Open access • Journal Article • DOI:10.1038/NATURE08927

Evolution of self-compatibility in Arabidopsis by a mutation in the male specificity gene — Source link

Takashi Tsuchimatsu, Keita Suwabe, Keita Suwabe, Rie Shimizu-Inatsugi ...+7 more authors

Institutions: University of Zurich, Tohoku University, Mie University, Ludwig Maximilian University of Munich ...+3 more institutions

Published on: 29 Apr 2010 - Nature (Nature Publishing Group)

Topics: Selfing and Locus (genetics)

Evolution of self-compatibility in Arabidopsis by a mutation in the male specificity gene

Tsuchimatsu, T; Suwabe, K; Shimizu-Inatsugi, R; Isokawa, S; Pavlidis, P; Städler, T; Suzuki, G; Takayama, S; Watanabe, M; Shimizu, KK

Evolution of self-compatibility in *Arabidopsis* by a mutation in the male specificity gene

Takashi Tsuchimatsu[1,2]*, Keita Suwabe[3,4]*, Rie Shimizu-Inatsugi[1], Sachiyo Isokawa[3,5], Pavlos Pavlidis[6], Thomas Städler[7], Go Suzuki[8], Seiji Takayama[9], Masao Watanabe[3,5] & Kentaro K. Shimizu[1]

[1]Institute of Plant Biology, University Research Priority Program in Systems Biology/Functional Genomics & Zürich-Basel Plant Science Center, University of Zurich, Zollikerstrasse 107, CH-8008 Zurich, Switzerland. [2]Department of General Systems Studies, University of Tokyo, 3-8-1 Komaba, Meguro, Tokyo 153-8902, Japan. [3]Graduate School of Life Sciences, Tohoku University, Katahira, Sendai 980-8577, Japan. [4]Graduate School of Bioresources, Mie University, Tsu 514-8507, Japan. [5]Faculty of Science, Tohoku University, Aoba, Sendai 980-8578, Japan. [6]Section of Evolutionary Biology, BioCenter, University of Munich (LMU), Grosshaderner Strasse 2, D-82152 Planegg-Martinsried, Germany. [7]Plant Ecological Genetics, Institute of Integrative Biology, ETH Zurich, Universitätstrasse 16, CH-8092 Zurich, Switzerland. [8]Division of Natural Science, Osaka Kyoiku University, Kashiwara 582-8582, Japan. [9]Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma 630-0101, Japan.

*These authors contributed equally to this work.

Ever since Darwin's pioneering research, the evolution of self-fertilisation (selfing) has been regarded as one of the most prevalent evolutionary transitions in flowering plants[1,2]. A major mechanism to prevent selfing is the self-incompatibility (SI) recognition system, which consists of male and female specificity genes at the *S*-locus and SI modifier genes[2–4]. Under conditions that

favour selfing, mutations disabling the male recognition component are predicted to enjoy a relative advantage over those disabling the female component, because male mutations would increase through both pollen and seeds whereas female mutations would increase only through seeds[5,6]. Despite many studies on the genetic basis of loss of SI in the predominantly selfing plant *Arabidopsis thaliana*[7–15], it remains unknown whether selfing arose through mutations in the female specificity gene (*S*-receptor kinase, *SRK*), male specificity gene (*S*-locus cysteine-rich protein, *SCR*, also known as *S*-locus protein 11, *SP11*) or modifier genes, and whether any of them rose to high frequency across large geographic regions. Here we report that a disruptive 213-bp inversion in the *SCR* gene (or its derivative haplotypes with deletions encompassing the entire *SCR-A* and a large portion of *SRK-A*) is found in 95% of European accessions, which contrasts with the genome-wide pattern of polymorphism in European *A. thaliana*[16,17]. Importantly, interspecific crossings using *Arabidopsis halleri* as pollen donor reveal that some *A. thaliana* accessions, including Wei-1, retain the female SI reaction, suggesting that all female components including SRK are still functional. Moreover, when the 213-bp inversion in *SCR* was inverted and expressed in transgenic Wei-1 plants, the functional *SCR* restored the SI reaction. The inversion within *SCR* is the first mutation disrupting SI shown to be nearly fixed in geographically wide samples, and its prevalence is consistent with theoretical predictions regarding the evolutionary advantage of mutations in male components.

