View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Bradford Scholars

1 # Genetic basis and timing of a major mating system shift in *Capsella*

2

3 Jörg A. Bachmann[1,¶], Andrew Tedder[1,¶], Benjamin Laenen[1,¶], Marco Fracassetti[1], Aurélie
4 Désamoré[1], Clément Lafon-Placette[2], Kim A. Steige[1], Caroline Callot[3], William Marande[3],
5 Barbara Neuffer[4], Hélène Bergès[3], Claudia Köhler[2], Vincent Castric[5], Tanja Slotte[1]*
6 [1]Department of Ecology, Environment and Plant Sciences, Science for Life Laboratory,
7 Stockholm University, SE-106 91 Stockholm, Sweden, [2]Department of Plant Biology,
8 Swedish University of Agricultural Sciences & Linnean Center for Plant Biology, SE-750 07
9 Uppsala, Sweden[3]Institut National de la Recherche Agronomique UPR 1258, Centre National
10 des Ressources Génomiques Végétales, Castanet-Tolosan, France, [4]Department of Botany,
11 University of Osnabruck, Osnabruck, Germany, [5]Univ. Lille, CNRS, UMR 8198 - Evo-Eco-
12 Paleo, F-59000 Lille, France
13
14 *Corresponding author: Email: tanja.slotte@su.se, Phone: +46-76-1252109
15 ¶These authors contributed equally to this work.
16

17 ## Summary

18 • A crucial step in the transition from outcrossing to self-fertilization is the loss of
19 genetic self-incompatibility (SI). In the Brassicaceae, SI involves interaction of female
20 and male specificity components, encoded by the genes *SRK* and *SCR* at the self-
21 incompatibility locus (*S*-locus). Theory predicts that *S*-linked mutations, and
22 especially dominant mutations in *SCR*, are likely to contribute to loss of SI. However,
23 few studies have investigated the contribution of dominant mutations to loss of SI in
24 wild plant species.

25 • Here, we investigate the genetic basis of loss of SI in the self-fertilizing crucifer
26 species *Capsella orientalis,* by combining genetic mapping, long-read sequencing of
27 complete *S*-haplotypes, gene expression analyses, and controlled crosses.
28 • We show that loss of SI in *C. orientalis* occurred less than 2.6 Mya and maps as a
29 dominant trait to the *S*-locus. We identify a fixed frameshift deletion in the male
30 specificity gene *SCR* and confirm loss of male SI specificity. We further identify an *S*-
31 linked small RNA that is predicted to cause dominance of self-compatibility.
32 • Our results agree with predictions on the contribution of dominant *S*-linked mutations
33 to loss of SI, and thus provide new insights into the molecular basis of mating system
34 transitions.
35
36 **Keywords:** *Capsella*, dominance modifier, long-read sequencing, parallel evolution, plant
37 mating system shift, self-compatibility, S-locus, small RNA

38 ## Introduction

39 The shift from outcrossing to self-fertilization is one of the most common evolutionary
40 transitions in flowering plants (Darwin, 1876; Wright et al., 2013). This transition is favored
41 when the benefits of reproductive assurance (Darwin, 1876; Pannell & Barrett, 1998; Eckert
42 et al., 2006) and the transmission advantage of selfing (Fisher, 1941) outweigh the cost of
43 inbreeding depression (Charlesworth, 2006).

44 The transition to self-fertilization often involves breakdown of self-incompatibility
45 (SI). SI systems allow plants to recognize and reject self pollen through the action of male and
46 female specificity components and modifier loci (Takayama & Isogai, 2005). In the
47 Brassicaceae, SI is controlled by two tightly linked genes at the *S*-locus, the *S*-locus receptor
48 kinase gene *SRK* and *SCR*, which encode the female and male SI specificity determinants,
49 respectively (de Nettancourt, 2001). SRK is a transmembrane serine-threonine receptor kinase
50 located on the stigma surface (Stein et al., 1991; 1996). SCR is a small cysteine-rich protein
51 that is deposited on the pollen coat and acts as a ligand to the SRK receptor (Schopfer et al.,
52 1999; Takayama et al., 2001). Direct interaction between SRK and SCR from the same *S*-
53 haplotype results in inhibition of pollen germination (Takasaki et al., 2000; Takayama et al.,
54 2001; Ma et al., 2016) through a signaling cascade involving several proteins (Nasrallah &
55 Nasrallah 2014). This SI response prevents close inbreeding and promotes outcrossing. At the
56 *S*-locus, recombination is suppressed and rare allele advantage maintains alleles with different
57 specificities (Wright, 1939; Castric & Vekemans, 2004; Vekemans et al., 2014). SI
58 populations often harbor dozens of highly diverged *S*-haplotypes as a result of negative
59 frequency-dependent selection (Mable et al., 2003; Guo et al., 2009). In the sporophytic
60 Brassicaceae SI system, expression of a single *S*-specificity provides greater compatibility
61 with other individuals (Schoen & Busch, 2009). Therefore, *S*-haplotypes often form a
62 dominance hierarchy that determines which specificity is expressed in *S*-heterozygotes

63 (Durand et al., 2014). At the pollen level, dominance is governed by dominance modifiers in
64 the form of sRNAs expressed by dominant alleles. These sRNAs target sequence motifs
65 specific to recessive alleles of *SCR*, resulting in transcriptional silencing (Tarutani et al.,
66 2010; Durand et al., 2014).

67 Despite the advantages of outcrossing, SI has been lost repeatedly in many different
68 lineages. There is a strong theoretical and empirical interest in the role of parallel molecular
69 changes for repeated shifts to self-compatibility (SC) (Vekemans et al., 2014; Shimizu &
70 Tsuchimatsu, 2015). While the numerous genes that act as unlinked modifiers of SI
71 potentially constitute a larger mutational target than the *S*-locus itself, theory predicts that
72 mutations that result in degeneration of components of the *S*-locus should have an advantage
73 (Porcher & Lande, 2005). Theory further predicts that the probability of spread of mutations
74 disrupting SI depends on whether they affect male or female SI function, or both functions
75 jointly (Charlesworth & Charlesworth, 1979). In particular, mutations that disrupt male
76 specificity should have an advantage over those mutations that disrupt female specificity,
77 because male specificity mutations can spread faster through both pollen and seeds
78 (Uyenoyama et al., 2001; Tsuchimatsu & Shimizu, 2013). Finally, dominant advantageous
79 mutations should have a higher fixation probability in outcrossers, as expected from
80 Haldane's sieve (Haldane 1927). However, dominant *S*-alleles typically have low population
81 frequencies (Llaurens et al., 2008), resulting in a lower probability that SC mutations occur on
82 dominant than on recessive alleles. While degeneration of male specificity has contributed to
83 loss of SI in a few Brassicaceae species (Tsuchimatsu et al., 2010; 2012; Shimizu &
84 Tsuchimatsu, 2015) Chantha et al., 2013), more examples are needed. So far, few empirical
85 studies of wild species have examined the contribution of dominant *S*-haplotypes to the loss
86 of SI (but see Nasrallah et al. 2007). To understand the role of parallel molecular changes for
87 recurrent loss of SI, identification of causal mutations is required. This has been a challenging

88 task, due to the difficulty of sequencing the up to 110 kb long, highly polymorphic and
89 repetitive *S*-locus. However, thanks to the advent of long-read sequencing, contiguous *S*-
90 haplotypes can now be assembled with low error rates (Bachmann et al., 2018).

91 The crucifer genus *Capsella* is an emerging model for genomic studies of plant mating
92 system evolution. In *Capsella*, SI is the ancestral state, as there is trans-specific shared *S*-
93 locus polymorphism between the outcrossing SI species *Capsella grandiflora* and outcrossing
94 SI *Arabidopsis* species (Guo et al., 2009). Nevertheless, SC has evolved repeatedly in
95 *Capsella*, resulting in two self-compatible and highly selfing diploid species, *Capsella rubella*
96 and *Capsella orientalis*, as well as the selfing allotetraploid *Capsella bursa-pastoris*, which
97 formed by hybridization between *C. orientalis* and *C. grandiflora* accompanied by genome
98 duplication (Douglas et al., 2015). These species also differ greatly in their geographical
99 distributions, with *C. bursa-pastoris* having a nearly worldwide distribution, whereas *C.*
100 *rubella* is mainly found in Central and Southern Europe, and *C. orientalis* has a distribution
101 ranging from Eastern Europe to Central Asia (Hurka et al., 2012). Finally, the SI outcrosser
102 *C. grandiflora* mainly occurs in northwestern Greece and Albania, and in northern Italy
103 (Hurka et al., 2012).

104 In *C. rubella*, the transition to selfing has been intensely studied (Foxe et al., 2009;
105 Guo et al., 2009; Slotte et al., 2013; Brandvain et al., 2013) and involved the fixation of a
106 relatively dominant *S*-haplotype (Nasrallah et al., 2007; Guo et al., 2009; Paetsch et al. 2010)
107 most likely within the past 100-170 ky (Slotte et al., 2013; Koenig et al., 2019). Knowledge
108 on the mode, timing and demographics of the transition to selfing in *C. rubella* has provided
109 an evolutionary context for the study of genomic (Gos et al., 2012; Slotte et al., 2013;
110 Brandvain et al., 2013; Koenig et al., 2019), regulatory (Steige et al. 2015) and phenotypic
111 (Slotte et al., 2012; Sicard et al., 2016) consequences of selfing. In contrast, we know little
112 about the genetic basis and timing of loss of SI and transition to selfing in *C. orientalis*. Such

113 information is important for proper interpretation of genomic studies of the effects of selfing
114 and can provide insights into the role of parallel molecular changes for convergent loss of SI.
115 Here, we combined genetic mapping, long-read sequencing of *S*-haplotypes,
116 controlled crosses, population genomic and expression analyses to investigate the loss of SI in
117 *C. orientalis*, with the specific aims to: 1) test whether loss of SI maps to the *S*-locus, 2)
118 identify candidate causal mutations for the loss of SI, 3) investigate the role of sRNA-based
119 dominance modifiers, and 4) estimate the timing of loss of SI in *C. orientalis*.
120

121 # Materials and Methods

122 **Plant material and growth conditions**

123 We surface-sterilized seeds of *Capsella orientalis* Klokov*, Capsella bursa-pastoris* (L.)
124 Medik. and *Capsella grandiflora* (Fauché & Chaub.) Boiss. (S1 Table, Supporting
125 Information), plated them on ½ MS medium (Murashige and Skoog basal salt mixture,
126 Sigma-Aldrich Co. MI, USA) and stratified seeds at 2-4°C in the dark for two weeks. Plates
127 were then moved to controlled climate chambers (16 h light at 20ºC / 8 h dark at 18 ºC, 70 %
128 maximum humidity, 122 uE light intensity). After one week, seedlings were transplanted to
129 soil in pots. For genotyping and whole-genome resequencing, leaf samples for DNA
130 extractions were collected from >3 week old plants and dried in silica gel. For bacterial
131 artificial chromosome (BAC) library construction, leaf samples were collected after 48 h dark
132 treatment and were immediately flash-frozen in liquid N_2. For RNA extractions, mixed-stage
133 floral buds and leaf samples were collected in the middle of the light period and immediately
134 flash-frozen in liquid N_2.
135

136 **Genetic mapping of loss of SI in *C. orientalis***

137 To test whether loss of SI mapped to the *S*-locus, we generated an interspecific *C. orientalis* ×
138 *C. grandiflora* F2 mapping population which segregated for SI/SC by crossing *C. orientalis*
139 accession Co2008-1 as seed parent to *C. grandiflora* accession Cg88.15 as pollen donor (S1
140 Table, Supporting Information). Because *C. orientalis* × *C. grandiflora* F1 seeds were aborted
141 prior to full development, generating viable F1 seeds required embryo rescue (Methods S1,
142 Supporting Information). F1 individuals were SC, and we collected F2 seeds from one
143 autonomously self-pollinated F1 individual. Our mapping population consisted of a total of
144 350 F2 individuals. We extracted DNA from all F2 individuals using a Qiagen DNeasy kit
145 (Qiagen, Venlo, The Netherlands) and genotyped them at 998 SNPs at SciLifelab Stockholm
146 (Methods S1, Supporting Information).

147 We scored SI/SC in a total of 321 F2 individuals. SI/SC was scored as presence or
148 absence of silique formation on mature individuals. In addition, we assessed the success of 3-
149 6 manual self-pollinations for 204 F2 individuals. In the case of a discrepancy between seed
150 set after manual self-pollination and silique formation after autonomous self-pollination, we
151 used the scoring based on manual self-pollination. To validate that the SI phenotype was due
152 to pollen tube growth arrest and the lack of seed development following self-pollination was
153 not due to e.g. inbreeding depression or later-acting genetic incompatibilities, we assessed
154 pollen tube growth in the pistil after manual self-pollination in a subset of 10 F2 individuals
155 scored as SI (Methods S1, Supporting Information).

156 We generated a linkage map and mapped quantitative trait loci (QTL) for SI/SC status
157 in R/Qtl (Broman et al., 2003). The final linkage map had 549 SNPs after removal of SNPs
158 with segregation distortion or redundant genotype information. We mapped QTL for SI/SC,
159 encoded as a binary trait, using interval mapping and the Haley & Knott regression method
160 (Haley & Knott, 1992) in intervals of 1 cM, based on 304 F2 individuals for which we had
161 both phenotype and genotype data. A 1% genome-wide significance threshold was obtained

162 by 1000 permutations and we estimated credible intervals of significant QTL as 1.5-LOD
163 drop intervals. We estimated the additive allelic effect and dominance deviation using the
164 R/Qtl effectscan function.
165

166 **Sequencing, assembly and annotation of the *S*-locus in *Capsella***

167 To identify putative causal genetic changes responsible for loss of SI in *C. orientalis*, we
168 conducted targeted sequencing and assembly of *S*-haplotypes by long-read sequencing of
169 BAC clones containing the *S*-locus, as in Bachmann et al. (2018) (Methods S1, Supporting
170 Information). We conducted targeted long-read sequencing and assembly of *S*-haplotypes of
171 two SC *C. orientalis* accessions, four SC *C. bursa-pastoris* accessions and two SI *C.*
172 *grandiflora* accessions. The two *C. grandiflora S*-haplotypes presented here were chosen
173 from a larger set of 15 *S*-haplotypes to represent the *C. grandiflora S*-haplotype segregating in
174 our F2 population as well as a *C. grandiflora S*-haplotype from the same haplogroup as the *S*-
175 haplotype of *C. orientalis* (see "Phylogenetic analyses of *S*-locus sequences" below; S1 Table,
176 Supporting Information). In total, we here present eight full-length *S*-locus haplotypes
177 obtained by targeted long-read sequencing (S1 Table and S2 Table, Supporting Information).
178 As far as possible, we use the same accession designations as in previous studies. All
179 accession information is listed in S1 Table, Supporting Information.

180 BAC clones were sequenced to high coverage (150-400x) using PacBio SMRT
181 sequencing (S2 Table, Supporting Information). Short-read sequencing data for all BACs
182 were generated on an Illumina MiSeq (>380 x; S2 Table, Supporting Information) and used
183 for indel error correction as in Bachmann et al. (2018). All sequencing was done at the
184 SciLifeLab National Genomics Infrastructure in Uppsala, Sweden. Sequences were assembled
185 in HGAP3.0 (Chin et al., 2013), except for the *S*-haplotype of Cg88.15, for which Canu v.1.7
186 (Koren et al., 2017) was used.

187 We annotated our *S*-locus assemblies as in Bachmann et al. (2018). Briefly, we used
188 Augustus v3.2.3 (Stanke et al., 2004) and RepeatMasker v4.0.7;
189 http://www.repeatmasker.org), run via Maker v2.31.9 (Holt & Yandell, 2011) with
190 *Arabidopsis thaliana* as a model prediction species and using protein homology data for *SRK*,
191 *U-box* and *ARK3* from *Arabidopsis lyrata* and *Arabidopsis halleri*. Due to high levels of
192 sequence diversity at the key *S*-locus genes *SRK* and *SCR*, they were difficult to annotate
193 automatically. Sequence similarity to known *SRK* exon 1 sequences was used to accept
194 candidate loci as *SRK*, while we used similarity to *ARK3* as a rejection criterion. To annotate
195 *SCR*, we used a window-based approach to screen for the characteristic pattern of cysteine
196 residues after translation of the DNA sequence in all three frames (Bachmann et al., 2018).
197 Using this approach, we identified a region highly similar to *A. halleri SCR* in *S*-locus
198 haplotype *S12* (GenBank accession number KJ772374.1) in our *C. orientalis S*-locus BAC
199 sequences.
200
201 **Phylogenetic analyses of *S*-locus sequences**
202 To examine the phylogenetic placement of the *S*-haplotypes sequenced here, we used a
203 dataset of Brassicaceae *SRK* exon 1 and *ARK3* sequences downloaded from Genbank as
204 described in Bachmann et al. (2018) and generated an alignment of *SRK* exon 1 sequences
205 using the MAFFT v7.245 & E-INS-I algorithm (Katoh et al., 2002), with manual curation in
206 SeaView v4.6 (Gouy et al., 2010). We generated a maximum likelihood phylogenetic tree
207 from the *SRK* alignment with RaXMl v8.2.3. In this phylogeny, the *C. grandiflora S*-
208 haplotype from accession Cg2-2 clustered with the *S*-haplotypes of *C. orientalis* and the *C.*
209 *orientalis*-derived subgenome of *C. bursa-pastoris* (i.e. the *C. bursa-pastoris* B subgenome).
210 Due to the high sequence similarity (93.4% protein sequence identity at *SRK*) of the Cg2-2 *C.*
211 *grandiflora S*-haplotype to *A. halleri S12* (GenBank accession number KJ772374.1) we

212 hereafter term this *S*-haplotype *CgS12.* We assessed sequence conservation across the entire
213 ~100 kbp *S*-locus by aligning *S*-locus sequences using LASTZ v1.03.54 (Harris, 2007) and
214 calculating pairwise sequence conservation in 250 bp sliding windows.
215

216 **Candidate mutations for the loss of SI in *C. orientalis***

217 To identify candidate causal mutations for the loss of SI in *C. orientalis*, we analyzed
218 sequence alignments of the two key *S*-locus genes *SRK* and *SCR*, as well as the *S*-linked *U-*
219 *box* gene, which may act as a modifier of the SI response (Liu et al., 2007). Specifically, we
220 searched for major-effect variants resulting in frameshifts, premature stop codons or non-
221 consensus splice sites, present in sequences from the SC *C. orientalis* and/or in the SC *C.*
222 *bursa-pastoris* B subgenome, which is derived from *C. orientalis* (Douglas et al., 2015), but
223 not in sequences from the same haplogroup found in the SI species *C. grandiflora* and *A.*
224 *halleri.* For *SRK* we identified nonsynonymous changes in hypervariable regions important
225 for SRK specificity (Nishio & Kusaba 2000; Kusaba et al. 1997), based on a protein sequence
226 alignment of *Capsella* SRK with *Brassica rapa* SRK9 which represents a different
227 haplogroup, but whose protein structure and interaction with SCR has been resolved recently
228 (Ma et al., 2016).
229

230 **Bioinformatic processing of RNAseq data and expression of *S*-locus genes in *C. orientalis***

231 RNAseq data were trimmed with Trimmomatic v.0.36 (Bolger et al. 2014) and reads mapped
232 using STAR v.2.2.1 (Dobin et al., 2013). For small RNA sequencing, we mapped reads of
233 length 18-27 nt using STAR v.2.2.1. Expression was quantified as RPKM (number of reads
234 per kb per million mapped reads; Mortazavi et al., 2008).

235 To assess whether *SRK, SCR* and *U-box* were expressed in *C. orientalis* flower buds,
236 we generated RNAseq data from mixed-stage flower buds of two *C. orientalis* accessions (S1

237 Table, Supporting Information) as previously described (Steige et al., 2017). For comparison,
238 we also generated RNAseq data from leaf samples from the same individuals. Trimmed reads
239 were mapped to a modified v1.0 reference *C. rubella* assembly (Slotte et al., 2013), where the
240 *S*-locus region (scaffold_7 7523601:7562919) was masked and our *S*-locus assembly from *C.*
241 *orientalis* Co1719/11 was added. We conducted qualitative RT-PCR with specific primers to
242 *SCR* in *C. orientalis* and *C. grandiflora CgS12*, to assess the expression of *SCR* in flower
243 buds of both *C. orientalis* accessions, as well as in three *C. grandiflora* individuals harboring
244 *CgS12* (Methods S1, Supporting Information).
245
246 **Assessing the functionality of *C. orientalis* SCR by interspecific crosses**
247 We performed controlled crosses to verify that *C. grandiflora CgS12* conferred SI, and to
248 assess the functionality of SCR in *C. orientalis*. To verify functional SI in *C. grandiflora*
249 carrying *CgS12,* we performed a total of 24 manual self-pollinations of four *C. grandiflora*
250 individuals carrying the *CgS12 S*-haplotype. While the identity of the other *S*-haplotype in
251 these individuals is unknown and we were unable to identify it using PCR-based screening,
252 we verified expression of *CgSCR12*, indicating that the other *S*-allele is not dominant over
253 *CgS12* at the pollen level. We assessed the success of manual self-pollination of *C. orientalis*
254 by performing 6-12 manual self-pollinations of each of three accessions (Table S1,
255 Supporting Information). To assess whether *C. orientalis* SCR is functional, we crossed *C.*
256 *grandiflora* harboring *CgS12* as a seed parent to *C. orientalis* as a pollen donor. We
257 performed a total of 144 crosses of this type, with three different *C. orientalis* accessions as
258 pollen donors and six different *CgS12*-carrying *C. grandiflora* individuals as seed parents
259 (Table S1, Supporting Information). If *C. orientalis* SCR is functional, and provided that
260 *CgS12 SRK* is expressed, then we expect this cross to be incompatible, whereas if *C.*
261 *orientalis* SCR is nonfunctional, the cross should be compatible. The reciprocal cross of the

262 same individuals was also carried out with the same accessions (total 104 crosses of this
263 type), to test whether female SI specificity is functional in *C. orientalis*. Finally, we
264 performed 12 crosses of *C. grandiflora* harboring other *S*-haplotypes to *C. grandiflora*
265 harboring *CgS12*, and 12 to *C. orientalis*. These crosses are expected to be successful.

266 We observed pollen tube growth in the pistil 12 hours after pollination. Pistils were
267 fixed in EtOH: acetic acid 9:1 for > 2 hours, softened in 1N NaOH 60°C for 20 minutes and
268 stained with 0.01% decolorised aniline blue in 2% solution of K_3P0_4 for 2 hours. Pollen tubes
269 were visualised by mounting pistils on a microscope slide which was examined under an
270 epifluorescence microscope (Zeiss Axiovert 200M). We compared the number of pollen tubes
271 among different types of crosses using a Kruskal-Wallis test.

272

273 **The role of small RNA-based dominance modifiers for dominance of SC**

274 To test whether dominant expression of SC could be mediated by small RNA-based
275 dominance modifiers, we conducted additional sequence and expression analyses, using a
276 strategy similar to that of Nasrallah et al. (2007). We identified a region in our *C. orientalis S*-
277 haplotypes with high sequence similarity (91.3%) to the *A. halleri S12* small RNA precursor
278 *Ah12mirS3* identified previously (Durand et al., 2014). We generated small RNA and RNA
279 sequencing data from flower buds of 19 F2s, representing all three *S*-locus genotypes in our
280 F2 mapping population (12 heterozygotes, 4 and 3 individuals homozygous for the *C.*
281 *orientalis* or the *C. grandiflora S*-haplotype, respectively). We quantified expression of
282 sRNAs in the *Ah12mirS3*-like sRNA precursor region (hereafter termed *ComirS3* sRNAs) and
283 tested whether *ComirS3* sRNAs were expressed specifically in F2s with a *C. orientalis S*-
284 allele.

285 To test whether *C. grandiflora SCR* was repressed in F2s heterozygous at the *S*-locus
286 we quantified the expression of *C. orientalis* and *C. grandiflora SCR* in our F2s. We mapped

287 F2 RNAseq reads from flower buds to a modified *C. rubella* reference containing both the
288 Co1719/11 *S*-haplotype and the *C. grandiflora* Cg88.15 *S*-haplotype segregating in our F2
289 population, and quantified the expression of *C. orientalis* and *C. grandiflora SCR* in all three
290 genotypes, respectively.

291 To identify targets of *ComirS3* small RNAs we took all 18-27 nt *ComirS3* sRNAs and
292 searched for small RNA targets within 1 kb of *SCR* of the *C. grandiflora* Cg88.15 *S*-
293 haplotype. Small RNA targets were identified using a Smith & Waterman algorithm (Smith &
294 Waterman, 1981) with scoring matrix: match=01, mismatch=-1, gap=-2, G:U wobble=-0.5 as
295 previously described (Durand et al., 2014).

296

297 **Timing of loss of SI in *C. orientalis***

298 To assess whether major-effect mutations at the *S*-locus were fixed in *C. orientalis*, we
299 analyzed whole-genome resequencing data from additional *C. orientalis* accessions, in total
300 covering 30 accessions from 18 populations and including publicly available *C. orientalis*
301 genome resequencing data (Douglas et al., 2015; Huang et al., 2018; Koenig et al., 2019)
302 (Table S1, Supporting Information). We mapped trimmed data to a *C. rubella* reference
303 modified to include the *C. orientalis* haplotype of accession Co1719/11 using BWA-MEM
304 (Li, 2013) and used GATK 3.8 (McKenna et al., 2010; DePristo et al., 2011; Van der Auwera
305 et al., 2013) Unified Genotyper with the option --output_mode
306 EMIT_ALL_CONFIDENT_SITES to call all sites. We filtered sites following GATK
307 recommended hard filtering with the following parameters; QD < 2.0 || FS > 60.0 || MQ <
308 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0. We required a minimum read depth
309 of 15 and a maximum of 200. Finally, we scored the presence or absence of major-effect
310 mutations at the *S*-locus in our samples. Because *C. orientalis* is highly homozygous, SC, and
311 has low levels of polymorphism genome-wide (Douglas et al., 2015), this approach is

312 expected to work well, as long as a *C. orientalis S*-haplotype is included in the reference
313 genome.
314 We used a strategy similar to that in Guo et al. (2009) to estimate a lower and upper
315 bound of the timing of the loss of SI in *C. orientalis.* We obtained a lower bound for the
316 timing of the loss of SI by estimating the time to the most recent common ancestor (TMRCA)
317 based on full-length *C. orientalis* and *C. bursa-pastoris* B *S*-locus sequences. Genome-wide
318 haplotype sharing between *C. orientalis* and the *C. bursa-pastoris* B subgenome suggests that
319 the ancestor of *C. orientalis* that contributed to formation of *C. bursa-pastoris* was self-
320 compatible (Douglas et al., 2015) and including *C. bursa-pastoris* B sequences can thus
321 increase the precision of our estimates. To obtain an upper bound for the timing of the loss of
322 SI we estimated the TMRCA for *C. orientalis*, *C. bursa-pastoris* B and *C. grandiflora CgS12.*
323 For analyses of the timing of loss of SI, our final alignment contained 37 *S*-locus
324 sequences including the *C. grandiflora* ancestral *S*-haplotype (*CgS12*), 4 *C. bursa-pastoris*
325 subgenome B *S*-haplotypes and *S*-haplotype data for 32 *C. orientalis* individuals (Supporting
326 Information). Sequences were aligned using block alignment using Muscle v.3.8.31 (Edgar
327 2004) as implemented in AliView v.1.20 (Larsson 2014). The total length of the *S*-locus
328 alignment was 33,485 bp, 22,689 bp had indels in at least one sequence, 9,835 sites were
329 invariant and 876 sites were polymorphic. The alignment was partitioned into coding and
330 non-coding regions and sites with indels and missing data were pruned in further analysis.
331 We estimated the timing of the splits between *C. grandiflora*, *C. bursa-pastoris* and *C.*
332 *orientalis* using a strict molecular clock in a Bayesian framework in BEAST2 (Bouckaert et
333 al. 2014). We used a fixed clock rate assuming a mutation rate of $7x10^{-9}$ substitutions per site
334 per generation (Ossowski et al., 2010) and a generation time of one year. We ran both a model
335 with exponential changes in population size and a model with a constant population size, and
336 compared models using Akaike's information criterion through Markov chain Monte Carlo,

337 AICM (Baele et al. 2012) (Methods S1, Supporting Information). We ran two chains of 10
338 millions generations sampled every 1000 generations and checked convergence by visual
339 inspection of the log-likelihood profile and assuring an effective sample size (ESS) value
340 above 200. The posterior distribution of trees was used to build a maximum clade credibility
341 tree and estimate node age and 95% confidence interval using TreeAnnotator (Drummond et
342 al. 2012).
343

344 **Results**

345 **SC maps to the *S*-locus as a dominant trait**

346 We first asked whether loss of SI in *C. orientalis* maps to the canonical Brassicaceae *S*-locus.
347 We therefore generated an F2 mapping population by crossing SC *C. orientalis* to a SI *C.*
348 *grandiflora* accession. Interspecific F1 individuals were SC, indicating that SC is dominant.
349 Our F2 mapping population segregated for SC, and we detected a single, significant
350 ($P<0.001$) quantitative trait locus (QTL) for this trait, based on 304 F2 individuals genotyped
351 at 549 markers (Fig. 1a; Fig. S1, Supporting Information). The credible interval for this QTL
352 includes the *S*-locus on chromosome 7 (Fig. 1a), and SC was dominant over SI (Fig. 1b). SC
353 in *C. orientalis* thus maps as a dominant trait to a region encompassing the *S*-locus.
354

355 **Sequencing the S-haplotype of *C. orientalis* and a highly similar but functional *S*-**
356 **haplotype from *C. grandiflora***

357 We next sought to identify candidate causal loss-of-function mutations at the *C. orientalis S*-
358 locus. For this purpose, we assembled full-length *S*-haplotype sequences of two *C. orientalis*
359 accessions based on long-read sequencing of BACs (Tables S1-S2, Supporting Information).
360 To facilitate identification of candidate mutations for the loss of SI, it is beneficial to be able
361 to contrast functional and non-functional *S*-haplotypes that belong to the same *S*-haplogroup

362 and ancestrally shared the same SI specificity. Here, we identified and sequenced a functional
363 *C. grandiflora S*-haplotype (for details, see Materials and Methods), which had 98.3% protein
364 sequence identity at *SRK* to that of *C. orientalis* (Fig. 2a-c, Table S3, Supporting
365 Information). According to criteria used in outcrossing *Arabidopsis* species (Castric et al.,
366 2008; Chantha et al., 2013), this *C. grandiflora* haplotype is expected to represent the same SI
367 specificity as that of *C. orientalis*. This *C. grandiflora S*-haplotype is also similar (93.4%
368 protein sequence identity at *SRK*) to the functional *Arabidopsis halleri S12* haplotype (Durand
369 et al., 2014) (Fig. 2a-b, Fig. S2, Supporting Information), and we therefore designate it
370 *CgS12*. Sequence similarity between *CgS12* and *C. orientalis* is not limited to *SRK*, as other
371 *S*-linked genes showed the same phylogenetic topology as those for *SRK* (Fig. 2b), and there
372 were peaks of sequence conservation between *CgS12* and *C. orientalis* in both genic and
373 intergenic parts of the *S*-locus (Fig. 2c). We verified that *C. grandiflora* individuals with
374 *CgS12* expressed *CgSCR12* and were SI by scoring pollen tube germination after controlled
375 self-pollination (Fig. 3a, Table S4, Fig. S3-S5, Supporting Information).
376

377 **A frameshift deletion in the male specificity gene *SCR* is fixed in *C. orientalis***

378 By comparing *S*-haplotype sequences from *C. orientalis* (SC) to *C. grandiflora CgS12* and *A.*
379 *halleri S12* (both SI), we identified a single-base frameshift deletion in the *SCR* coding
380 sequence of *C. orientalis* (Fig. 2d). This frameshift is predicted to result in loss of 5 out of 8
381 conserved cysteine residues essential to the function of SCR (Fig. 2e), likely resulting in loss
382 of male specificity. To assess whether the deletion was fixed in *C. orientalis*, as we would
383 expect for mutations that spread early during the transition to selfing, we analyzed whole-
384 genome resequencing data from additional *C. orientalis* accessions (S1 Table, Supporting
385 Information). We found that the *SCR* frameshift deletion was fixed across 32 samples of *C.*
386 *orientalis* from 18 populations, consistent with expectations if the deletion was fixed in

387 association with the loss of SI. The same deletion was found in *SCR* of the *C. bursa-pastoris*
388 B subgenome, which is derived from *C. orientalis* (Fig. 2d, Fig. 2e). This suggests that *C.*
389 *orientalis* was self-compatible when it contributed to the origin of the allotetraploid *C. bursa-*
390 *pastoris*.

391 In contrast to *SCR*, we observed no major loss-of-function mutations in *C. orientalis*
392 *SRK* or at the *S*-linked *U-box* gene, which may modify the female SI response (Liu et al.,
393 2007). There were two nonsynonymous substitutions in *C. orientalis SRK* that were likely
394 located within hypervariable regions of SRK (Fig. S6, Supporting Information). However,
395 without resolving the detailed protein structure of the CgS12 SRK/SCR complex the exact
396 consequences of these nonsynonymous changes cannot be determined. Finally, *SRK, U-box*
397 and the truncated version of *SCR* are all expressed in flower buds of *C. orientalis* (Table S4,
398 Fig. S5, Supporting Information) and we currently cannot rule out that more subtle changes to
399 their sequence or expression affect their function.
400
401 **Assessment of SI specificity**
402 To assess whether male SI specificity is degenerated in *C. orientalis*, as we expect if SCR is
403 nonfunctional, we crossed *C. orientalis* to *C. grandiflora* individuals harboring *CgS12*, which
404 likely ancestrally shared the same SI specificity (Fig. 2). As expected if the frameshift
405 deletion impaired the function of *SCR*, pollen from *C. orientalis* successfully germinated on
406 the stigma of *C. grandiflora* individuals harboring *CgS12* (Fig. 3, Fig. S3-S4, Supporting
407 Information). However, we also found evidence for degeneration of female specificity in *C.*
408 *orientalis*, as pollen from *C. grandiflora* harboring *CgS12* germinated on the *C. orientalis*
409 stigma (Fig. 3; Fig. S3-S4, Supporting Information). Similar results were obtained for crosses
410 with *C. orientalis* accessions from three different populations (Fig. S4, Supporting
411 Information).

412

413 **A conserved *S*-linked sRNA is associated with dominant expression of *C. orientalis SCR***

414 Under most circumstances, loss of function mutations are predicted to be recessive, as a single
415 copy of a functional allele is often sufficient to result in a complete phenotype (Kacser &
416 Burns, 1981). Here, SC is associated with a frameshift deletion at *SCR*, yet it is dominant in
417 our F2s. Hence, we investigated whether the small RNA-based mechanism that governs
418 dominance hierarchies among *S*-alleles in *Arabidopsis* (Durand et al., 2014) could also
419 explain dominance of SC in our case. Specifically, if the *C. orientalis S*-haplotype encodes a
420 trans-acting sRNA that represses expression of *C. grandiflora SCR* in *S*-locus heterozygotes,
421 SC could be dominant even if it is due to a loss of function mutation in *C. orientalis SCR*.

422 In *A. halleri,* the *S12* haplotype belongs to the second most dominant class of *S*-alleles
423 and harbors an *S*-linked sRNA-based dominance modifier termed *Ah12mirS3* (Durand et al.,
424 2014). In *C. orientalis,* we found the corresponding *mirS3* sRNA precursor region to be
425 conserved (91.3% sequence identity) (Fig. 4a, Fig. S2, Supporting Information). The region
426 harboring the *mirS3* sRNA precursor was also conserved between *C. grandiflora CgS12* and
427 *C. orientalis* (Fig. 2c). To assess whether expression of *C. orientalis Ah12mirS3*-like sRNA
428 (*ComirS3*) was associated with repression of the *C. grandiflora SCR* allele passed on in our
429 cross through the F1 plant, we sequenced and assembled the *C. grandiflora S*-haplotype
430 segregating in our F2 population, and analyzed *SCR* and sRNA expression in flower buds of
431 19 F2s representing all three possible *S*-locus genotypes. We detected expression of *ComirS3*
432 sRNAs (Fig. 4a) in F2s harboring the *C. orientalis S*-haplotype, but not in *C. grandiflora S*-
433 homozygotes (Fig. 4b). The most abundant *ComirS3* sRNA was highly similar to the
434 *Ah12mirS3* sRNA and had a predicted target within the intron of *C. grandiflora SCR* allele
435 (Fig. 4c). The sRNA-target affinity was similar to that of functional *Arabidopsis* dominance
436 modifiers (Durand et al., 2014; Burghraeve et al., 2018). As expected if *ComirS3* sRNAs

437 silence *C. grandiflora SCR*, *C. grandiflora SCR* was specifically downregulated in *S*-locus
438 heterozygotes (Fig. 4d). Our F2 *S*-locus heterozygotes thus only express the truncated *C.*
439 *orientalis SCR* at detectable levels. These results are consistent with *S*-linked sRNAs
440 conferring dominance of the SC *C. orientalis S*-haplotype through transcriptional silencing of
441 recessive *SCR* alleles.
442
443 **Timing of loss of SI in *C. orientalis***
444 The timing of loss of SI can be estimated based on polymorphism accumulated at the *S*-locus
445 after loss of SI (Guo et al., 2009). We analyzed 37 full-length *S*-locus sequences and
446 estimated an upper bound for the timing of loss of SI in *C. orientalis* as the time to the most
447 recent common ancestor (TMRCA) of *C. orientalis, C. bursa-pastoris* B and *C. grandiflora*
448 *CgS12 S*-haplotypes. Based on these analyses, we infer an upper bound of the timing loss of
449 SI in *C. orientalis* at 2.6 Mya (2.2-2.9 Mya, 95% CI) and a lower bound at 70 kya (50-100
450 kya, 95% CI) (Fig. 5, Table S5, Supporting Information) under an exponential population size
451 change model. Very similar estimates were obtained under a constant population size model
452 and after subsampling the *C. orientalis* accessions to obtain a scattered sample (S5 Table,
453 Supporting Information). Our timing estimates thus appear to be robust to sampling strategy
454 and assumptions regarding population size changes.
455
456 ## Discussion
457 Here, we show that loss of SI in *C. orientalis* maps as a dominant trait to the *S*-locus. This
458 result is consistent with the theoretical prediction that *S*-linked mutations should often
459 contribute to the loss of SI (Porcher & Lande, 2005). We identify candidate mutations for the
460 loss of SI, including a frameshift deletion in the male specificity gene *SCR.* Our finding that
461 SC is dominant agrees with Haldane's prediction that dominant alleles enjoy a higher fixation

462 probability in outcrossers (Haldane 1927). Finally, we identify an sRNA that could be
463 responsible for dominance of SC and that is conserved between *Capsella* and *Arabidopsis*
464 *halleri.*
465 Theory predicts that mutations that disrupt male SI specificity should be more strongly
466 selected for during the transition to selfing than those that disrupt female SI specificity
467 (Uyenoyama et al., 2001; Busch & Schoen, 2008; Tsuchimatsu & Shimizu, 2013). Indeed,
468 mutations that disrupt male SI specificity should have an advantage both when spreading
469 through seeds and pollen, because they avoid recognition and rejection when they spread
470 through outcross pollen (Uyenoyama et al., 2001; Busch & Schoen, 2008; Tsuchimatsu &
471 Shimizu, 2013). In contrast, mutations that disrupt female specificity only have an advantage
472 over those that disrupt male specificity when there is pollen limitation of seed set, i.e. reduced
473 reproductive success due to inadequate quantity or quality of pollen (Uyenoyama et al., 2001;
474 Busch & Schoen, 2008; Tsuchimatsu & Shimizu, 2013). The *C. orientalis SCR* deletion is
475 expected to lead to the loss of 5 of 8 conserved cysteine residues in the SCR protein, which
476 could cause loss of male SI specificity. The *SCR* deletion was fixed in a broad sample of *C.*
477 *orientalis*, as we would expect if it arose early during the transition to selfing. It was also
478 found in the allopolyploid *C. bursa-pastoris*, suggesting that the shift to SC in *C. orientalis*
479 predated the origin of *C. bursa-pastoris*. Through crosses between *C. orientalis* and *C.*
480 *grandiflora* individuals harboring highly similar *S*-haplotypes, we confirmed that male SI
481 specificity was lost in *C. orientalis*, as the pollen of *C. orientalis* germinated on the stigma of
482 individuals harboring the highly similar but functional *CgS12* haplotype. However, we cannot
483 strictly rule out a contribution of *S*-linked mutations that disrupt female SI specificity to the
484 loss of SI in *C. orientalis,* as our controlled crosses indicated that female SI specificity was
485 also impaired in *C. orientalis*. We identified two fixed nonsynonymous substitutions in likely
486 functionally important regions of *SRK* that might have contributed to the breakdown of

487 female SI specificity in *C. orientalis*. However, without further work it is difficult to predict
488 the functional consequences of these nonsynonymous substitutions. One scenario that would
489 be consistent with our crossing results is one where the *C. grandiflora CgS12 S*-haplotype
490 represents a different SI specificity than that of *C. orientalis*. Due to the very high sequence
491 similarity between these *S*-haplotypes, we consider this unlikely. Instead, we believe that our
492 crossing results illustrate a general challenge for studies that aim to identify causal changes
493 for the loss of SI. Indeed, after SI has been lost, additional mutations that impair the function
494 of *S*-locus genes can accumulate without cost, unless there are pleiotropic constraints.
495 Ancestral reconstruction would be the only way to tease apart the role of these individual
496 mutations to the breakdown of SI (Tsuchimatsu et al., 2010).

497 Information on the timing of loss of SI is currently available for less than a handful of
498 Brassicaceae systems (e.g. Guo et al. 2009; Busch et al. 2011; Tsuchimatsu et al., 2012).
499 Accurately estimating bounds for the timing of loss of SI is challenging, as it requires
500 identifying and sequencing shared *S*-haplotypes in closely related SI and SC species. Here, we
501 identify shared *S*-haplotypes in the SI *C. grandiflora* and the SC *C. orientalis.* We estimate
502 that the loss of SI in *C. orientalis* occurred between 2.6 Mya and 70 kya, based on TMRCA
503 analyses of full-length *S*-haplotypes. While our estimates cover a broad range of times, we
504 argue that the most likely estimate of the timing of loss of SI is probably closer to the upper
505 bound, 2.6 Mya. For instance, in comparison to the recently derived selfer *C. rubella*, *C.*
506 *orientalis* has strongly reduced genome-wide polymorphism levels (Douglas et al., 2015;
507 Koenig et al. 2019), shows increased reproductive isolation through endosperm development
508 defects in crosses to *C. grandiflora* (Lafon-Placette et al., 2018), and possibly exhibits a lower
509 genomic content of transposable elements (Ågren et al., 2014). An older origin of selfing in
510 *C. orientalis* than in *C. rubella* would be compatible with these findings, as selfing is
511 expected to result in reduced polymorphism genome-wide and affect TE content (Wright et

512 al., 2013). While the shift to SC was clearly independent in *C. orientalis* and *C. rubella*,
513 which harbor different *S*-haplotypes (Fig. 2a), both transitions involved fixation of a single *S*-
514 haplotype (Guo et al., 2009; Slotte et al., 2012). These scenarios contrast with the situation in
515 *A. thaliana*, where multiple *S*-haplogroups are still segregating (Durvasula et al., 2017;
516 Tsuchimatsu et al., 2017). Our study thus contributes to an improved understanding of the
517 timing and mode of loss of SI in a system that is widely used for genomic studies.

518 Population geneticists have long predicted that dominant beneficial mutations should
519 have a higher fixation probability than recessive ones (Haldane, 1927), a phenomenon termed
520 "Haldane's sieve". Our finding that SC is dominant over SI is consistent with this prediction,
521 and agrees with results for several other wild Brassicaceae species (e.g. *L. alabamica*; Busch
522 et al., 2011, *A. kamchatica*; Tsuchimatsu et al., 2012, *C. rubella*; Nasrallah et al., 2007; Slotte
523 et al., 2012). However, not all transitions involve dominant *S*-haplotypes, and for instance in
524 *A. lyrata*, a transition involving a recessive loss of SI has recently been documented (Mable et
525 al., 2017). Our results further suggest that a small RNA-based mechanism could explain
526 dominance of SC. If this is the case, dominance of the SC phenotype will depend on the exact
527 combination of *S*-alleles and their position in the dominance hierarchy. Interestingly, in both
528 *C. orientalis* and *C. rubella*, SC is linked to relatively dominant *S*-haplotypes. Taken together,
529 these findings suggest that dominant SC mutations on average have an advantage over
530 recessive mutations, at least early during the transition to selfing. Thus, the lower population
531 frequencies or higher *S*-linked load (Llaurens et al., 2009) of dominant *S*-alleles do not
532 prevent mutations in such alleles from contributing to recurrent loss of SI.

533

534 ## Acknowledgements

535 We thank Timothy Paape and Matthew Hartfield for helpful discussion, Daniel Koenig and
536 Detlef Weigel for making *C. orientalis* resequencing data publicly available prior to

537 publication, Cindy Canton for help with plant care, Olga Vinnere Pettersson and Christian
538 Tellgren-Roth for help with long-read sequencing and BAC assembly. The authors
539 acknowledge support from the National Genomics Infrastructure (NGI) / Uppsala Genome
540 Center / SNP&SEQ Technology Platform and UPPMAX for assistance in massive parallel
541 sequencing and computational infrastructure. Work at Uppsala Genome Center is funded by
542 RFI / VR and Science for Life Laboratory, Sweden. The SNP&SEQ Platform is supported by
543 the Swedish Research Council and the Knut and Alice Wallenberg Foundation. V.C.
544 acknowledges support by a grant from the European Research Council (NOVEL project,
545 grant #648321). The authors thank the French Ministère de l'Enseignement Supérieur et de la
546 Recherche, the Hauts de France Region and the European Funds for Regional Economical
547 Development for their financial support to this project. This work was supported by a grant
548 from the Swedish Research Council (grant #D0432001) and by a grant from the Science for
549 Life Laboratory, Swedish Biodiversity Program to T.S. The Swedish Biodiversity Program is
550 supported by the Knut and Alice Wallenberg Foundation.
551

552 **Author contributions**

553 T.S. designed the experiments. J.B., A.T., C.L.-P., K.A.S., C.C. and W.M. performed the
554 experiments, J.B., A.T. and A.D. generated the data. J.B. analyzed sRNA expression and
555 targets, A.T. analyzed and annotated *S*-locus BACs, B.L. analyzed and annotated *S*-locus
556 BACs and performed BEAST analyses, M.F. analyzed QTL mapping and expression data,
557 B.N. contributed reagents/materials/analysis tools, and A.D. generated full-length *S*-locus
558 alignments, H.B. supervised the work of C.C. and W.M., C.K. supervised the work of C.L.-P.,
559 V.C. supervised the work of J.B, T.S. supervised the work of A.D., A.T., B.L., J.B., K.A.S.
560 and M.F. All authors contributed to the writing of the paper.
561

562 ## References

563 **Bachmann JA, Tedder A, Laenen B, Steige KA, Slotte T. 2018.** Targeted long-read
564 sequencing of a locus under long-term balancing selection in *Capsella*. G3 8: 1327-1333.

565 **Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. 2012.** Accurate model selection
566 of relaxed molecular clocks in Bayesian phylogenetics. Molecular Biology and
567 Evolution 30:239-243.

568 **Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina
569 sequence data. Bioinformatics. 30:2114–2120.

570 **Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut
571 A, Drummond AJ. 2014.** BEAST 2: a software platform for Bayesian evolutionary analysis.
572 PLoS Computational Biology 10:e1003537.

573 **Brandvain Y, Slotte T, Hazzouri KM, Wright SI, Coop G. 2013.** Genomic identification of
574 founding haplotypes reveals the history of the selfing species *Capsella rubella*. PLos Genetics
575 9: e1003754.

576 **Broman KW, Wu H, Sen S, Churchill GA. 2003.** R/qtl: QTL mapping in experimental
577 crosses. Bioinformatics 19: 889-890.

578 **Busch JW, Schoen DJ. 2008.** The evolution of self-incompatibility when mates are limiting.
579 Trends in Plant Science 13: 128–136.

580 **Busch JW, Joly S, Schoen DJ. 2011.** Demographic signatures accompanying the evolution
581 of selfing in *Leavenworthia alabamic*a. Molecular Biology and Evolution 28: 1717–1729.

582 **Burghgraeve N, Simon S, Barral S, Fobis-Loisy I, Holl AC, Ponitzki C, Schmitt C,
583 Vekemans X, Castric V. 2018.** Base-pairing requirements for small RNA-mediated gene

584 silencing of recessive self-incompatibility alleles in *Arabidopsis halleri*. BioRxiv doi:
585 10.1101/370239

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

588 **Castric V, Bechsgaard J, Schierup MH, Vekemans X. 2008.** Repeated adaptive
589 introgression at a gene under multiallelic balancing selection. PLoS Genetics 4: e1000168.

590 **Chantha S-C, Herman AC, Platts AE, Vekemans X, Schoen DJ. 2013.** Secondary
591 evolution of a self-incompatibility locus in the Brassicaceae genus *Leavenworthia*. PLoS
592 Biology 11:e1001560.

593 **Charlesworth D. 2006.** Evolution of plant breeding systems. Current Biology 16: R726–35.

594 **Charlesworth D. & Charlesworth B. 1979.** The evolution and breakdown of *S*-allele
595 systems. Heredity. 43: 41-55.

596 **Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A,**
597 **Copeland A, Huddleston J, Eichler EE, et al. 2013.** Nonhybrid, finished microbial genome
598 assemblies from long-read SMRT sequencing data. Nature Methods. 10: 563–569.

599 **Darwin C. 1876.** The effects of cross and self fertilisation in the vegetable kingdom. London:
600 John Murray.

601 **Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,**
602 **Gingeras TR. 2013.** STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29: 15–21.

603 **de Nettancourt D. 2001.** Incompatibility and Incongruity in Wild and Cultivated Plants.
604 Berlin: Springer.

605 **DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA,**
606 **del Angel G, Rivas MA, Hanna M, et al. 2011.** A framework for variation discovery and
607 genotyping using next-generation DNA sequencing data. Nature Genetics 43: 491–498.

608 **Douglas GM, Gos G, Steige KA, Salcedo A, Holm K, Josephs EB, Arunkumar R, Agren**
609 **JA, Hazzouri KM, Wang W, et al. 2015.** Hybrid origins and the earliest stages of
610 diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. Proceedings
611 of the National Academy of Sciences of the USA 112: 2806-2811.

612 **Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012.** Bayesian phylogenetics with
613 BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29:1969–1973.

614 **Durand E, Méheust R, Soucaze M, Goubet PM, Gallina S, Poux C, Fobis-Loisy I,**
615 **Guillon E, Gaude T, Sarazin A, et al. 2014.** Dominance hierarchy arising from the evolution
616 of a complex small RNA regulatory network. Science 346: 1200–1205.

617 **Edgar RC. 2004.** MUSCLE: multiple sequence alignment with high accuracy and high
618 throughput. Nucleic Acids Research 32:1792–1797.

619 **Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, Tsuchimatsu**
620 **T, Burbano HA, Picó FX, Alonso-Blanco C, et al. 2017.** African genomes illuminate the
621 early history and transition to selfing in *Arabidopsis thaliana*. Proceedings of the National
622 Academy of Sciences of the USA 114: 5213–5218.

623 **Eckert C, Samis K, Dart S. 2006.** Reproductive assurance and the evolution of uniparental
624 reproduction in flowering plants. In: Harder L, Barrett S, eds. *Ecology and evolution of*
625 *flowers*. Oxford: Oxford University Press.

626 **Fisher RA. 1941.** Average excess and average effect of a gene substitution. Annals of Human

627 Genetics 11: 53–63.

628 **Foxe JP, Slotte T, Stahl EA, Neuffer B, Hurka H, Wright SI. 2009.** Recent speciation
629 associated with the evolution of selfing in *Capsella*. Proceedings of the National Academy of
630 Sciences of the USA 106: 5241–5245.

631 **Gos G, Slotte T, Wright SI. 2012.** Signatures of balancing selection are maintained at
632 disease resistance loci following mating system evolution and a population bottleneck in the
633 genus *Capsella*. BMC Evolutionary Biology 12: 152.

634 **Gouy M, Guindon S, Gascuel O. 2010.** SeaView version 4: A multiplatform graphical user
635 interface for sequence alignment and phylogenetic tree building. Molecular Biology and
636 Evolution 27: 221–224.

637 **Guo Y-L, Bechsgaard JS, Slotte T, Neuffer B, Lascoux M, Weigel D, Schierup MH.**
638 **2009.** Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss
639 of self-incompatibility and an extreme bottleneck. Proceedings of the National Academy of
640 Sciences of the USA 106: 5246–5251.

641 **Haldane JBS. 1927.** A mathematical theory of natural and artificial selection, part V:
642 selection and mutation. Mathematical Proceedings of the Cambridge Philosophical Society
643 23: 838–844.

644 **Haley CS, Knott SA. 1992.** A simple regression method for mapping quantitative trait loci in
645 line crosses using flanking markers. Heredity. 69: 315–324.

646 **Harris, R.S. 2007.** Improved pairwise alignment of genomic DNA. PhD Thesis,
647 Pennsylvania State University, Pennsylvania, USA.

648 **Holt C, Yandell M. 2011.** MAKER2: an annotation pipeline and genome-database

649 management tool for second-generation genome projects. BMC Bioinformatics. 12: 491.

650 **Huang H-R, Liu J-J, Xu Y, Lascoux M, Ge X-J, Wright SI. 2018.** Homeologue-specific
651 expression divergence in the recently formed tetraploid *Capsella bursa-pastoris*
652 (Brassicaceae). The New Phytologist 42: e46–635.

653 **Hurka H, Friesen N, German DA, Franzke A, Neuffer B. 2012.** 'Missing link' species
654 *Capsella orientalis* and *Capsella thracica* elucidate evolution of model plant genus *Capsella*
655 (Brassicaceae). Molecular Ecology 21: 1223–1238.

656 **Kacser H, Burns JA. 1981.** The molecular basis of dominance. Genetics 97: 639-666.

657 **Katoh K, Misawa K, Kuma K-I, Miyata T. 2002.** MAFFT: a novel method for rapid
658 multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30:
659 3059–3066.

660 **Koenig D, Hagmann J, Li R, Bemm F, Slotte T, Neuffer B, Wright SI, Weigel D. 2019.**
661 Long-term balancing selection drives evolution of immunity genes in *Capsella*. eLife 8:
662 e43606.

663 **Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017.** Canu:
664 scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation.
665 Genome Research 27: 722–736.

666 **Kusaba M, Nishio T, Satta Y, Hinata K, Ockendon D. 1997.** Striking sequence similarity
667 in inter- and intra-specific comparisons of class I SLG alleles from *Brassica oleracea* and
668 *Brassica campestris*: Implications for the evolution and recognition mechanism. Proceedings
669 of the National Academy of Sciences of the USA 94: 7673–7678.

670 **Lafon-Placette C, Hatorangan MR, Steige KA, Cornille A, Lascoux M, Slotte T, Köhler**

671 **C. 2018.** Paternally expressed imprinted genes associate with hybridization barriers in
672 *Capsella*. Nature Plants 4: 352–357.

673 **Larsson A. 2014.** AliView: a fast and lightweight alignment viewer and editor for large
674 datasets. Bioinformatics. 30: 3276–3278.

675 **Li H. 2013.** Aligning sequence reads, clone sequences and assembly contigs with BWA-
676 MEM. arXiv:1303.3997v2, https://arxiv.org/abs/1303.3997v2.

677 **Liu P, Sherman-Broyles S, Nasrallah ME, Nasrallah JB. 2007.** A cryptic modifier causing
678 transient self-incompatibility in *Arabidopsis thaliana*. Current Biology 17: 824–824.

679 **Llaurens V, Billiard S, Leducq J-B, Castric V, Klein EK, Vekemans X. 2008.** Does
680 frequency-dependent selection with complex dominance interactions accurately predict allelic
681 frequencies at the self-incompatibility locus in *Arabidopsis halleri*? Evolution 62: 2545–2557.

682 **Llaurens V, Gonthier L, Billiard S. 2009.** The sheltered genetic load linked to the S locus in
683 plants: new insights from theoretical and empirical approaches in sporophytic self-
684 incompatibility. Genetics 183: 1105–1118.

685 **Ma R, Han Z, Hu Z, Lin G, Gong X, Zhang H, Nasrallah JB, Chai J. 2016.** Structural
686 basis for specific self-incompatibility response in *Brassica*. Cell Research 26: 1320–1329.

687 **Mable BK, Hagmann J, Kim S-T, Adam A, Kilbride E, Weigel D, Stift M. 2017.** What
688 causes mating system shifts in plants? *Arabidopsis lyrata* as a case study. Heredity 118: 110.

689 **Mable BK, Schierup MH, Charlesworth D. 2003.** Estimating the number, frequency, and
690 dominance of *S*-alleles in a natural population of *Arabidopsis lyrata* (Brassicaceae) with
691 sporophytic control of self-incompatibility. Heredity 90: 422–431.

692 **McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella**
693 **K, Altshuler D, Gabriel S, Daly M, et al. 2010.** The Genome Analysis Toolkit: a
694 MapReduce framework for analyzing next-generation DNA sequencing data. Genome
695 Research 20: 1297–1303.

696 **Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008.** Mapping and
697 quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5: 621–628.

698 **Nasrallah JB, Nasrallah ME. 2014.** *S*-locus receptor kinase signalling. Biochemical Society
699 Transactions 42: 313–319.

700 **Nasrallah JB, Liu P, Sherman-Broyles S, Schmidt R, Nasrallah ME. 2007.** Epigenetic
701 mechanisms for breakdown of self-incompatibility in interspecific hybrids. Genetics 175:
702 1965–1973.

703 **Nishio T, Kusaba M. 2000.** Sequence Diversity of *SLG* and *SRK* in *Brassica oleracea* L .
704 Annals of Botany 85: 141–146.

705 **Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG,**
706 **Weigel D, Lynch M. 2010.** The rate and molecular spectrum of spontaneous mutations in
707 *Arabidopsis thaliana*. Science 327: 92–94.

708 **Pannell J, Barrett S. 1998.** Baker's law revisited: Reproductive assurance in a
709 metapopulation. Evolution 52: 657–668.

710 **Porcher E, Lande R. 2005.** The evolution of self-fertilization and inbreeding depression
711 under pollen discounting and pollen limitation. Journal of Evolutionary Biology 18: 497–508.

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

714 **Schoen DJ, Busch JW. 2009**. The evolution of dominance in sporophytic self-
715 incompatibility systems. II. Mate availability and recombination. Evolution 63: 2099–2113.

716 **Schopfer CR, Nasrallah ME, Nasrallah JB. 1999.** The male determinant of self-
717 incompatibility in *Brassica*. Science 286: 1697–1700.

718 **Shimizu KK, Tsuchimatsu T. 2015.** Evolution of selfing: recurrent patterns in molecular
719 adaptation. Annual Review in Ecology, Evolution, and Systematics 46: 593–622.

720 **Sicard A, Kappel C, Lee YW, Woźniak NJ, Marona C, Stinchcombe JR, Wright SI,**
721 **Lenhard M. 2016.** Standing genetic variation in a tissue-specific enhancer underlies selfing-
722 syndrome evolution in *Capsella*. Proceedings of the National Academy of Sciences of the
723 USA. 113: 13911-13916.

724 **Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo Y-L, Steige K, Platts AE,**
725 **Escobar JS, Newman LK, et al. 2013.** The *Capsella rubella* genome and the genomic
726 consequences of rapid mating system evolution. Nature Genetics 45: 831-835.

727 **Slotte T, Hazzouri KM, Stern D, Andolfatto P, Wright SI. 2012.** Genetic architecture and
728 adaptive significance of the selfing syndrome in *Capsella*. Evolution 66: 1360–1374.

729 **Smith TF, Waterman MS. 1981.** Identification of common molecular subsequences. Journal
730 of Molecular Biology 147: 195-197.

731 **Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004.** AUGUSTUS: a web server for
732 gene finding in eukaryotes. Nucleic Acids Research 32: W309–12.

733 **Steige KA, Reimegård J, Koenig D, Scofield DG, Slotte T. 2015.** *Cis*-regulatory changes
734 associated with a recent mating system shift and floral adaptation in *Capsella*. Molecular
735 Biology and Evolution 32: 2501–2514.

736 **Steige KA, Laenen B, Reimegård J, Scofield DG, Slotte T. 2017.** Genomic analysis reveals
737 major determinants of *cis*-regulatory variation in *Capsella grandiflora*. Proceedings of the
738 National Academy of Sciences of the USA. 114: 1087-1092.

739 **Stein JC, Dixit R, Nasrallah ME, Nasrallah JB. 1996.** SRK, the stigma-specific S locus
740 receptor kinase of Brassica, is targeted to the plasma membrane in transgenic tobacco. The
741 Plant Cell 8: 429–445.

742 **Stein JC, Howlett B, Boyes DC, Nasrallah ME, Nasrallah JB. 1991.** Molecular cloning of
743 a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica*
744 *oleracea*. Proceedings of the National Academy of Sciences of the USA 88: 8816–8820.

745 **Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K. 2000.** The S
746 receptor kinase determines self-incompatibility in *Brassica* stigma. Nature 403: 913–916.

747 **Takayama S, Shimosato H, Shiba H, Funato M, Che FS, Watanabe M, Iwano M, Isogai**
748 **A. 2001.** Direct ligand-receptor complex interaction controls Brassica self-incompatibility.
749 Nature. 413: 534–538.

750 **Takayama S, Isogai A. 2005.** Self-incompatibility in plants. Annual Review of Plant Biology
751 56: 467–489.

752 **Tarutani Y, Shiba H, Iwano M, Kakizaki T, Suzuki G, Watanabe M, Isogai A,**
753 **Takayama S. 2010.** *Trans*-acting small RNA determines dominance relationships in *Brassica*
754 self-incompatibility. Nature 466: 983–986.

755 **Tsuchimatsu T, Shimizu KK. 2013.** Effects of pollen availability and the mutation bias on
756 the fixation of mutations disabling the male specificity of self-incompatibility. Journal Of
757 Evolutionary Biology 26: 2221–2232.

758 **Tsuchimatsu T, Goubet PM, Gallina S, Holl A-C, Fobis-Loisy I, Bergès H, Marande W,**
759 **Prat E, Meng D, Long Q, et al. 2017.** Patterns of polymorphism at the self-incompatibility
760 locus in 1,083 *Arabidopsis thaliana* genomes. Molecular Biology and Evolution 34: 1878–
761 1889.

762 **Tsuchimatsu T, Kaiser P, Yew C-L, Bachelier JB, Shimizu KK. 2012.** Recent loss of self-
763 incompatibility by degradation of the male component in allotetraploid *Arabidopsis*
764 *kamchatica*. PLoS Genetics 8: e1002838.

765 **Tsuchimatsu T, Suwabe K, Shimizu-Inatsugi R, Isokawa S, Pavlidis P, Städler T, Suzuki**
766 **G, Takayama S, Watanabe M, Shimizu KK. 2010.** Evolution of self-compatibility in
767 *Arabidopsis* by a mutation in the male specificity gene. Nature 464: 1342–1346.

768 **Uyenoyama MK, Zhang Y, Newbigin E. 2001.** On the origin of self-incompatibility
769 haplotypes: transition through self-compatible intermediates. Genetics 157: 1805–1817.

770 **Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A,**
771 **Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013.** From FastQ data to high confidence
772 variant calls: the Genome Analysis Toolkit best practices pipeline. Current Protocols in
773 Bioinformatics 11: 11.10.1–11.10.33.

774 **Vekemans X, Poux C, Goubet PM, Castric V. 2014.** The evolution of selfing from
775 outcrossing ancestors in Brassicaceae: what have we learned from variation at the *S*-locus?
776 Journal of Evolutionary Biology 27: 1372–1385.

777 **Wright S. 1939.** The distribution of self-sterility alleles in populations. Genetics 24: 538–
778 552.

779 **Wright SI, Kalisz S, Slotte T. 2013.** Evolutionary consequences of self-fertilization in

780 plants. Proceedings of the Royal Academy of Sciences Series B Biological Sciences 280:
781 20130133.

782 **Ågren JA, Wang W, Koenig D, Neuffer B, Weigel D, Wright SI. 2014.** Mating system
783 shifts and transposable element evolution in the plant genus *Capsella*. BMC genomics 15:
784 602.

785 # Figure Legends

786

787 **Figure 1. Self-compatibility is dominant and maps to the *S*-locus.**

788 **a.** Logarithm of odds (LOD) profile resulting from interval mapping of self-compatibility in

789 an interspecific *Capsella orientalis* × *Capsella grandiflora* F2 population. The dotted and

790 dashed lines indicates the 1% vs. 5% genome-wide permutation-based significance threshold.

791 The red vertical line shows the location of the canonical Brassicaceae *S*-locus. The 1.5-LOD

792 confidence interval ranges from position 6,241,223 to 8,742,368, whereas the *S*-locus is

793 located between positions 7,523,602 and 7,562,919 on chromosome 7. **b.** Estimated

794 quantitative trait locus (QTL) additive effect (red line) and dominance deviation (blue line)

795 across chromosome 7. Light shaded regions indicate standard errors.

796

797 **Figure 2. Sequence comparison of full-length *S*-haplotype sequences results in**

798 **identification of a frameshift deletion in *Capsella orientalis SCR*.**

799 **a.** Phylogram of *SRK* sequences, showing the diversity of *S*-alleles among Brassicaceae and

800 the close similarity of *SRK* in the *Arabidopsis halleri S12*-haplotype to the clade containing

801 *Capsella grandiflora CgS12*, *Capsella orientalis* and *Capsella bursa-pastoris* (B subgenome)

802 sequences (marked by a brace).

803 **b.** Maximum likelihood gene trees for three *S*-locus genes: *SRK*, *SCR* and *U-BOX* showing

804 the relationship between *A. halleri S12, C. grandiflora CgS12*, two *C. orientalis* and *C. bursa-*

805 *pastoris* (B subgenome) accessions.

806 **c.** Plot showing the percentage of sequence similarity (sequence conservation) between *C.*

807 *grandiflora CgS12 and C.orientalis* 1979/09 *S*-haplotypes. Gene positions are indicated by

808 grey bars.

809 **d.** Alignment of *SCR* sequences from *A. halleri S12, C. grandiflora CgS12*, *C. orientalis* and
810 *C. bursa-pastoris* (*B* subgenome) S-haplotypes. A frameshift deletion in the coding sequence
811 (marked by a red arrow) is found in *C. orientalis* but not in the two SI species *A. halleri* and
812 *C. grandiflora*.
813 **e.** Predicted SCR amino acid sequences for *A. halleri S12, C. grandiflora CgS12*, *C. orientalis*
814 and *C. bursa-pastoris* (*B* subgenome). The predicted protein sequence of *C. orientalis* lacks
815 five conserved cysteine residues (indicated by black arrows and orange boxes). The position
816 of the frameshift deletion is marked by a red arrow.
817
818 **Figure 3. Success of controlled crosses based on pollen tube germination assays.** Arrows
819 point to pollen tubes growing through the style. Scale bars, 200 μm.
820 **a.** Self-pollination of *Capsella grandiflora* carrying *CgS12* allele results in no pollen tube
821 growth (incompatible reaction), demonstrating functional self-incompatibility.
822 **b.** Pollination of *C. grandiflora* carrying *CgS12* with pollen from an individual carrying
823 different *S*-haplotypes results in pollen tube growth (compatible reaction).
824 **c.** Pollination of *C. grandiflora* carrying *CgS12* with pollen from *Capsella orientalis* results in
825 pollen tube growth (compatible reaction), demonstrating that *C. orientalis* SCR is not
826 functional.
827 **d.** Pollination of *C. orientalis* with pollen from a *C. grandiflora* carrying *CgS12* results in
828 pollen tube growth (compatible reaction).
829
830 **Figure 4. A conserved, *S*-linked *Capsella orientalis* sRNA is associated with repression of**
831 ***Capsella grandiflora SCR* in *S*-locus heterozygotes.**
832 **a.** Read depth of *C. orientalis* expresses *S*-linked small RNAs (sRNAs) homologous to
833 *Arabidopsis halleri S12 Ah12mirS3* in flower buds. The grey box indicates the length of the

834 sRNA precursor region and the location of *Ah12mirS3* 24 bp sRNA with highest expression
835 in *A. halleri S12* is indicated in red.
836 **b.** Expression (reads per kilobase of transcript per million mapped reads, RPKM) of 18-27 nt
837 sRNAs in the *Ah12mirS3*-like RNA precursor region in flower buds differs between *Capsella*
838 *orientalis* × *Capsella grandiflora* F2s with different *S*-locus genotypes (Kruskal-Wallis
839 χ^2=7.830, *P*=0.012): "Cg/Cg" and "Co/Co" are homozygous for the *C. grandiflora* or *C.*
840 *orientalis S*-allele respectively, wheras "Co/Cg" are heterozygous. Only homozygotes or
841 heterozygotes for the *C. orientalis S*-allele express sRNAs in the *Ah12mirS3*-like RNA
842 precursor region (Dunn's test *P*<0.01 for both comparisons Cg/Cg vs. Co/Cg and Cg/Cg vs.
843 Co/Co).
844 **c.** *mirS3* 24-nt small RNA sequences of *A. halleri S12* (*Ah12mirS3*) and *C. orientalis*
845 (*ComirS3*) and the predicted target in *C. grandiflora* Cg88.15 *SCR*, located 665 bp from exon
846 1 and 183 bp from exon 2.
847 **d.** Relative expression (RPKM) of *C. grandiflora SCR* (blue) and *C. orientalis SCR*
848 (turquoise) in flower buds of F2 individuals with different *S*-locus genotypes: "Cg/Cg" and
849 "Co/Co" are homozygous for the *C. grandiflora* or *C. orientalis S*-allele respectively, wheras
850 "Co/Cg" are heterozygous. *C. grandiflora SCR* is repressed in *C. grandiflora*/*C. orientalis*
851 heterozygotes (Kruskal-Wallis $\chi^2(2)$ = 9.9383, *P* < 0.01, Dunn´s test Z(2) = 2.25, *P* = 0.012
852 for Co/Cg vs Cg/Cg). Values for *C. grandiflora* are relative to the median RPKM of *C.*
853 *grandiflora* homozygotes, whereas those for *C. orientalis SCR* are relative to the median
854 RPKM of *C. orientalis* homozygotes.
855
856 **Figure 5. The timing of loss of self-incompatibility in *Capsella orientalis***
857 Phylogenetic tree showing relationships among *S*-haplotypes and estimates of the timing of
858 the loss of self-incompatibility (SI) in *C. orientalis* based on analyses in BEAST2. Green bars

859 at nodes indicate 95% credible intervals of the time to the most recent common ancestor
860 (TMRCA). The TMRCA of *C. grandiflora CgS12* and *C. orientalis* + *C. bursa-pastoris* B
861 represents an upper bound for the timing of loss of SI in *C. orientalis*. Because *C. orientalis*
862 was self-compatible when it contributed to the origin of *C. bursa-pastoris*, the TMRCA of *C.*
863 *orientalis* and *C. bursa-pastoris* B represents a lower bound on the timing of loss of SI.

Supporting Information

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

Methods S1 Additional methodological detail on experiments and analyses.

Table S1 Information of plant accessions used in this study.

Table S2 Information on sequence length and coverage for SMRT and MiSeq sequencing of BAC clones.

Table S3 Polymorphism and divergence at the *S*-locus.

Table S4 Expression of *SRK*, *SCR* and *U-box* in flower buds and leaves of *C. orientalis*.

Table S5 Estimation of the timing of the loss of self-incompatibility performed in BEAST with two competing models of population size change, constant and exponential.

Fig. S1 QTL mapping on different phenotyping scores.

Fig. S2 Sequence conservation (% identity) between *C. grandiflora CgS12* and *A. halleri S12.*

Fig. S3 Success of controlled crosses based on pollen tube germination assays.

Fig. S4 Number of pollen tubes in controlled crosses of *C. orientalis*, *C. grandiflora* harboring *CgS12*, and *C. grandiflora* harboring other *S*-haplotypes.

Fig. S5 Qualitative RT-PCR demonstrating expression of *SCR* in flower buds of *C. orientalis* and *C. grandiflora* individuals harboring *CgS12*.

Fig. S6 Non-synonymous changes in the SRK hypervariable region.