Supplementary Material

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

Jörg A. Bachmann[1,9], Andrew Tedder[1,6,9], Benjamin Laenen[1,9], Marco Fracassetti[1], Aurélie Désamoré[1], Clément Lafon-Placette[2,7], Kim A. Steige[1,8], Caroline Callot[3], William Marande[3], Barbara Neuffer[4], Hélène Bergès[3], Claudia Köhler[2], Vincent Castric[5], Tanja Slotte[1]*

[1]Department of Ecology, Environment and Plant Sciences, Science for Life Laboratory, Stockholm University, SE-106 91 Stockholm, Sweden

[2]Department of Plant Biology, Swedish University of Agricultural Sciences & Linnean Center for Plant Biology, SE-75 007 Uppsala, Sweden

[3]Institut National de la Recherche Agronomique UPR 1258, Centre National des Ressources Génomiques Végétales, Castanet-Tolosan, France

[4]Department of Botany, University of Osnabruck, 49076 Osnabruck, Germany

[5]Unité Evo-Eco-Paléo (EEP) - UMR 8198, CNRS/Université de Lille - Sciences et Technologies, Villeneuve d'Ascq Cedex, F-59655, France

[6]Present address: School of Chemistry and Biosciences, Faculty of Life Sciences, University of Bradford, Bradford BD7 1DP, UK

[7]Present address: Department of Botany, Charles University, CZ-128 01 Prague, Czech Republic

[8]Present address: Institute of Botany, Biozentrum, University of Cologne, 50674 Cologne, Germany.

[9]These authors contributed equally.

*Author for correspondence: tanja.slotte@su.se

Supplementary Text

Genetic mapping of the loss of self-incompatibility in *Capsella orientalis*

Generation of an interspecific C. orientalis x C. grandiflora F2 mapping population

To map the genetic basis of self-compatibility in *Capsella orientalis*, we generated an interspecific *C. orientalis* × *C. grandiflora* F2 mapping population. This required an embryo rescue protocol, as F1 seeds were aborted prior to full development. To produce viable F1 hybrids, *C. orientalis* accession Co2008-1 (Table S1) was crossed as a seed plant with *C. grandiflora* Cg88.15 as pollen donor (Table S1). Fourteen days after pollination, siliques were surface sterilized by dipping them successively in 70% (v/v) ethanol and a sterilizing solution (5% (v/v) sodium hypochloride, 0.01% (v/v) Tween 20). Embryos were then dissected out of the seeds in sterile ½ MS solution using fine needles, and put on MS plates (½ MS, 2% (w/v) sucrose; Murashige and Skoog basal salt mixture, Sigma-Aldrich Co. MI, USA). A total of 29 embryos were collected. Plates were then covered and placed under long days condition (16 hrs light and 8 hrs darkness) at a temperature of 22°C light and 20°C darkness and a light intensity of 110 µE during two weeks. Afterwards, seedlings were transferred to soil and grown in a growth chamber at 60% humidity and daily cycles of 16 hrs light at 21°C and 8 hrs darkness at 18°C. The resulting F1 individuals were self-compatible, and we obtained F2 seeds from one autonomously self-pollinated F1 individual.

F2 seeds were surface-sterilised and stratified at 2-4°C in the dark for two weeks on ½ MS medium (Murashige and Skoog basal salt mixture, Sigma-Aldrich Co. MI, USA). Plates were then moved to a climate chamber for seeds to germinate (16 h light at 20ºC / 8 h dark at 18 ºC, 70 % max. humidity, 122 uE light intensity). and after one week, seedlings were transplanted onto soil in the same growth chamber conditions. Germination rate was >80%, and we used a total of 350 F2 individuals for QTL mapping.

Phenotyping of F2s

We recorded self-incompatibility / self-compatibility (SI/SC) in 321 F2 individuals. First, SI/SC was visually scored as presence or absence of silique formation on mature individuals. Second, we manually self-pollinated 3-6 flowers of 204 F2 individuals to verify that SI individuals are functionally SI, and are not simply kept from autonomous selfing due to e.g. a flower structure closer to the outcrossing parent *C. grandiflora* than the autonomous selfer *C. orientalis*. We recorded the number of enlarged siliques one week after self-pollination. As flower structure seemed to have a major impact on the efficacy of autonomous self-pollination in our F2 population (approximately 30% of plants initially scored as SI based on silique formation after autonomous self-pollination formed seeds after manual self-pollination, whereas only ~2% of F2s that were called as SC after autonomous self-pollination formed no siliques after manual self-pollination), we preferentially used the scoring based on manual self-pollination in our analyses. Third, to validate that the SI phenotype is due to pollen tube growth arrest and the lack of seed development following self-pollination is not due to other reasons, e.g. inbreeding depression or genetic incompatibilities between *C. orientalis* and *C. grandiflora*, we assessed pollen tube growth in the pistil after manual self-pollination in a subset of F2 individuals scored as SI based on silique formation after autonomous self-pollination. For 9 out of 10 individuals scored as SI, manual self-pollination did not result in pollen tube growth. Pistils were fixed in EtOH: acetic acid 9:1 for > 2 hours, softened in 1N NaOH 60°C for 20 minutes and stained with 0.01% decolorised aniline blue in 2% solution of K_3P0_4 for 2 hours. Pollen tubes were visualised by mounting the pistils on a microscope slide and observation under an epifluorescence microscope (Zeiss Axiovert 200M). In total, 321 F2 individuals were scored for SI/SC status.

Genotyping of F2s

We extracted DNA from young leaves of 350 F2s using a standard Qiagen DNeasy kit (Qiagen). Following digestion of genomic DNA with EcoRI, RAD-seq libraries were generated at the National Genomics Infrastructure at SciLifelab Stockholm. EcoRI-digested samples were prepared using the Agilent Bravo system (protocol available at Github: https://github.com/ngi-automation/rad-seq). During preparation a custom adapter with complementary bases to the EcoRI overhang was ligated to the ends of the digested DNA using T4 DNA ligase and 10X T4 DNA ligase (New England Biolabs). The ligated adapter contains Illumina adapter sequences for incorporation amplification in a later step. Samples were purified using AMPure bead purification (Beckman Coulter), strand displacement using Bst 2.0 polymerase, 10X ThermoPol buffer (New England Biolabs) and dNTPs (Thermo Scientific), a second bead purification before the amplification of DNA and a final bead purification of the finished library. To evaluate the quality of the library a concentration measurement using Qubit dsDNA HS Assay kit (Thermo Fisher Scientific) and an analysis of library size using Fragment Analyzer (Kem-en-Tec) was performed. A calculated concentration in the range of 420-720 bp was used to even out the contribution of the different samples to the the pool before subjecting the pool to size selection on the LabChip XT (PerkinElmer) and concentration and size was analyzed before sequencing. Samples were sequenced on an illumina HiSeq 2500 High Output v4 mode with 2x125bp reads.

The demultiplexed reads were trimmed with Trimmomatic (Bolger et al. 2014) to remove remaining Nextera adapters, with a sliding window approach (step size 4, window size 20). Base pairs at trailing and leading sequence ends with a quality score below 20 were removed. After trimming only proper paired reads with a length above 50 base pairs were retained.

The reads were mapped to the *Capsella rubella* v 1.0 genome (Slotte et al. 2013), using BWA-MEM (Li 2013) with default parameters. We called SNPs and genotypes using the GATK (The Genome Analysis Toolkit) pipeline (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013). First, the base pair quality scores of the mapped reads where recalibrated (Base Quality Score Recalibration) using a set of 1,538,085 SNPs identified in *C. grandiflora* (Williamson et al. 2014). Second, we allow the presence of duplicated reads with the -drf option, since many reads mapped in the same genomic positions in GBS data differently from whole genome sequencing. Eventually, SNPs were called using the GATK UnifiedGenotyper. Only biallelic SNPs with a minimum sequencing depth of 8, maximum depth of 200 and present at least in 80% of the samples were retained. Furthermore we filtered out the SNPs present in repeated region identified with RepeatMasker v4.0.7 (http://www.repeatmasker.org). The SNPs were phased by reference to genome-resequencing data for the *C. orientalis* Co2008-1 parent (see section "Whole genome resequencing" below). The final genetic data set consisted of 998 SNPs in 350 individuals.