Selfing is disadvantageous when selfed offspring suffer reduced fitness (that is, inbreeding depression), but it may nevertheless be favoured due to reproductive assurance when pollinators or mates are scarce, as first proposed by Darwin[1,18]. It is also favoured by an inherent transmission advantage because a selfing individual can transmit gametes in three ways (as both ovule and pollen parent to its own selfed progeny and as pollen parent in outcrossing), whereas an outcrossing individual cannot serve as pollen parent in selfing[18,19]. Thus, an allele promoting selfing has a 3:2 transmission advantage relative to outcrossing when all other conditions are equal[18,19]. In many plant lineages, predominant selfing evolved through the loss of self-incompatibility (SI) and changes in floral traits[2,18]. The molecular basis of the SI system has been extensively studied using the family Brassicaceae, where it is controlled by *SRK* and *SCR*, encoding the female and male SI specificity determinants, respectively, at the *S*-locus[3,4,11]. SRK is a transmembrane serine/threonine receptor kinase that functions on the stigma, and SCR is a small cysteine-rich protein present in the pollen coat that acts as a ligand of the SRK receptor protein[20–22]. Dozens of highly divergent sequence groups, called *S*-haplogroups (or *S*-haplotypes or *S*-alleles), confer specificity in self-recognition: direct interaction between SCR and SRK of the same *S*-haplogroup leads to downstream signalling in female tissues, which inhibits pollen germination on the stigma[3,4]. Haplogroups are characterized by low nucleotide diversity within each haplogroup and high divergence and suppressed recombination between haplogroups; these patterns extend to flanking genes of the *S*-locus to some extent[11,23].

A. thaliana is a self-compatible, predominantly selfing species of Brassicaceae[24], whereas obligate outcrossing enforced by the SI system and inbreeding depression is observed in its closely related congeners[23,25]. Only three haplogroups at the *S*-locus have been found in *A. thaliana*; haplogroups A and C are distributed in Europe, North America, North Africa and Asia, whereas haplogroup B is found only in offshore African islands[9–11]. While there is no evidence for inter-haplogroup recombination between *SRK* and *SCR*[10,11] (Supplementary Note 1), several independent mutations and rearrangements disrupting *SRK* and *SCR* have been identified[7–13]. When pairs of *SRK* and *SCR* from self-incompatible *Arabidopsis lyrata* were introduced into *A. thaliana* heterologously by transgenic experiments, some accessions showed developmentally stable SI and produced very few seeds, suggesting the disruption of the *S*-locus in *A. thaliana*[8,12]. In other transformed accessions, SI was detected only in young flowers and no clear reduction in seed set was observed, which was considered as pseudo-self-compatibility[8,12,13]. This cryptic variation in the strength of SI among transgenic plants was attributed to polymorphism in the promoter region of a female modifier gene, *PUB8*, adjacent to the *S*-locus[13]. However, whether mutations in the male or female specificity gene at the *S*-locus have contributed to the loss of SI cannot be assessed by such heterologous transgenic experiments. Most importantly, it is difficult to distinguish the primary inactivating mutation from subsequent decay by further mutations[5,12,14]. Thus, it is still unclear which mutation(s) contributed to the loss of SI and the transition to predominant selfing, and whether any mutation conferring self-compatibility is geographically widespread[5–15].

To search for high-frequency mutations underlying self-compatibility and to delimit the affected genomic interval, we first analysed the geographic distribution of haplogroups in the genomic region encompassing the *S*-locus. Because allele frequencies are affected by linkage to neighbouring genes and by demographic factors that are expected to affect the entire genome, it is critical to take non-random patterns of genome-wide polymorphisms into account. We examined the European accessions of *A. thaliana* for which an east–west gradient or clustering of accessions, as well as the divergence of northern accessions, were described based on a large data set[16,17]. This population structure is generally attributed to a scenario in which *A. thaliana* persisted and diverged in eastern and western refugia during glacial periods and subsequently spread across Europe[16,17] (Supplementary Note 1). Previous studies of European accessions revealed two distinct haplogroups at the *S*-locus (haplogroups A and C), as well as two distinct haplogroups for the *ARK3* and *PUB8* genes which flank the *S*-locus on each side[10,11]. Our sequence data show that the flanking genes also exhibit high sequence divergence between haplogroups (0.0372 in *ARK3* and 0.0321 in *PUB8*), while nucleotide diversity within each haplogroup is up to two orders of magnitude lower (0.00031 in *ARK3-W*, 0.00078 in *ARK3-E*, 0.00063 in *PUB8-W* and 0.00103 in *PUB8-E*; see also Supplementary Figs 1, 2). To survey whether the haplogroup frequencies at each locus are concordant with genome-wide population structure, we exploited the clustering described in ref. 17 and calculated the correlation between haplogroups and the clusters. For both *ARK3* and *PUB8*, we found a significant correlation (Cramer coefficient 0.262 in *ARK3*, $P = 0.00177$; 0.224 in *PUB8*,

