S*-locus region in these species suggested that transitions occurred relatively recently in most cases. Although our conclusions are based on a larger data set, they are very similar to those proposed by Shimizu *et al.* (2011). These results are in agreement with a scenario proposed for the Solanaceae (Goldberg *et al.*, 2010) with high transition rates from SI to SC, producing the observed independent transitions, but high extinction rates of SC lineages, which would explain that most transitions occurred recently. Polymorphism data from additional species, including those investigated here with a single genome sequence, would be necessary to estimate the relative occurrence of the different transition scenarios described here.

Acknowledgments

We thank Denis Roze and Tanja Schwander for the organization of the Roscoff meeting and for the initiative of this special issue. We thank Jesper Bechsgaard and Mikkel Schierup for communicating unpublished results on allele frequencies in *Capsella grandiflora*. Two anonymous reviewers made relevant comments that helped improving the manuscript. This work has been greatly inspired by discussions with Dan Schoen and Takashi Tsuchimatsu, as well as partners of the ANR 'Trans' project. Financial support is from French Agence Nationale de la Recherche (ANR-11-BSV7-013-03).

References

Baker, H.G. 1955. Self-compatibility and establishment after 'long-distance' dispersal. *Evolution* **9**: 347–348.

Barrett, S.C. 1995. Mating-system evolution in flowering plants: micro and macro-evolutionary approaches. *Acta Bot. Neerl.* **44**: 385–402.

Barrett, S.C. 2002. The evolution of plant sexual diversity. *Nat. Rev. Genet.* **3**: 274–284.

Bataillon, T. & Kirkpatrick, M. 2000. Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter. *Genet. Res.* **75**: 75–81.

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. *Mol. Biol. Evol.* **23**: 1741–1750.

Beck, J.B., Schmuths, H. & Schaal, B.A. 2008. Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. *Mol. Ecol.* **17**: 902–915.

Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R. & Mathews, S. 2010. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **107**: 18724–18728.

Bergonzi, S., Albani, M.C., Ver Loren van Themaat, E., Nordström, K.J., Wang, R., Schneeberger, K. *et al.* 2013. Mechanisms of age-dependent response to winter temperature in perennial flowering of *Arabis alpina*. *Science* **340**: 1094–1097.

Biesmeijer, J.C., Roberts, S.P.M., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T. *et al.* 2006. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science* **313**: 351–354.

Billiard, S., Castric, V. & Vekemans, X. 2007. A general model to explore complex dominance patterns in plant sporophytic self-incompatibility systems. *Genetics* **175**: 1351–1369.

Boggs, N.A., Nasrallah, J.B. & Nasrallah, M.E. 2009a. Independent *S*-Locus mutations caused self-fertility in *Arabidopsis thaliana*. *PLoS Genet.* **5**: e1000426.

Boggs, N.A., Dwyer, K.G., Shah, P., McCulloch, A.A., Bechsgaard, J., Schierup, M.H. *et al.* 2009b. Expression of distinct self-incompatibility specificities in *Arabidopsis thaliana*. *Genetics* **182**: 1313–1321.

Brandvain, Y., Slotte, T., Hazzouri, K.M., Wright, S.I. & Coop, G. 2013. Genomic identification of founding haplotypes reveals the history of the selfing species *Capsella rubella*. *PLoS Genet.* **9**: e1003754.

Brennan, A.C., Harris, S.A. & Hiscock, S.J. 2006. The population genetics of sporophytic self-incompatibility in *Senecio squalidus* L. (Asteraceae): the number, frequency, and dominance interactions of S alleles across its British range. *Evolution* **60**: 213–224.

Busch, J.W. & Delph, L.F. 2012. The relative importance of reproductive assurance and automatic selection as hypotheses for the evolution of self-fertilization. *Ann. Bot.* **109**: 553–562.

Busch, J.W. & Schoen, D.J. 2008. The evolution of self-incompatibility when mates are limiting. *Trends Plant Sci.* **13**: 128–136.

Busch, J.W. & Urban, L. 2011. Insights gained from 50 years of studying the evolution of self-compatibility in Leavenworthia (Brassicaceae). *Evol. Biol.* **38**: 15–27.