QTL mapping of loss of self-incompatibility

We constructed a genetic map with the R/Qtl package (Broman et al. 2003) in R 3.4.2. We removed SNPs that had redundant genotype information, were assigned to the wrong linkage group based on position information from the *C. rubella* genome, or showed segregation distortion. The genetic map was based on 549 SNPs in 328 individuals. Using a logarithm of odds (LOD) score cutoff of 6, we obtained 8 linkage groups that correspond to the haploid chromosome number of 8 chromosomes of diploid *Capsella spp*. The QTL mapping was performed on 304 F2 individuals scored for SI status and present in the genetic map. We conducted interval mapping with a binary data model using the Haley & Knott regression method in intervals of 1 cM, with a 1% genome-wide significance threshold obtained by 1000

permutations. We obtained 1.5-LOD confidence intervals for the significant QTL. We estimated the additive allelic effect (half the phenotypic difference between *C. grandiflora* and *C. orientalis* homozygotes) and dominance deviation at the QTL peak using the Rqtl effectscan function. To assess whether scoring of SI after autonomous or manual self-pollination had an effect on the detected QTL, we repeated the QTL analysis using the two phenotyping scores separately with the same parameters used previously. Visual SI/SC score was available for 228 individuals, whereas the score based on number of siliques formed was available for 195 genotyped individuals. In both cases, there was a significant QTL for self-compatibility in the genomic region that harbors the *S*-locus (Figure S8).

Sequencing and annotation of the *S*-locus

BAC library construction and screening

Because the *S*-locus is a region under strong balancing selection, reference-based whole-genome resequencing approaches cannot be used to study *S*-locus evolution. Instead, we opted for targeted sequencing and assembly of *S*-haplotypes using long-read sequencing of *S*-locus containing bacterial artificial chromosome (BAC) clones. This strategy results in high-quality assemblies, especially when combining long-read assembly with short-read based error correction (Bachmann et al. 2018). To identify putative causal genetic changes responsible for loss of SI, we sequenced and annotated *S*-haplotypes from two *C. orientalis* accessions (Co1719/11 and Co1979/09; Table S1). We also sequenced two *C. grandiflora S*-haplotypes, one identified as harboring an *S*-haplotype highly similar to that of *C. orientalis* (accession Cg2-2-KS2; Table S1) and one (accession Cg88.15) representing the *C. grandiflora S*-haplotype segregating in our F2 mapping population. Finally, we obtained *S*-haplotypes of the *C. orientalis*-derived B-subgenome from five *C. bursa-pastoris* individuals covering the species range in Eurasia (Table S1), for use in analyses of the timing of the loss of SI in *C. orientalis* (see "Timing of the loss of self-incompatibility in *C. orientalis*" below). In total, we generated and annotated eight full-length *S*-locus haplotypes from *Capsella spp*. (two from *C. orientalis*, two from *C. grandiflora*, and four from the *C. bursa-pastoris* B-subgenome).

Plants for BAC library construction were grown as previously described (Bachmann et al. 2018) and at least 5 g of leaf material was collected per plant for bacterial artificial chromosome (BAC) library production, except in the case of Cg88.15, where combined leaf material of 11 SI F2 plants was used. One BAC library was generated based on material from two *C. bursa-pastoris* accessions (Table S1) the resulting *S*-haplotype from this library is termed CH1. BAC libraries were generated and screened at the French Plant Genomic Resource Centre (CNRGV). BACs containing *S*-locus haplotypes of interest were identified based on insert size and hybridisation with DNA probes of genes flanking the *S*-locus.

Sequencing, assembly and annotation of BAC clones

BAC clones with *S*-haplotypes of interest were sequenced to high coverage using SMRT sequencing (150-400 x; Table S2), and by short-read sequencing on an Illumina MiSeq (>380 x; Table S2) at the SciLifeLab National Genomics Infrastructure (NGI), Uppsala. BAC clone sequences were assembled in HGAP.3 based on SMRT sequencing data, except that from Cg88.15, which was instead assembled using Canu (v1.7) (Koren et al. 2017) with default parameter and target genome size of 1Mb. For all BAC assemblies, short-read sequencing data was subsequently used to correct indel errors in long-read assemblies, as previously described (Bachmann et al. 2018).

S-locus haplotypes were subsequently annotated as previously described (Bachmann et al. 2018). Briefly, we annotated our *S*-locus assemblies using Augustus v3.2.3 (Stanke et al.

2004) and RepeatMasker v4.0.7 http://www.repeatmasker.org), run via Maker v2.31.9 (Holt and Yandell 2011). We used *Arabidopsis thaliana* as a model prediction species and protein homology data for *SRK*, *U-box* and *ARK3* from *Arabidopsis lyrata* and *A. halleri*. Due to the high levels of sequence diversity at the key *S*-locus genes *SRK* and *SCR*, they were difficult to annotate automatically using Maker. To annotate *SRK*, we instead searched for *SRK* exon 1 sequences within the *S*-haplotypes by scanning for sequence similarity between general *SRK* exon 1 forward / reverse primers (*SRLF / SLGR*) (for more details, see (Bachmann et al. 2018)). Sequence similarity (BLASTN) to known *SRK* exon 1 sequences was used to accept the candidate loci as *SRK*, while we used close similarity to *ARK3* as a rejection criterion. To annotate *SCR*, we used a window-based approach to screen for the characteristic pattern of cysteine residues after translation of the DNA sequence in all three frames. Using this approach, we identified a region highly similar to *SCR* in *A. halleri S*-locus haplotype *S12* (GenBank accession number KJ772374.1) in our *S*-locus BAC sequences. Using BLASTN we found high similarity between the Cg88.15 *S*-haplotype and the *S*-haplotype *A. halleri* S4 (GenBank accession KJ461484), and the Cg88.15 *S*-haplotype was therefore annotated by reference to the *A. halleri* S4 sequence annotation.

Identification of a C. grandiflora S-haplotype highly similar to C. orientalis

Using a dataset of Brassicaceae *SRK* exon 1 and *ARK3* sequences downloaded from Genbank for a previous study (Bachmann et al. 2018) we made an alignment of *SRK* exon 1 sequences using MAFFT v7.245 & E-INS-I algorithm (Katoh et al. 2002). Manual curation and error correction of the alignment was performed in SeaView v4.6 (Gouy et al. 2010). We generated a maximum likelihood phylogenetic tree from the alignment between SRK sequences with RaXMl v8.2.3 (GTRGAMMA model and 1000 bootstraps replicates) and then plotted the *SRK* phylogeny in R v. 3.3.1.

Based on the phylogenetic tree, we identified an *SRK* sequence in *C. grandiflora* accession Cg2-2-KS2 with high similarity to *SRK* from *C. orientalis* and *C. bursa-pastoris B*. Sequence identity between of this *SRK* allele from Cg2-2-KS2 and *C. orientalis SRK* was 98.3% (estimates were identical for both *C. orientalis* accessions), which is above the level of sequence similarity considered to represent the same specificity in Arabidopsis (Castric et al. 2008; Tsuchimatsu et al. 2012). Due to the high sequence similarity of this *C. grandiflora S*-haplotype to *A. halleri S12*, we termed this *S*-haplotype *CgS12.* We further assessed sequence conservation across the entire *S*-locus to check whether the region of high sequence similarity between this *S*-haplotype and the *C. orientalis S*-haplotype was limited to *SRK*. For this purpose, we aligned *S*-locus sequences in AliView v. 1.20 (Larsson 2014), using Muscle v. 3.8.31 (Edgar 2004) with block alignment, calculated pairwise sequence conservation in 250 bp sliding windows, and visualized the results in R v. 3.3.1.