P = 0.0152; Fig. 1; Supplementary Note 1 and Supplementary Figs 1–3). Haplogroups *ARK3-W* and *PUB8-W* were found mainly in western Europe, in the east–west transition zone and northern Europe, whereas haplogroups *ARK3-E* and *PUB8-E* were more common in eastern Europe (Fig. 1 and Supplementary Table 1). Similarly, we found a significant correlation between population structure and length polymorphism in the *PUB8* promoter region (Cramer coefficient 0.223, *P* = 0.0252), which has been reported to be responsible for cryptic SI variation[13]. These correlations reflect significant differences in haplogroup frequencies depending on the clusters, indicating that the haplogroup patterns at these genes are consistent with the genome-wide pattern of polymorphism.

In contrast to the intermediate representation of haplogroups in the flanking genes, the frequency of haplogroup A at the *S*-locus (gauged by data on *SRK*; Supplementary Note 1) is markedly higher than that of haplogroup C and thus is widespread throughout Europe (Fig. 1), resulting in no significant correlation between our *SRK* haplogroup data and population structure (Cramer coefficient 0.0864, *P* = 0.750). These results indicate that although genome-wide differentiation can explain the pattern of polymorphism in the flanking regions of the *S*-locus, additional evolutionary forces might be necessary to explain the pattern of polymorphism at the *S*-locus itself. Because only *SRK*, *SCR* and repetitive elements have been identified in the genomic region between the *ARK3* and *PUB8* genes, the data are consistent with a

mutation conferring self-compatibility having occurred in either *SRK* or *SCR* whose increase in frequency might have been driven by positive selection.

To dissect the potential historical contributions of gene-disruptive mutations at the *S*-locus, we took advantage of trans-specific sharing of *S*-haplogroups among *Arabidopsis* species[23]. We obtained the entire coding sequences of *SCR-A* and *SRK-A* of the functional haplogroup A from *A. halleri*, a self-incompatible close relative of *A. thaliana*[23] (Fig. 2 and Supplementary Fig. 4). The predicted SCR amino acid sequence contains eight conserved cysteine residues known to be important for protein structure (Fig. 2b). *SCR-A* of *A. thaliana* was reported to be truncated, encoding only three of the eight conserved cysteine residues[7], but subsequent study showed that the truncated protein is not caused by a deletion but by a 213-bp inversion in the coding sequence[12]. Interestingly, if the inverted region is reverted, *SCR-A* of Col-0 would contain the entire coding information of SCR-A[11] (Fig. 2). To survey the frequency of this mutation across the European range of *A. thaliana*, we examined a larger collection of 277 accessions and found that 95% of them possess this inversion or its derivative deletion (named A-t2, with a ~23-kb deletion encompassing the entire *SCR-A* and a large portion of *SRK-A*; Supplementary Note 1)[11]. These results indicate that the *SCR-A* inversion mutation is nearly fixed in Europe. In addition, the inversion and A-t2 were also found in accessions from Tajikistan, Kazakhstan and Libya, demonstrating that they are also distributed outside Europe.

In contrast, alignment of the coding regions of *SRK-A* from *A. halleri* and *A. thaliana* shows that at least 12 accessions of *A. thaliana*, including Wei-1 and Old-1, still have the full-length *SRK-A* gene without any apparent disruptive mutations, such as frameshifts or inverted repeats (Supplementary Fig. 4). These accessions are scattered widely across Europe, indicating that the intact *SRK-A* variant is not restricted to isolated regions (Supplementary Fig. 5). Given its tight linkage to *SCR-A*, these observations indicate that the frequency of *SRK-A* might have increased by hitchhiking and that *SRK-A* of some accessions might still be functional, even though the majority of accessions have a pseudogenized *SRK-A* due to multiple disruptive mutations[7,11].