Busch, J.W., Joly, S. & Schoen, D.J. 2011. Demographic signatures accompanying the evolution of selfing in *Leavenworthia alabamica*. *Mol. Biol. Evol.* **28**: 1717–1729.

Byers, D.L. & Waller, D.M. 1999. Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. *Annu. Rev. Ecol. Syst.* **30**: 479–513.

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

Castric, V., Bechsgaard, J.S., Grenier, S., Noureddine, R., Schierup, M.H. & Vekemans, X. 2010. Molecular evolution within and between self-incompatibility specificities. *Mol. Biol. Evol.* **27**: 11–20.

Castric, V., Billiard, S. & Vekemans, X. 2014. Trait transitions in explicit ecological and genomic contexts: plant mating systems as case studies. In: *Ecological Genomics: Ecology and the Evolution of Genes and Genomes* (C. Landry & N. Aubin-Horth, eds), pp. 7–36. Springer Science, Dordrecht, the Netherlands.

Chantha, S.-C., Herman, A.C., Platts, A.E., Vekemans, X. & Schoen, D.J. 2013. Secondary evolution of a self-incompatibility locus in the Brassicaceae genus Leavenworthia. *PLoS Biol.* **11**: e1001560.

Charlesworth, D. 2010. Self-incompatibility. *F1000 Biol. Rep.* **2**: 68.

Charlesworth, D. & Charlesworth, B. 1979. The evolutionary genetics of sexual systems in flowering plants. *Proc. R. Soc. Lond. B Biol. Sci.* **205**: 513–530.

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

Charlesworth, D., Morgan, M.T. & Charlesworth, B. 1990. Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* **14**: 1469–1489.

de la Chaux, N., Tsuchimatsu, T., Shimizu, K.K. & Wagner, A. 2012. The predominantly selfing plant *Arabidopsis thaliana* experienced a recent reduction in transposable element abundance compared to its outcrossing relative *Arabidopsis lyrata*. *Mob. DNA* **3**: 2.

Couvreur, T.L., Franzke, A., Al-Shehbaz, I.A., Bakker, F.T., Koch, M.A. & Mummenhoff, K. 2010. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Mol. Biol. Evol.* **27**: 55–71.

Dassanayake, M., Oh, D., Haas, J., Hernandez, A., Hong, H., Ali, D. *et al.* 2011. The genome of an extremophile Arabidopsi*S*-relative: *Thellungiella parvula*. *Nat. Genet.* **43**: 913–918.

Dwyer, K.G., Berger, M.T., Ahmed, R., Hritzo, M.K., McCulloch, A.A., Price, M.J. *et al.* 2013. Molecular characterization and evolution of self-incompatibility genes in *Arabidopsis thaliana*: the case of the Sc haplotype. *Genetics* **193**: 985–994.

Ferrer, M.M. & Good-Avila, S.V. 2007. Macrophylogenetic analyses of the gain and loss of self-incompatibility in the Asteraceae. *New Phytol.* **173**: 401–414.

Foxe, J.P., Slotte, T., Stahl, E.A., Neuffer, B., Hurka, H. & Wright, S.I. 2009. Recent speciation associated with the evolution of selfing in Capsella. *Proc. Natl. Acad. Sci. USA* **106**: 5241–5245.

Foxe, J.P., Stift, M., Tedder, A., Haudry, A., Wright, S.I. & Mable, B.K. 2010. Reconstructing origins of loss of self-incompatibility and selfing in North American *Arabidopsis lyrata*: a population genetic context. *Evolution* **64**: 3495–3510.

François, O., Blum, M.G., Jakobsson, M. & Rosenberg, N.A. 2008. Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genet.* **4**: e1000075.

Goldberg, E.E., Kohn, J.R., Lande, R., Robertson, K.A., Smith, S.A. & Igić, B. 2010. Species selection maintains self-incompatibility. *Science* **330**: 493–495.

Goubet, P.M., Bergès, H., Bellec, A., Prat, E., Helmstetter, N., Mangenot, S. *et al.* 2012. Contrasted patterns of molecular evolution in dominant and recessive self-incompatibility haplotypes in Arabidopsis. *PLoS Genet.* **8**: e1002495.

Guo, Y.L., Bechsgaard, J.S., Slotte, T., Neuffer, B., Lascoux, M., Weigel, D. *et al.* 2009. Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proc. Natl. Acad. Sci. USA* **106**: 5246–5251.

Guo, Y.L., Zhao, X., Lanz, C. & Weigel, D. 2011. Evolution of the *S*-locus region in Arabidopsis relatives. *Plant Physiol.* **157**: 937–946.

Haudry, A., Zha, H.G., Stift, M. & Mable, B.K. 2012. Disentangling the effects of breakdown of self-incompatibility and transition to selfing in North American *Arabidopsis lyrata*. *Mol. Ecol.* **21**: 1130–1142.

Haudry, A., Platts, A.E., Vello, E., Hoen, D.R., Leclercq, M., Williamson, R.J. *et al.* 2013. An atlas of over 90 000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat. Genet.* **45**: 891–898.

Herman, A.C., Busch, J.W. & Schoen, D.J. 2012. Phylogeny of Leavenworthia *S*-alleles suggests unidirectional mating system evolution and enhanced positive selection following an ancient population bottleneck. *Evolution* **66**: 1849–1861.

Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.F., Clark, R.M. *et al.* 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43**: 476–481.

Igic, B. & Busch, J.W. 2013. Is self-fertilization an evolutionary dead end? *New Phytol.* **198**: 386–397.

Igic, B., Bohs, L. & Kohn, J.R. 2004. Historical inferences from the self-incompatibility locus. *New Phytol.* **161**: 97–105.

Igic, B., Bohs, L. & Kohn, J.R. 2006. Ancient polymorphism reveals unidirectional breeding system shifts. *Proc. Natl. Acad. Sci. USA* **103**: 1359–1363.

Igic, B., Lande, R. & Kohn, J.R. 2008. Loss of self-incompatibility and its evolutionary consequences. *Int. J. Plant Sci.* **169**: 93–104.

Indriolo, E., Tharmapalan, P., Wright, S.I. & Goring, D.R. 2012. The *ARC1* E3 ligase gene is frequently deleted in self-compatible Brassicaceae species and has a conserved role in *Arabidopsis lyrata* self-pollen rejection. *Plant Cell* **24**: 4607–4620.

Joly, S. & Schoen, D.J. 2011. Migration rates, frequency-dependent selection and the self-incompatibility locus in Leavenworthia (Brassicaceae). *Evolution* **65**: 2357–2369.

Kamau, E. & Charlesworth, D. 2005. Balancing selection and low recombination affect diversity near the self-incompatibility loci of the plant *Arabidopsis lyrata*. *Curr. Biol.* **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.

Koch, M.A. & German, D.A. 2013. Taxonomy and systematics are key to biological information: *Arabidopsis*, *Eutrema* (*Thellungiella*), *Noccaea* and *Schrenkiella* (Brassicaceae) as examples. *Front. Plant Sci.* **4**: 267.

Koch, M.A., Haubold, B. & Mitchell-Olds, T. 2001. Molecular systematics of the Brassicaceae: evidence from coding plastidic *MATK* and nuclear *CHS* sequences. *Am. J. Bot.* **88**: 534–544.

Kondo, K., Yamamoto, M., Itahashi, R., Sato, T., Egashira, H., Hattori, T. *et al.* 2002. Insights into the evolution of self-compatibility in Lycopersicon from a study of stylar factors. *Plant J.* **30**: 143–153.

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.

Lande, R. & Schemske, D.W. 1985. The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution* **39**: 24–40.

Lawrence, M.J. 2000. Population genetics of homomorphic self-incompatibility polymorphisms in flowering plants. *Ann. Bot.* **85**: 221–226.

Leducq, J.B., Llaurens, V., Castric, V., Saumitou-Laprade, P., Hardy, O.J. & Vekemans, X. 2011. Effect of balancing selection on spatial genetic structure within populations: theoretical investigations on the self-incompatibility locus and empirical studies in *Arabidopsis halleri*. *Heredity* **106**: 319–329.