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

To identify candidate causal mutations for the loss of SI in *C. orientalis*, we analyzed sequence alignments of the two key *S*-locus genes *SRK* and *SCR*, as well as of the *S*-linked *U-box* gene, which may act as a modifier of the SI response. Specifically, we searched for major-effect changes such as frameshifts, premature stop codons or non-consensus splice sites that were found in sequences from the self-compatible *C. orientalis* and/or in the *C. bursa-pastoris* B subgenome, which is derived from *C. orientalis*, but not in the self-incompatible *C. grandiflora CgS12* or *A. halleri S12* (GenBank accession number KJ772374.1).

There were no major-effect changes in our *C. orientalis S*-haplotypes compared to *C. grandiflora CgS12* or *A. halleri S12* in either the *SRK* or *U-box* genes. However, in *SCR*, a frameshift deletion was predicted to be present in both *C. orientalis S*-haplotypes and in two of our *C. bursa-pastoris* B sequences (the other two *C. bursa-pastoris* B sequences had an

overlapping larger deletion in this region). The *SCR* frameshift deletion was not found in *C. grandiflora CgS12*. To check whether this deletion was fixed in *C. orientalis* we analyzed Illumina whole-genome resequencing data from an additional 25 accessions (Supplementary table S1). We mapped the trimmed data to a *C. rubella* reference modified to include a *C. orientalis S*-haplotype, as described in the section "Timing of loss of self-incompatibility in C. orientalis" bellow and scored visual presence or absence of the deletion on .bam files in IGV v2.4.2 (Robinson *et al.* 2011)

Assessing the functionality of *C. orientalis* SCR

We performed controlled crossed to verify that *C. grandiflora CgS12* confers SI, and to assess the functionality of SCR in *C. orientalis*. To verify functional SI in *C. grandiflora* carrying *CgS12,* we performed 12 manual self-pollinations of a *C. grandiflora* individual carrying the *CgS12 S*-haplotype. We further assessed the success of manual self-pollination of *C. orientalis* by performing 6 manual self-pollinations. To assess whether *C. orientalis* SCR is functional, we crossed *C. grandiflora* harboring *CgS12* as a seed parent to *C. orientalis* as a pollen donor. We performed a total of 112 crosses of this type, with two different *C. orientalis* accessions as pollen donors and three different *CgS12*-carrying *C. grandiflora* as seed parents (Supplementary Table S1). If *C. orientalis* SCR is functional, and provided that *CgS12 SRK* is expressed, then we expect this cross to be incompatible, whereas if *C. orientalis* SCR is nonfunctional, the cross should be compatible. The reciprocal cross of the same individuals was also carried out with the same accessions (total 84 crosses of this type). Finally, we performed 12 crosses of *C. grandiflora* harboring other *S*-haplotypes to *C. grandiflora* harboring *CgS12*, and 12 to *C. orientalis*. These crosses are expected to be successful. We observed pollen tube growth in the pistil 12 hours after pollination. Pollen tube germination was visualized as described in the section "Genetic mapping of loss of SI in *C. orientalis"*. We compared the number of pollen tubes among different types of crosses using a Kruskal-Wallis test (Fig. S3).

Expression of SCR, SRK and U-box

To assess whether *SRK, SCR* and *U-box* were expressed in *C. orientalis* flower buds, we generated RNAseq data from the same two *C. orientalis* accessions (Co1719/11 and Co1979/09) for which we had sequenced the *S*-locus using targeted long-read sequencing. Briefly, three biological replicates (different individuals) of two *C. orientalis* accessions (Co1719/11 and Co1979/09) were grown at long-day conditions and mixed-stage flower buds were collected for each sample, as previously described (Steige et al. 2017). For comparison, we also collected leaf samples from the same individuals. RNA was extracted using a Qiagen RNEasy Plant Mini Kit (Qiagen) and RNAseq libraries were constructed using the TruSeq RNA v2 kit. Sequencing of 100-bp paired-end reads was done on an Illumina HiSeq 2000 instrument. Reads were trimmed with Trimmomatic (Bolger et al. 2014) with a sliding window approach (step size 4, window size 20). The reads were mapped using STAR software v.2.2.1 (Dobin et al. 2013) against a modified v1.0 reference *C. rubella* assembly (Slotte et al. 2013), where the *S*-locus region (scaffold_7 7523601:7562919) was masked and the S-locus from *C. orientalis* Co1719/11 was added.

Expression was quantified as RPKM (the number of reads per kb per million mapped reads (Mortazavi et al. 2008)) and mean values and standard errors over our three biological replicates are presented in Table S4. *SCR* is known to be specifically expressed in anther tapetum and in pollen (Shiba et al. 2006), which constitute a small and variable fraction of the mixed-stage flower buds used for RNAseq here. Despite this limitation, we still found evidence for expression of *SCR* in flower buds but not leaves of both *C. orientalis* accessions examined (Table S4), although at a higher level in accession Co1719/11 than in accession

Co1979/09, in this data set. We also found evidence for expression of *SRK* and *U-box* in flowers (Table S4).

To validate our inference that *SCR* is expressed in *C. orientalis* and assess whether *SCR* is expressed in *C. grandiflora* harboring the *CgS12* haplotype, we conducted qualitative RT-PCR with primers specific to *SCR* in *C. orientalis* and *C. grandiflora CgS12*. In this assay, we detected expression of *SCR* in flower buds of both the Co1719/11 and Co1979/09 accessions, as well as in three *C. grandiflora* individuals harboring *CgS12* (Fig. S4). We collected fresh mixed-stage flower buds, <100 mg per sample, and extracted total RNA with the RNeasy Plant Mini Kit (Qiagen, Venlo, The Netherlands). Apart from the standard protocol, we included a DNA digestion step with RNase-Free DNase (Qiagen, Venlo, The Netherlands). A DNA digestion with RNase-Free DNase (Qiagen, Venlo, The Netherlands) was added to the standard RNeasy protocol for plant tissue. We eluted the RNA in RNase-free water (Qiagen, Venlo, The Netherlands) and measured 10x diluted RNA with the RNA 6000 Nano Kit (Agilent Technologies Inc. Waldbronn, Germany) on an Agilent 2100 Bioanalyzer (Agilent Technologies Inc. Waldbronn, Germany). Only samples with RIN value > 8 were kept. For cDNA generation, we took 0.8 µg of RNA from each sample and followed the first-strand cRNA synthesis protocol of Invitrogen SuperScript III (Thermo Fisher Scientific, Waltham MA, USA), using oligo(dT)20 primers (Thermo Fisher Scientific, Waltham MA, USA) and omitting the RNase inhibitor step. We diluted the samples 1:20 and amplified gene sequences with AmpliTaq Gold DNA polymerase (Thermo Fisher Scientific, Waltham MA, USA) in a three-step cycling program (5 min at 95°C, 35 cycles of 40 sec at 95°C / 60 sec at 60°C / 40 sec at 72°C, 10 min at 72°C) and the following concentrations: MgCL2: 20 ng/µl, dNTPs: 2 mM, BSA: 20mg/µl, forward primer: 10 µM, reverse primer 10 µM, Taq: 5 U/µl, Buffer II: 1x.) We designed specific PCR primers for SCR transcript of CgrS12 S-allele: 5A13_SCR_D_f (5'-CGGCAACTCTGCTCTTTGTTT) & 5A13_SCR_D_r (5'-AGTACGGATACACTCGCATTCC). As a control, we amplified a region of the *TUB* gene, which is expected to be expressed equally in all individuals using the primers AlyrTUB_1F (5'-ACCACTCCTAGCTTTGGTGATCTG) and AlyrTUB_2R (5'-AGGTTCACTGCG-AGCTTCCTCA).