To test whether SRK and the components of the female signalling pathway are functional in *A. thaliana*, we carried out interspecific crosses between *A. halleri* (male) and *A. thaliana* (female), as such crosses are successful when *A. thaliana* is used as the female parent (Methods; Supplementary Table 2). Because *S*-allele lineages are shared trans-specifically among several *Arabidopsis* species[23], an incompatible reaction should occur even in interspecific crosses where pollen and stigma share the same haplogroup[26]. Indeed, when pollen of *A. halleri* bearing *SCR-A* was used to pollinate emasculated pistils of the 12 *A. thaliana* accessions with intact *SRK-A* (see above), an incompatible reaction was observed in seven accessions (Fig. 3; Supplementary Fig. 6 and Supplementary Table 2). This indicates that female SI is functional in these accessions. We further examined the strength of incompatibility during flower development. As selfing in *A. thaliana* occurs when stamens touch stigmas autonomously

(autopollination) and thus pollination afterwards would contribute little to fertilisation, an incompatibility reaction before and during the flower developmental stage of autopollination would be most relevant physiologically to prevent selfing. Because autopollination occurs during a certain range of pistil length, we used pistil length as a continuous variable to measure developmental stage in emasculated flowers. A constitutive incompatible reaction up to the stage of autopollination was observed in four of the seven accessions including Wei-1 (Fig. 3b; Supplementary Fig. 6 and Supplementary Note 2), whereas the incompatible reaction was attenuated in older flowers.

To exclude the possibility that pollen tube growth was inhibited not by an incompatible reaction but by interspecific reproductive isolation, we also conducted interspecific crosses with *A. halleri* bearing other *S*-alleles and confirmed that all such crossings were compatible. In addition, crossings between *A. thaliana* with degenerated *SRK* and *A. halleri* with haplogroup A were compatible (Fig. 3; Supplementary Table 2). These results indicate that reduced pollen growth, where observed, was not caused by interspecific reproductive isolation but by the SI system. Taken together, these crossing experiments confirm that haplogroup A of *A. thaliana* and *A. halleri* share the same functional specificity and, most importantly, demonstrate that at least these accessions of *A. thaliana* harbour functional alleles of *SRK* and other genes required for the female SI reaction. Thus, degradation of the male component was responsible for their loss of SI.

Because our crosses revealed that the female SI function has been retained in at least four *A. thaliana* accessions, we reasoned that they would become self-incompatible if the normal function of the male *SCR* gene could be restored. To restore the function of *SCR-A*, we inverted the 213-bp segment that disrupts the coding region to yield a full-length open reading frame (ORF). To alleviate the effect of co-suppression, which is pronounced for the *S*-locus genes[3,27], we used the tapetum-specific *ATA7* promoter instead of the *SCR* promoter to express the ORF (see Methods and Supplementary Fig. 7), because the intrinsic *SCR-A* promoter is still active in *A. thaliana*[11]. In this way, *ATA7::restored SCR* was introduced into Wei-1. Upon selfing, pollen of the transgenic plants was rejected on the stigmas (Fig. 4a, b). This finding strongly indicates that the disruption of *SCR-A* resulted in the evolutionary loss of SI. In addition, our data indicate that the inversion was the responsible mutation, although we cannot exclude the possibility that changes in *SCR-A* expression also played a role. Although other minor-frequency mutations (possibly in haplogroup C) might have also contributed to self-compatibility, our findings are consistent with the hypothesis that the non-functional *SCR-A* led to the near-fixation of haplogroup A at the *S*-locus (*SCR-A* and *SRK-A*). This scenario for the increase in the frequency of self-compatibility within *A. thaliana* contrasts with the proposed loss of SI in *Capsella rubella* through an extreme bottleneck associated with speciation, inferred to have occurred without replacing other *S*-locus haplogroups by selection[28].

The transgenic plants produced shorter siliques than non-transgenic Wei-1 plants and fewer seeds per silique, consistent with our finding that the female SI function was observed up to the developmental stage of autopollination (Fig. 4c–e; Supplementary Note 4). Our crossing and transgenic experiments that used plants possessing native *SRK* sequences demonstrate the prevention of selfing by a SI reaction combined with later attenuation of SI, which was not reported from the use of heterologous transgenic plants[8,12,13]. We suggest that the disruption of *SCR-A* was a critical step in the evolution of predominant selfing in *A. thaliana* in conjunction with other morphological and physiological changes. Our results also reveal the diversity in SI response along developmental stages and among accessions (Supplementary Note 2, Supplementary Table 2 and Supplementary Fig. 6). This might indicate that mutations conferring partial selfing with pseudo-self-compatibility have arisen before the loss of SI, although it is conceivable that they represent secondary decay (Supplementary Note 3).

A fairly recent origin of self-compatibility in *A. thaliana* has been suggested based on the K_a/K_s ratio (the ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions) for the *S*-locus genes (<413,000 years ago)[23]. Our results are consistent with the recent spread of self-compatibility. First, except for *SCR*, all SI components are still functional in a number of accessions, although it is possible that functional alleles have been maintained due to pleiotropy for biological functions other than SI. Second, it is generally thought that the current population structure of European *A. thaliana* was shaped by range expansions from multiple refugia during

recent glacial–interglacial periods[16,17]. Significant deviation at the *S*-locus from the genome-wide population structure implies that haplogroup A (harbouring the *SCR-A* inversion) increased non-randomly, replacing other haplogroups during the process of admixture from multiple source populations. Range expansions tend to be accompanied by reduced mate availability and increased pollen limitation[2,29]. Coupled with an inherent transmission advantage[19], reproductive assurance might have driven the evolution of selfing enabled by the loss of SI as first proposed by Darwin[1], thus overcoming inbreeding depression in ancestral outcrossing populations[18,25].

If evolutionary constraints or selection were unimportant, female mutations leading to self-compatibility would be prevalent especially in the Brassicaceae, because the *SRK* coding region is ~10 times longer than that of *SCR*. Hence, the rate of disruptive mutations, such as premature stop or frameshift mutations, is expected to be higher (assuming similar deleterious mutation rates). Indeed, a number of loss-of-function mutations in the female gene have been identified, including those that might have occurred after the *SCR* inversion and those responsible for self-compatibility in cultivated and geographically restricted populations[3,5,7,10,11,26,30]. In cultivated *Brassica*, genetic crosses and molecular experiments have identified loss-of-function mutations in *SRK* as well as in *SCR*[3,26,30], and it is conceivable that these were selected in the course of human cultivation based on the self-compatible phenotype *per se*. Using crossing and transgenic experiments, we have shown that all female SI components are functional in a number of European *A. thaliana* accessions. The pseudogenized *SCR-A*

represents the first mutation conferring self-compatibility shown to be nearly fixed in geographically wide samples. Our findings are consistent with theoretical predictions that male mutations causing loss of SI are more likely to spread than those disrupting female components under both reproductive assurance and transmission advantage[5,6]. Our work shows that the experimental reversal of evolution is a powerful tool in disentangling mutational history, especially when combined with analyses of genomic polymorphism.

METHODS SUMMARY

Plant materials

A. thaliana ecotypes listed in Supplementary Table 3 were obtained from the *Arabidopsis* Biological Resource Center (ABRC; www.biosci.ohio-state.edu/pcmb/Facilities/abrc/abrchome.htm). *A. halleri* was collected in Japan, Germany and Switzerland.

Pollination assay

Anthers were removed from flower buds and stigmas were examined for the absence of contaminating pollen by microscopy. At 0, 12, 24 and 36 h later, pollen grains were manually applied to the stigmas. After fixation, softening and staining, the pollen tubes were observed using epifluorescence microscopy. All statistical analyses were performed using R 2.8.1 (www.r-project.org).

Isolation of genomic/complementary DNA, genotyping and sequencing

Genomic DNA and total RNA were extracted using the Plant DNeasy Mini and RNeasy kits (Qiagen), respectively. Complementary DNA was synthesized with the RETROscript reverse transcription kit (Ambion). PCR was performed with Taq DNA polymerase (Roche), Go-Taq polymerase (Promega) or ExTaq (TaKaRa). *SCR-A* and *SRK-A* sequences of *A. halleri* were obtained using the BD GenomeWalker Universal Kit (BD Biosciences).

Transformation of restored *SCR*

The complete coding sequence of the *SCR* gene of Col-0 was restored by a series of PCR amplifications. It was then joined to the *ATA7* promoter and the construct was subcloned into the pBI121 vector, introduced into *Agrobacterium tumefaciens* strain EHA105 by electroporation and transformed into *Arabidopsis* plants (Wei-1).

1. Darwin, C. *The Effects of Cross and Self Fertilisation in the Vegetable Kingdom* (J. Murray, London, 1876).
2. Stebbins, G. L. *Flowering Plants: Evolution Above the Species Level* (Harvard University Press, Cambridge, Massachusetts, 1974).
3. de Nettancourt, D. *Incompatibility and Incongruity in Wild and Cultivated Plants, 2nd ed.* (Springer, Berlin, 2001).
4. Takayama, S. & Isogai, A. Self-incompatibility in plants. *Annu. Rev. Plant Biol.* **56,** 467–489 (2005).
5. Busch, J. W. & Schoen, D. J. The evolution of self-incompatibility when mates are limiting. *Trends Plant Sci.* **13,** 128–136 (2008).
6. Uyenoyama, M. K., Zhang, Y. & Newbigin, E. On the origin of self-incompatibility haplotypes: transition through self-compatible intermediates. *Genetics* **157,** 1805–1817 (2001).
7. Kusaba, M. *et al.* Self-incompatibility in the genus *Arabidopsis*: characterization of the *S* locus in

the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *Plant Cell* **13,** 627–643 (2001).

8. Nasrallah, M. E., Liu, P. & Nasrallah, J. B. Generation of self-incompatible *Arabidopsis thaliana* by transfer of two *S* locus genes from *A. lyrata*. *Science* **297,** 247–249 (2002).
9. Sherman-Broyles, S. *et al. S* locus genes and the evolution of self-fertility in *Arabidopsis thaliana. Plant Cell* **19,** 94–106 (2007).
10. Tang, C. *et al.* The evolution of selfing in *Arabidopsis thaliana*. *Science* **317,** 1070–1072 (2007).
11. Shimizu, K. K., Shimizu-Inatsugi, R., Tsuchimatsu, T. & Purugganan, M. D. Independent origins of self-compatibility in *Arabidopsis thaliana*. *Mol. Ecol.* **17,** 704–714 (2008).
12. Boggs, N. A., Nasrallah, J. B. & Nasrallah, M. E. Independent *S*-locus mutations caused self-fertility in *Arabidopsis thaliana*. *PLoS Genet.* **5,** e1000426 (2009).
13. Liu, P., Sherman-Broyles, S., Nasrallah, M. E. & Nasrallah, J. B. A cryptic modifier causing transient self-incompatibility in *Arabidopsis thaliana*. *Curr. Biol.* **17,** 734–740 (2007).
14. Igic, B., Lande, R. & Kohn, J. R. Loss of self-incompatibility and its evolutionary consequences. *Int. J. Plant Sci.* **169,** 93–104 (2008).
15. Mable, B. K. Genetic causes and consequences of the breakdown of self-incompatibility: case studies in the Brassicaceae. *Genet. Res.* **90,** 47–60 (2008).
16. Sharbel, T. F., Haubold, B. & Mitchell-Olds, T. Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Mol. Ecol.* **9,** 2109–2118 (2000).
17. François, O., Blum, M. G. B., Jakobsson, M. & Rosenberg, N. A. Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genet*. **4,** e1000075 (2008).
18. Goodwillie, C., Kalisz, S. & Eckert, C. G. The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* **36,** 47–79 (2005).
19. Fisher, R. A. Average excess and average effect of a gene substitution. *Ann. Eugen*. **11,** 53–63 (1941).
20. Schopfer, C. R., Nasrallah, M. E. & Nasrallah, J. B. The male determinant of self-incompatibility in *Brassica. Science* **286,** 1697–1700 (1999).
21. Takasaki, T. *et al.* The *S* receptor kinase determines self-incompatibility in *Brassica* stigma. *Nature* **403,** 913–916 (2000).
22. Takayama, S. *et al*. Direct ligand–receptor complex interaction controls *Brassica* self-incompatibility. *Nature* **413,** 534–538 (2001).
23. Bechsgaard, J. S., Castric, V., Charlesworth, D., Vekemans, X. & Schierup, M. H. The transition to self-compatibility in *Arabidopsis thaliana* and evolution within *S*-haplotypes over 10 Myr. *Mol.*

Biol. Evol. **23,** 1741–1750 (2006).

24. Abbott, R. J. & Gomes, M. F. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* **62,** 411–418 (1989).

25. Kärkkäinen, K. *et al*. Genetic basis of inbreeding depression in *Arabis petraea*. *Evolution* **53,** 1354–1365 (1999).

26. Okamoto, S. *et al*. Self-compatibility in *Brassica napus* is caused by independent mutations in *S*-locus genes. *Plant J.* **50,** 391–400 (2007).

27. Shiba, H., Hinata, K., Suzuki, A. & Isogai, A. Breakdown of self-incompatibility in *Brassica* by the antisense RNA of *SLG* gene. *Proc. Japan Acad. