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 Évolution 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, Igie 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 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;
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 et al., 2012).

![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, 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, Ledueq 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).](image-url)
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; Slote 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 bottlenecks 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 ARCI missing in individuals from the a4 SC race (Indriolo et al.,
S-locus inherited from S-the new species. Patterns of allelic diversity at the and the putative reproductive isolation expected for SI occurred at speciation time, but the strong bottleneck matsu about 250 000 years ago and was apparently associated sugi A. halleri JOURNAL OF EVOLUTIONARY BIOLOGY © 2014 THE AUTHORS. Indriolo tions at the pollen/pistil recognition genes themselves, mutations in an unlinked modifier rather than to muta-
tions, and not a single SC individuals was found to be almost as high as in geographically nearby SI popula-
tions. 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 sel-
fing populations, Foxe et al. (2010) suggested that selling 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 pop-
ulations, Haudry et al. (2012) found a significantly lower nucleotide diversity in a gene (B80, a U-Box gene) flanking the S-locus in selling 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 selling. 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 ances-
tral 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 compo-
nents of diversity at B80 beyond the sole reduction in effective population size due to selling, 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 Ameri-
can 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.

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 dif-
ferent 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 muta-
tions at the pollen/pistil recognition genes themselves, but this has not been confirmed yet. Indriolo et al. (2012) recently found that the ARCI 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 pop-
ulations 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 sel-
fing populations, Foxe et al. (2010) suggested that selling mutations evolved multiple times or that they secondarily spread to multiple genetic backgrounds.

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 gen-
ome 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 RNAsseq 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).

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

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
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 Brassicaceae 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?**

*Arabidopsis arabicum* belongs to the early branching sister group to the remainder of the core Brassicaceae (Couvrour 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. kamchatka, 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 kamchatka-Aka in blue, Capsella rubella-Cru in green) and from SC taxa with a single genome available (Sisymbrium alatum-Sal, 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. kamchatka 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 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 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 initiation 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


**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