Expression of *SCR* in interspecific F2 individuals

To be able to separately assess the expression of *C. grandiflora* and *C. orientalis SCR* in the F2s we sequenced the *S*-haplotype from *C. grandiflora* Cg88.15 segregating the in the interspecific F2s (see section "*Sequencing, assembly and annotation of BAC clones*" above). The resulting sequence was added to the modified *C. rubella* reference described above. Based on the F2 genotyping we selected 19 F2s belonging to different *S*-locus genotype classes (3 with the Cg/Cg genotype, 12 with the Cg/Co genotype and 4 with the Co/Co genotype). Total RNA was extracted from flower buds and RNAseq libraries were generated, sequenced and analyzed following the same protocol and pipeline described above. The reads were aligned to the new reference that contained the *S*-locus sequences from *C. grandiflora* (Cg88.15) and *C. orientalis* (Co1719/11).

We separately quantified expression levels for *C. grandiflora* and *C. orientalis SCR* in the F2 individuals of different *S*-locus genotype and performed tests for differential expression among *S*-genotypes in R 3.5.0 (R Core Team 2017) with a Kruskal-Wallis rank sum test and Dunn's test of multiple comparisons using rank sums.

Small RNA expression

To test if the *C. orientalis S*-allele expressed small RNAs similar to *A. halleri AhmirS3*, we extracted RNA using the mirVana Kit (Thermo Fisher Scientific Inc, Waltham MA, USA) from mixed stage flower buds of the same 19 F2 individuals as described in the section above

"Expression of SCR in the F2", and obtained sequencing data for small RNA. We prepared libraries using the Illumina TruSeq small RNA kit, and sequenced them on an Illumina HiSeq2500, using a 1x51 setup with the HiSeq Rapid SBS Kit v2 sequencing chemistry (Illumina, Inc., San Diego, USA). One sample failed during sequencing and was not included in the analysis. We trimmed sequencing adapters and low quality reads with Trimmomatic v0.36 and mapped small RNA of length 18-27 nt to a modified *C. rubella* reference including the *C. orientalis S*-locus, as described in the above section "Expression of *S*-locus genes in *C. orientalis*" with STAR.

We identified and annotated a region highly similar to the *Ah12mirS3* small RNA precursor involved in the dominance of of *A. halleri S12* (Durand et al. 2014) in our *C. orientalis S*-haplotypes, and examined expression levels of small RNAs uniquely mapping to this region. We compared the expression of sRNAs in the *Ah12mirS3*-like sRNA precursor region (hereafter termed *ComirS3*s) between *S*-genotypes with a Kruskal-Wallis rank sum test followed by a Dunn's post-hoc test. We further examined the location of the expressed *C. orientalis* small RNAs in relation to the position of *Ah12mirS3* small RNAs in *A. halleri.* For these analyses, we combined small RNAs of 3 F2 individuals homozygous for the *C. orientalis S*-haplotype.

Small RNA targets

We took all *ComirS3* small RNAs of *C. orientalis* in three F2 individuals homozygous for the *C. orientalis S*-haplotype (described in the section above), and searched for small RNA targets within 1kb of *SCR* of the *C. grandiflora* Cg88.15 *S*-haplotype segregating in the F2 individuals using a Smith and Waterman algorithm (Smith & Waterman 1981) with the scoring matrix: matches = +1; mismatches = -1; gaps = -2; G:U wobbles = -0.5, as previously described[25].

Whole genome resequencing

We generated Illumina whole-genome resequencing data for three additional *C. orientalis* individuals (Co1981/5, Co1719/03 and Co2008/1; Table S1) for evolutionary genetic analyses of the *S*-locus and for assigning alleles as *C. orientalis* or *C. grandiflora* in the F2 mapping population. Plant growth conditions and sequencing was performed as described previously (Steige et al. 2017), and all individuals were sequenced to at least 25x coverage.

Timing of loss of self-incompatibility in *C. orientalis*

We used a strategy similar to that in (Guo et al. 2009) to estimate a lower and upper bound of the timing of the loss of SI in *C. orientalis.* We obtained a lower bound for the timing of the loss of SI by estimating the time to the most recent common ancestor (TMRCA) based on *C. orientalis* and *C. bursa-pastoris* B subgenome *S*-locus sequences. This is possible because extensive haplotype sharing between *C. orientalis* and the *C. bursa-pastoris* B subgenome, indicates that the ancestor of *C. orientalis* that contributed to formation of *C. bursa-pastoris* was already selfing (Douglas et al. 2015). To gain an upper bound for the timing of the loss of SI we estimated the TMRCA for *C. orientalis*, *C. bursa-pastoris* B and the ancestral *C. grandiflora S*-haplotypes.

We supplemented our *C. orientalis S*-locus polymorphism data from BAC sequencing by mapping Illumina whole-genome resequencing data for 30 additional *C. orientalis* individuals (Table S1) against a *C. rubella* reference modified to include a *C. orientalis S*-haplotype, as described in the section "Expression of *SCR, SRK* and *U-box* in *C. orientalis*" above. We mapped our genomic reads with BWA-MEM and called SNPs with GATK HaplotypeCaller. Because *C. orientalis* is highly homozygous, self-compatible, and has low levels of polymorphism genome-wide (Douglas et al. 2015), this approach is expected to

work well, as long as a *C. orientalis S*-haplotype is included in the reference genome. We filtered sites following GATK recommended hard filtering with the following parameters; QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0. We required at minimum read depth of 15 and maximum of 200. As mapping can be uncertain in repeat regions and in the presence of transposable element with filtered out those regions using repeatMasker. Heterozygous SNPs were also hard masked in the alignment. Finally, the vcf file for the three *C.orientalis* samples was transformed to fasta format.

We produced an alignment of 37 *S*-locus sequences including the *C. grandiflora CgS12 S*-haplotype, 4 *C. bursa-pastoris* subgenome B *S*-haplotypes and *S*-haplotype data for 32*C. orientalis* individuals. Sequences were aligned using block alignment using Muscle v.3.8.31 (Edgar 2004) as implemented in AliView v.1.20 (Larsson 2014). The total length of the *S*-locus alignment (excluding *C. orientalis* orthologs of the genes At4g21350 and At4g21380) was 33485 bp, 22689 bp had indels in at least one sequence, 9835 of invariant and 876 polymorphic sites. The alignment was partitioned into coding and non-coding regions following the annotation and sites with indels were pruned before further analysis of polymorphism and divergence in dnaSP v6 (Rozas et al. 2017).

To obtain estimates of the upper and lower bound of the timing of the loss of self-incompatibility, we followed an approach similar to that in (Guo et al. 2009), although here we base our analyses on data for the entire *S*-locus, not only *SRK*. We estimated the timing of the splits between *C. grandiflora*, *C. bursa-pastoris* and *C. orientalis* as well as the crown age of *C. orientalis* using a strict molecular clock in a Bayesian framework as implemented in BEAST2 (Bouckaert et al. 2014). The TMRCA for all included samples provides an estimate of the upper bound of the timing of the loss of SI in *C. orientalis,* whereas the TMRCA of *C. orientalis* and *C. bursa-pastoris* provides a lower bound of the timing of the loss of SI. This approach assumes that the ancestor of *C. bursa-pastoris* from the *C. orientalis* lineage was already self-fertilizing, as inferred previously based on genome-wide haplotype conservation between the *C. bursa-pastoris* B subgenome and *C. orientalis* (Douglas et al. 2015).

In our analyses, the clock rate was fixed assuming a mutation rate of $7x10^{-9}$ substitutions per sites per generation (Ossowski et al. 2010) and we used a generation time of one year. The best substitution models inferred in PartitionFinder2 v.2.1.1 (Lanfear et al. 2012; Lanfear et al. 2017) for the coding and non-coding partition were GTR + G and HKY + I respectively. We ran both a complex model with exponential changes in population size and a model with a constant population size, and assessed whether the more complex model gave a significant improvement in likelihood using aicm (Baele et al 2012). The exponential growth model gave a slight improvement in aicm and we therefore present results from both model (Tables S5). We ran two chains of 10 millions generations sampled every 1000 generations and checked the convergence by visual inspection of the log-likelihood profile and assuring ESS value above 200. The posterior distribution of trees was used to build a maximum clade credibility tree and estimate node age and 95% credibility interval using TreeAnnotator (Drummond et al. 2012).

Supplementary Tables

Table S1. Overview of plant accessions used in this study.

Species	Accession	Population	Origin[1]	ENA numbers	Purpose
C. orientalis	Co2008-1	2008	CH	SRR1481629 (WGS)	Seed parent of F1 parent of F2 mapping population; whole-genome resequencing.
C. orientalis	Co1719/11	1719	MN	tbd	*S*-locus BAC sequencing; Expression analyses; Controlled crosses
C. orientalis	Co1979/09	1979	RU	tbd	*S*-locus BAC sequencing; Expression analyses; Controlled crosses
C. orientalis	DKCo_1_1718-9	1718	MN	ERR636109	Whole-genome resequencing
C. orientalis	DKCo_2_1719-1	1719	MN	ERR636115	Whole-genome resequencing
C. orientalis	Co1719/03	1719	MN	tbd	Whole-genome resequencing
C. orientalis	1719-4		MN	SRR1463025	Whole-genome resequencing
C. orientalis	DKCo_3_1938-1_1	1938	KZ	ERR636116	Whole-genome resequencing
C. orientalis	DKCo_4_1939-1_6	1939	KZ	ERR636117	Whole-genome resequencing
C. orientalis	DKCo_5_1940-1_1	1940	KZ	ERR636118	Whole-genome resequencing
C. orientalis	DKCo_6_1978-6	1978	RU	ERR636119	Whole-genome resequencing
C. orientalis	1979-1	1979	RU	SRR1481500	Whole-genome resequencing
C. orientalis	DKCo_7_1979-02	1979	RU	ERR636120	Whole-genome resequencing
C. orientalis	1979-7	1979	RU	SRR1481499	Whole-genome resequencing
C. orientalis	DKCo_8_1980-1	1980	RU	ERR636121	Whole-genome resequencing

C. orientalis	DKCo_9_1981-3	1981	RU	ERR636122	Whole-genome resequencing
C. orientalis	Co1981/5	1981	RU	tbd	Whole-genome resequencing
C. orientalis	1981-6	1981	RU	SRR1481625	Whole-genome resequencing
C. orientalis	1981-10	1981	RU	SRR1481618	Whole-genome resequencing
C. orientalis	DKCo_10_1982-9	1982	RU	ERR636107	Whole-genome resequencing
C. orientalis	DKCo_11_1983-6	1983	RU	ERR636108	Whole-genome resequencing
C. orientalis	DKCo_12_1984-2	1984	RU	ERR636110	Whole-genome resequencing
C. orientalis	1985-1	1985	RU	SRR1481626	Whole-genome resequencing
C. orientalis	DKCo_13_1985-11	1986	RU	ERR636111	Whole-genome resequencing
C. orientalis	DKCo_15_2006-01	2006	CH	ERR636112	Whole-genome resequencing
C. orientalis	DKCo_16_2007-03	2007	CH	ERR636113	Whole-genome resequencing
C. orientalis	Co2008-2_1	2008	CH	tbd	Whole-genome resequencing
C. orientalis	2008-07-01	2008	CH	SRR1481628	Whole-genome resequencing
C. orientalis	2008-09-01	2008	CH	SRR1481627	Whole-genome resequencing
C. orientalis	FY1	FY1	CH	SRR6179226	Whole-genome resequencing
C. orientalis	QH1	QH1	CH	SRR6179228	Whole-genome resequencing
C. orientalis	QH2	QH2	CH	SRR6179227	Whole-genome resequencing
C. grandiflora	Cg88.15	88	GR	tbd	Pollen parent of F1 parent of F2 mapping population; *S*-locus BAC sequencing
C. grandiflora	Cg2-2-KS2	2	GR	tbd	*S*-locus BAC sequencing; offspring used in controlled crosses.

C. bursa-pastoris	CbpWEDE9.136	DE9	DE	tbd	*S*-locus BAC sequencing
C. bursa-pastoris	CbpWESE11.139	SE11	SE	tbd	*S*-locus BAC sequencing
C. bursa-pastoris	CbpKMB205	KMB	CH	tbd	*S*-locus BAC sequencing[2]
C. bursa-pastoris	CbpGY36	GY	CH	tbd	*S*-locus BAC sequencing[2]
C. bursa-pastoris	CbpAQ416	AQ	CH	tbd	*S*-locus BAC sequencing

[1]Geographical origin; CH=China, DE=Germany, KZ=Kazakhstan, MN=Mongolia, RU=Russia, SE=Sweden, GR=Greece. More information on the geographical origin of *C. orientalis* samples is given in Hurka et al. (2012) and more information on the origin of the *C. bursa-pastoris* samples is available in Slotte et al. (2009). We further used publicly available *C. orientalis* genome resequencing data produced by Douglas et al. (2015), Huang et al. (2018) and Daniel Koenig and Detlef Weigel, Max Planck Institute for Developmental Biology.

[2] One BAC library containing material from both accessions was produced.

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

Species	**Accession**	**Coverage SMRT raw assembly (x)**	**SMRT N50 read length (bp)**	**MiSeq coverage (x)**	**Final length of *S*-locus contig**
C. grandiflora	Cg2-2-KS2	239	28 737	3 777	115 410
C. grandiflora	Cg88.15	454	24 362	386	128 278
C. orientalis	Co1719/11	150	25 308	2 689	167 282
C. orientalis	Co1979/09	256	27 620	4 693	93 370
C. bursa-pastoris	CbpWEDE9.136	246	24 893	4 215	114 127
C. bursa-pastoris	CbpWESE11.139	399	25 568	4 380	99 499
C. bursa-pastoris	CbpAQ416	338	25 653	3 157	131 978
C. bursa-pastoris	CbpCH1 (CbpKMB205, CbpGY36)[1]	252	27 925	1 967	189 744

[1] One BAC library containing material from both accessions CbpKMB205 and CbpGY36 was produced, and the resulting *S*-haplotype sequence is designated CbpCH1 here.

Table S3. Polymorphism and divergence at the *S*-locus. S is the number of segregating sites, and π is nucleotide diversity. Nonsynonymous and synonymous polymorphism was estimated (Nei & Gojobori 1986) in an alignment with codons containing indels removed.

Species	Region	*n*	Sites	Syn sites[1]	Nonsyn sites[2]	*S*	*S* syn[3]	*S* nonsyn[4]	π syn[5]	π nonsyn[6]	π tot[7]	*Ks* to *CgS12*[8]	*Ka* to *CgS12*[9]	*Ks* to *AhS12*[10]	*Ka* to *AhS12*[11]
C. orientalis	*SRK*	29	2516	562.2	1912.8	2	0	2	0	0.00053	0.00040	0.040	0.0107	0.176	0.0431
	SCR	29	226	51.5	173.5	0	0	0	0	0	0	0.103	0.0173	0.237	0.1008
	U-BOX	29	1125	279.1	845.9	0	0	0	0	0	0	0.063	0.0065	0.208	0.0266
	Non-coding	29	8693	-	-	22	-	-	-	-	0.00091	-	-	-	-
C. bursa-pastoris	*SRK*	4	2550	576.7	1961.3	4	2	2	0.00202	0.00059	0.00092	0.040	0.0108	0.175	0.0424
	SCR	4	215	49.9	163.1	0	0	0	0	0	0	0.107	0.0184	0.247	0.1010
	U-BOX	4	1125	279.0	846.0	1	0	1	0	0.00079	0.00059	0.063	0.0071	0.208	0.0272
	Non-coding	4	10382	-	-	24	-	-	-	-	0.00140	-	-	-	-

[1]Synonymous sites
[2]Nonsynonymous sites
[3]Synonymous segregating sites
[4]Nonsynonymous segregating sites
[5]Synonymous nucleotide diversity
[6]Nonsynonymous nucleotide diversity
[7]Total nucleotide diversity
[8]Synonymous divergence between the focal species and *C. grandiflora CgS12*.
[9]Nonsynonymous divergence between the focal species and *C. grandiflora CgS12*.
[10]Synonymous divergence between the focal species and *A. halleri AhS12*.
[11]Nonsynonymous divergence between the focal species and *A. halleri AhS12*.

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

Accession name	**Tissue**	**Gene name**	**Mean RPKM (standard error)**
Co1719/11	Flower	*SRK*	1.693(0.879)
Co1719/11	Leaf	*SRK*	0.019(0.009)
Co1979/09	Flower	*SRK*	2.585(0.322)
Co1979/09	Leaf	*SRK*	0(0)
Co1719/11	Flower	*SCR*	13.336(6.767)
Co1719/11	Leaf	*SCR*	0(0)
Co1979/09	Flower	*SCR*	0.5(0.277)
Co1979/09	Leaf	*SCR*	0(0)
Co1719/11	Flower	*U-box*	4.157(0.427)
Co1719/11	Leaf	*U-box*	9.582(0.007)
Co1979/09	Flower	*U-box*	5.737(0.213)
Co1979/09	Leaf	*U-box*	8.313(0.075)

Table S5. Results of the timing of the loss of self-incompatibility performed in BEAST with two competing models of population size change, constant and exponential respectively. Positive value of delta aicm indicates a better fit. Median ages and 95% confidence intervals in Ma are given for the MRCA of *C.grandiflora CgS12, C.orientalis* and *C.bursa-pastoris B* representing the upper bound for loss of SI, as well the MRCA of *C.orientalis* and *C.bursa-pastoris* B corresponding to the lower bound.

	Lnl	aicm	delta aicm	Upper bound	Lower bound
Exponential growth	-13966.8	27970.7	8.08	2.57 (2.21 – 2.9)	0.07 (0.05 - 0.1)
Constant	-13969.4	27978.8		2.48 (2.15 - 2.8)	0.08 (0.05 - 0.12)

Supplementary Figures

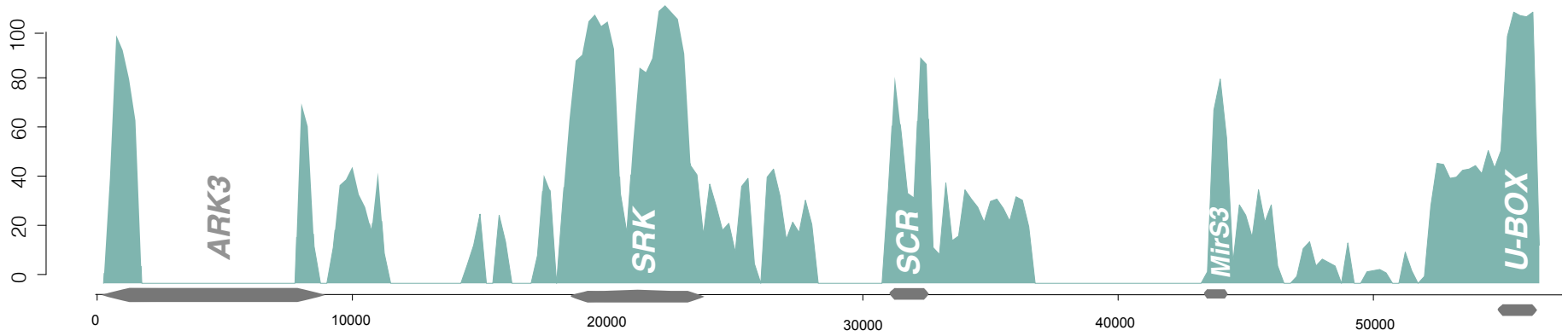

Figure S1. Sequence conservation (% identity) between *C. grandiflora CgS12* and *A. halleri S12*. The low sequence similarity in the middle of *ARK3* is due to the presence of a 6367 bp indel difference between the two sequences.

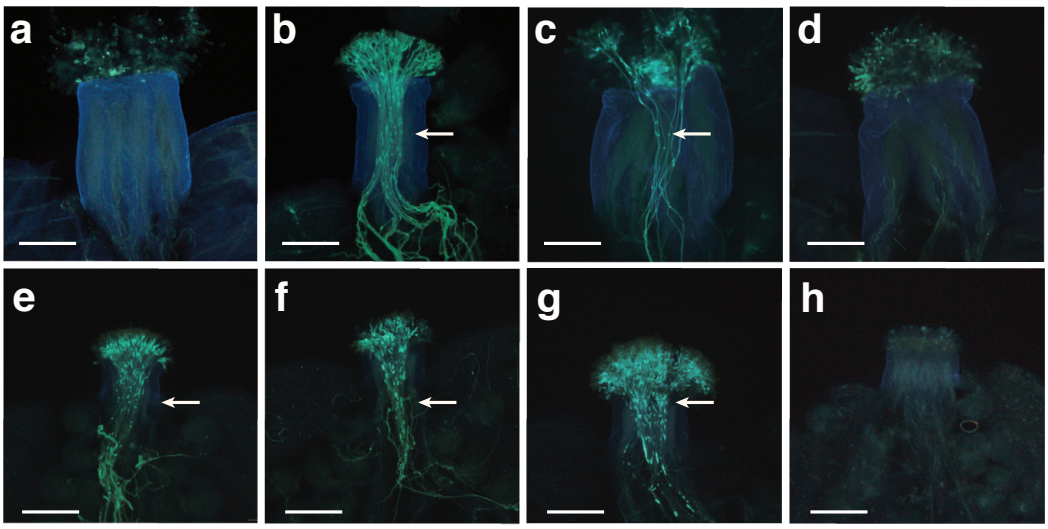

Figure S2. Success of controlled crosses based on pollen tube germination assays. a. Self-pollination of *C. grandiflora* carrying *CgS12* results in no pollen tube growth (incompatible reaction), demonstrating functional SI. **b.** Pollination of *C. grandiflora* carrying *CgS12* with pollen from an individual carrying a different S-haplotype results in pollen tube growth (compatible reaction). **c.** Pollination of *C. grandiflora* carrying *CgS12* with pollen from *C. orientalis* results in pollen tube growth (compatible reaction), demonstrating that *C. orientalis SCR* is not functional. **d.** No pollination of *C. grandiflora* carrying *CgS12* results in no pollen tube growth, demonstrating low risk of accidental cross-pollination in the crossing experiment. **e.** Self-pollination of *C. orientalis* shows self-compatibility (compatible reaction). **f.** Pollination of *C. orientalis* with pollen from a *C. grandiflora* individual carrying a different *S*-haplotype results in pollen tube growth (compatible reaction). **g.** Pollination of *C. orientalis* with pollen from a *C. grandiflora* carrying *CgS12* results in pollen tube growth (compatible reaction). **h.** No pollination of *C. orientalis* results in no pollen tube growth, demonstrating low risk of accidental cross-pollination in the crossing experiment.

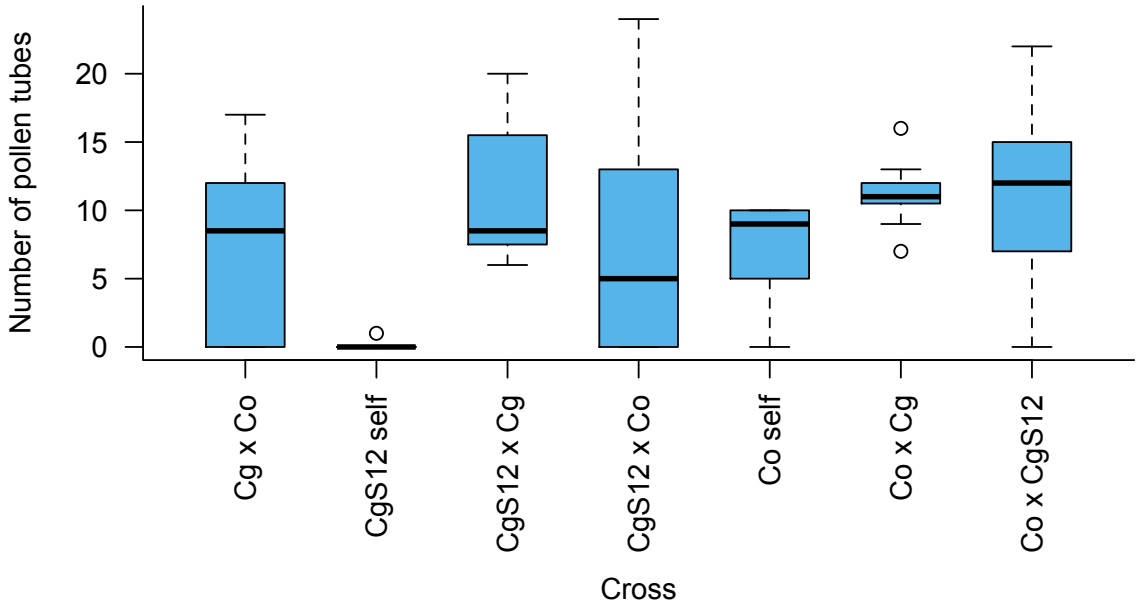

Figure S3. Number of pollen tubes in controlled crosses of *C. orientalis*, *C. grandiflora* harboring *CgS12*, and *C. grandiflora* harboring other *S*-haplotypes. There is a significant difference in the number of pollen tubes depending on the type of cross (Kruskal-Wallis χ^2= 35.126, df=6, *P<0.001*).

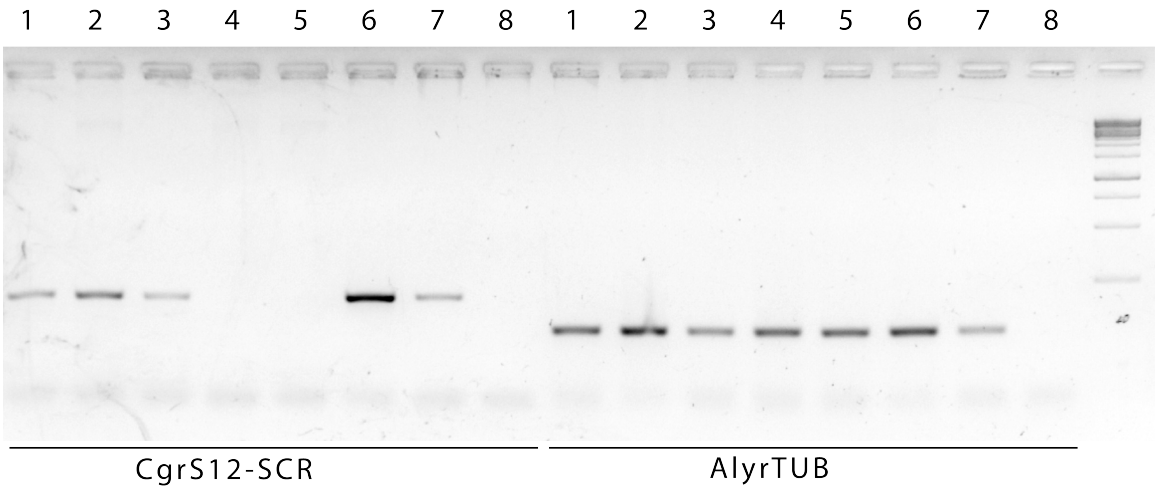

Figure S4. Qualitative RT-PCR demonstrating expression of *SCR* in flower buds of *C. orientalis* and *C. grandiflora* individuals with *CgS12*.

Samples 1-3: different *C. grandiflora* individuals harbouring the *CgS12 S*-allele, Samples 4-5: different *C. grandiflora* individuals with other *S*-alleles, Samples 6-7: two different *C. orientalis* genotypes - Co1719/11 and Co1979/09, respectively, Sample 8: negative control. 1kb GeneRuler DNA ladder for size reference. *CgS12-SCR* is expressed in *C. grandiflora* harbouring the *CgS12 S*-allele and in *C. orientalis,* but not in other *C. grandiflora* individuals. *TUB* is expressed in all samples.

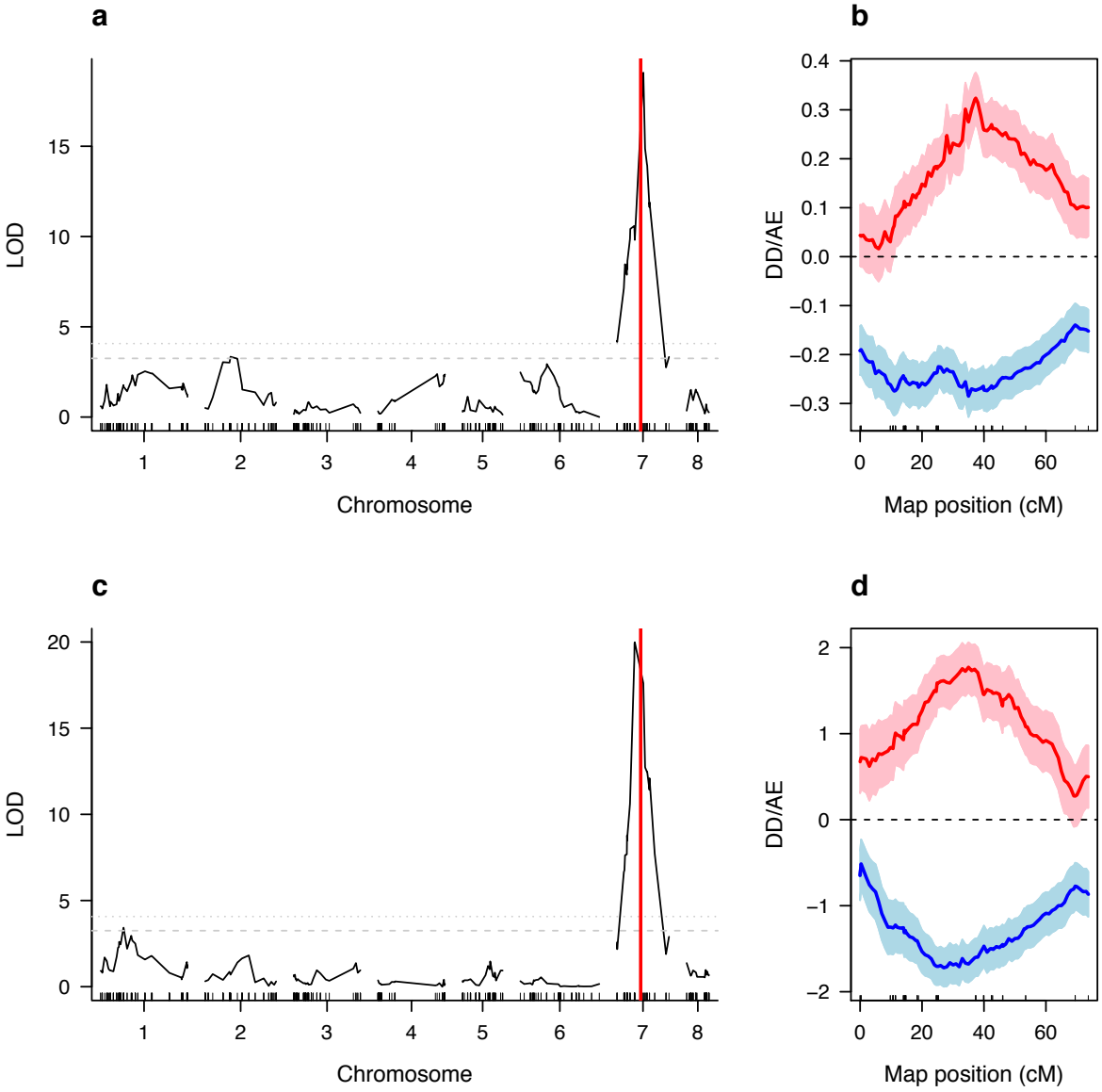

Figure S5. QTL mapping on phenotyping scores separately.
a. LOD profile resulting from interval mapping of SC/SI scores after autonomous self-pollination. The dotted and dashed lines indicate the 1% vs. 5% genome-wide permutation-based significance threshold. The red vertical line shows the location of the canonical Brassicaceae *S*-locus. The 1.5-LOD confidence interval ranges from position 6,241,223 to 8,742,368, whereas the *S*-locus is located between positions 7,523,602 and 7,562,919 on chromosome 7. **b**. Estimated QTL additive effect (red line) and dominance deviation (blue line) across chromosome 7 for the visual SC/SI score. Light shaded regions indicate standard errors. **c**. LOD profile resulting from interval mapping of the number of siliques formed after manual hand-pollination. The 1.5-LOD confidence interval ranges from position 5,061,589 to 8,304,045. **d**. Estimated QTL additive effect (red line) and dominance deviation (blue line) across chromosome 7 for the number of siliques formed. Light shaded regions indicate standard errors.

512 **References**
513
514 Bachmann JA, Tedder A, Laenen B, Steige KA, Slotte T. 2018. Targeted long-read
515 sequencing of a locus under long-term balancing selection in *Capsella*. *G3* 8:1327–1333.
516 Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. A., & Lemey, P. 2012. Accurate
517 model selection of relaxed molecular clocks in Bayesian phylogenetics. Mol. Biol.
518 Evol. 30(2): 239-243.
519 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
520 sequence data. *Bioinformatics* 30:2114–2120.
521 Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A,
522 Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis.
523 *PLoS Comp Biol* 10:e1003537.
524 Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental
525 crosses.
526 Castric V, Bechsgaard J, Schierup MH, Vekemans X. 2008. Repeated adaptive introgression
527 at a gene under multiallelic balancing selection. *PLoS Genet* 4:e1000168.
528 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del
529 Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and
530 genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498.
531 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
532 Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–
533 21.
534 Douglas GM, Gos G, Steige KA, Salcedo A, Holm K, Josephs EB, Arunkumar R, Agren JA,
535 Hazzouri KM, Wang W, et al. 2015. Hybrid origins and the earliest stages of
536 diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proc*
537 *Natl Acad Sci USA* 112:2806–2811.
538 Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti
539 and the BEAST 1.7. *Mol Biol Evol* 29:1969–1973.
540 Durand E, Méheust R, Soucaze M, Goubet PM, Gallina S, Poux C, Fobis-Loisy I, Guillon E,
541 Gaude T, Sarazin A, et al. 2014. Dominance hierarchy arising from the evolution of a
542 complex small RNA regulatory network. *Science* 346:1200–1205.
543 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
544 throughput. *Nucleic Acids Res* 32:1792–1797.
545 Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform graphical user
546 interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27:221–
547 224.
548 Guo Y-L, Bechsgaard JS, Slotte T, Neuffer B, Lascoux M, Weigel D, Schierup MH. 2009.
549 Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of
550 self-incompatibility and an extreme bottleneck. *Proc Natl Acad Sci USA* 106:5246–5251.
551 Hurka H, Friesen N, German DA, Franzke A, Neuffer B. 2012. “Missing link” species
552 *Capsella orientalis* and *Capsella thracica* elucidate evolution of model plant genus
553 *Capsella* (Brassicaceae). *Mol Ecol* 21:1223–1238.
554 Huang H-R, Liu J-J, Xu Y, Lascoux M, Ge X-J, Wright SI. 2018. Homeologue-specific
555 expression divergence in the recently formed tetraploid *Capsella bursa-pastoris*
556 (Brassicaceae). *New Phytol* 42:e46–e635.
557 Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database
558 management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.
559 Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple
560 sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066.
561 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable

562 and accurate long-read assembly via adaptive k-mer weighting and repeat separation.
563 *Genome Res* 27:722–736.
564 Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. Partitionfinder: combined selection of
565 partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol*
566 29:1695–1701.
567 Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: New
568 methods for selecting partitioned models of evolution for molecular and morphological
569 phylogenetic analyses. *Mol Biol Evol* 34:772–773.
570 Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large
571 datasets. *Bioinformatics* 30:3276–3278.
572 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-
573 MEM. arXiv:1303.3997v2, https://arxiv.org/abs/1303.3997v2.
574 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
575 Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce
576 framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–
577 1303.
578 Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying
579 mammalian transcriptomes by RNA-Seq. *Nat Methods* 5:621–628.
580 Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and
581 nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
582 Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel
583 D, Lynch M. 2010. The rate and molecular spectrum of spontaneous mutations in
584 *Arabidopsis thaliana. Science* 327:92–94.
585 R Core Team 2017. R: A language and environment for statistical computing.
586 Robinson JT, Thorvaldsdóttir H, Winckler H, Guttman M, Lander ES, Getz G & Mesirov JP.
587 2011. Integrative genomics viewer. *Nature Biotechnology* 29:24–26
588 Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE,
589 Sanchez-Gracia A. 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large
590 Data Sets. Mol Biol Evol 34:3299–3302.
591 Shiba H, Kakizaki T, Iwano M, Tarutani Y, Watanabe M, Isogai A, Takayama S. 2006.
592 Dominance relationships between self-incompatibility alleles controlled by DNA
593 methylation. *Nat Genet* 38:297–299.
594 Slotte T, Huang H-R, Holm K, Ceplitis A, Onge KS, Chen J, Lagercrantz U, Lascoux M.
595 2009. Splicing variation at a *FLOWERING LOCUS C* homeolog is associated with
596 flowering time variation in the tetraploid *Capsella bursa-pastoris*. Genetics 183:337–345.
597 Slotte T, Hazzouri KM, Agren JA, Koenig D, Maumus F, Guo Y-L, Steige K, Platts AE,
598 Escobar JS, Newman LK, et al. 2013. The *Capsella rubella* genome and the genomic
599 consequences of rapid mating system evolution. *Nat Genet* 45:831–835.
600 Smith, TF, Waterman, MS. 1981. Identification of common molecular subsequences. *J Mol*
601 *Biol* 147:195-197.
602 Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene
603 finding in eukaryotes. *Nucleic Acids Res* 32:W309–W312.
604 Steige KA, Laenen B, Reimegård J, Scofield DG, Slotte T. 2017. Genomic analysis reveals
605 major determinants of *cis*-regulatory variation in *Capsella grandiflora*. *Proc Natl Acad*
606 *Sci USA* 114:1087–1092.
607 Tsuchimatsu T, Kaiser P, Yew C-L, Bachelier JB, Shimizu KK. 2012. Recent loss of self-
608 incompatibility by degradation of the male component in allotetraploid *Arabidopsis*
609 *kamchatica. PLoS Genet* 8:e1002838.
610 Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A,
611 Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high

612 confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr
613 Protoc Bioinformatics 11:11.10.1–11.10.33.
614 Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, Wright SI.
615 2014. Evidence for widespread positive and negative selection in coding and conserved
616 noncoding regions of *Capsella grandiflora*. PLoS Genet 10:e1004622.