Lysak, M.A., Koch, M.A., Pecinka, A. & Schubert, I. 2005. Chromosome triplication found across the tribe Brassiceae. *Genome Res.* **15**: 516–525.

Mable, B.K. 2008. Genetic causes and consequences of the breakdown of self-incompatibility: case studies in the Brassicaceae. *Genet. Res. (Camb)* **90**: 47–60.

Mable, B.K., Robertson, A.V., Dart, S., Di Berardo, C. & Witham, L. 2005. Breakdown of self-incompatibility in the perennial *Arabidopsis lyrata* (Brassicaceae) and its genetic consequences. *Evolution* **59**: 1437–1448.

Nasrallah, M.E., Liu, P. & Nasrallah, J.B. 2002. Generation of self-incompatible *Arabidopsis thaliana* by transfer of two S locus genes from *A. lyrata*. *Science* **297**: 247–249.

Nasrallah, M.E., Liu, P., Sherman-Broyles, S., Boggs, N. & Nasrallah, J.B. 2004. Natural variation in expression of self-incompatibility in *Arabidopsis thaliana*: implications for the evolution of selfing. *Proc. Natl. Acad. Sci. USA* **101**: 16070–16074.

Nasrallah, J.B., Liu, P., Sherman-Broyles, S., Schmidt, R. & Nasrallah, M.E. 2007. Epigenetic mechanisms for breakdown of self-incompatibility in interspecific hybrids. *Genetics* **175**: 1965–1973.

Pannell, J.R. & Barrett, S.C.H. 1998. Baker's law revisited: reproductive assurance in a metapopulation. *Evolution* **53**: 664–676.

Porcher, E. & Lande, R. 2005a. The evolution of self-fertilization and inbreeding depression under pollen discounting and pollen limitation. *J. Evol. Biol.* **18**: 497–508.

Porcher, E. & Lande, R. 2005b. Loss of gametophytic self-incompatibility with evolution of inbreeding depression. *Evolution* **59**: 46–60.

Pujol, B., Zhou, S.R., Sanchez-Vilas, J. & Pannell, J.R. 2009. Reduced inbreeding depression after species range expansion. *Proc. Natl. Acad. Sci. USA* **106**: 15379–15383.

Roselius, K., Stephan, W. & Städler, T. 2005. The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics* **171**: 753–763.

Roux, C., Pauwels, M., Ruggiero, M.V., Charlesworth, D., Castric, V. & Vekemans, X. 2013. Recent and ancient signature of balancing selection around the *S*-locus in *Arabidopsis halleri* and *A. lyrata*. *Mol. Biol. Evol.* **30**: 435–447.

Ruggiero, M.V., 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*. *Genet. Res. (Camb)* **90**: 37–46.

Schierup, M.H., Vekemans, X. & Christiansen, F.B. 1997. Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. *Genetics* **147**: 835–846.

Schierup, M.H., Bechsgaard, J.S. & Christiansen, F.B. 2008. Selection at work in self-incompatible *Arabidopsis lyrata*. II. Spatial distribution of S haplotypes in Iceland. *Genetics* **180**: 1051–1059.

Schoen, D.J., Johnston, M.O., L'Heureux, A.-M. & Marsolais, J.V. 1997. Evolutionary history of the mating system in Amsinckia (Boraginaceae). *Evolution* **51**: 1090–1099.

Schranz, M.E., Lysak, M.A. & Mitchell-Olds, T. 2006. The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends Plant Sci.* **11**: 535–542.

Shimizu, K.K., Cork, J.M., Caicedo, A.L., Mays, C.A., Moore, R.C., Olsen, K.M. *et al.* 2004. Darwinian selection on a selfing locus. *Science* **306**: 2081–2084.

Shimizu, K.K., Shimizu-Inatsugi, R., Tsuchimatsu, T. & Purugganan, M.D. 2008. Independent origins of self-compatibility in *Arabidopsis thaliana*. *Mol. Ecol.* **17**: 704–714.

Shimizu, K.K., Kudoh, H. & Kobayashi, M.J. 2011. Plant sexual reproduction during climate change: gene function in natura studied by ecological and evolutionary systems biology. *Ann. Bot.* **108**: 777–787.

Shimizu-Inatsugi, R., Lihová, J., Iwanaga, H., Kudoh, H., Marhold, K., Savolainen, O. *et al.* 2009. The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *Arabidopsis lyrata* and *Arabidopsis halleri*. *Mol. Ecol.* **18**: 4024–4048.

Sicard, A., Stacey, N., Hermann, K., Dessoly, J., Neuffer, B., Bäurle, I. *et al.* 2011. Genetics, evolution, and adaptive significance of the selfing syndrome in the genus Capsella. *Plant Cell* **23**: 3156–3171.

Slotte, T., Hazzouri, K.M., Stern, D., Andolfatto, P. & Wright, S.I. 2012. Genetic architecture and adaptive significance of the selfing syndrome in Capsella. *Evolution* **66**: 1360–1374.

Slotte, T., Hazzouri, K.M., Ågren, J.A., Koenig, D., Maumus, F., Guo, Y.L. *et al.* 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* **45**: 831–835.

Soltis, D.E., Soltis, P.S. & Tate, J.A. 2003. Advances in the study of polyploidy since plant speciation. *New Phytol.* **161**: 173–191.

St Onge, K.R., Kallman, T., Slotte, T., Lascoux, M. & Palme, A.E. 2011. Contrasting demographic history and population structure in *Capsella rubella* and *Capsella grandiflora*, two closely related species with different mating systems. *Mol. Ecol.* **20**: 3306–3320.

Stebbins, G.L. 1957. Self-fertilization and population variability in the higher plants. *Am. Nat.* **91**: 337–354.

Takayama, S. & Isogai, A. 2005. Self-incompatibility in plants. *Annu. Rev. Plant Biol.* **56**: 467–489.

Tang, C., Toomajian, C., Sherman-Broyles, S., Plagnol, V., Guo, Y.L., Hu, T.T. *et al.* 2007. The evolution of selfing in *Arabidopsis thaliana*. *Science* **317**: 1070–1072.

Tedder, A., Ansell, S.W., Lao, X., Vogel, J.C. & Mable, B.K. 2011. Sporophytic self-incompatibility genes and mating system variation in *Arabis alpina*. *Ann. Bot.* **108**: 699–713.

Tsuchimatsu, T. & Shimizu, K.K. 2013. Effects of pollen availability and the mutation bias on the fixation of mutations disabling the male specificity of self-incompatibility. *J. Evol. Biol.* **26**: 2221–2232.

Tsuchimatsu, T., Suwabe, K., Shimizu-Inatsugi, R., Isokawa, S., Pavlidis, P., Städler, T. *et al.* 2010. Evolution of self-compatibility in Arabidopsis by a mutation in the male specificity gene. *Nature* **464**: 1342–1346.

Tsuchimatsu, T., Kaiser, P., Yew, C.L., Bachelier, J.B. & Shimizu, K.K. 2012. Recent loss of self-incompatibility by degradation of the male component in allotetraploid *Arabidopsis kamchatica*. *PLoS Genet.* **8**: e1002838.

Vekemans, X. & Slatkin, M. 1994. Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics* **137**: 1157–1165.

Wang, X.W., Wang, H., Wang, J., Sun, R., Wu, J., Brassica rapa Genome Sequencing Project Consortium *et al.* 2011.

The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* **43**: 1035–1039.
Wright, S. 1939. The distribution of self-sterility alleles in populations. *Genetics* **24**: 538–552.
Yang, R., Jarvis, D.E., Chen, H., Beilstein, M.A., Grimwood, J. & Jenkins, J. 2013. The reference genome of the halophytic plant *Eutrema salsugineum*. *Front. Plant Sci.* **4**: 46.
Zhou, C.-M., Zhang, T.-Q., Wang, X., Yu, S., Lian, H., Tang, H. *et al.* 2013. Molecular basis of age-dependent vernalization in *Cardamine flexuosa*. *Science* **340**: 1097–1100.

Supporting information

Additional Supporting Information may be found in the online version of this article:
Appendix S1 Methods for annotation of the *S*-locus regions and for phylogenetic reconstructions.

Received 1 November 2013; revised 6 March 2014; accepted 10 March 2014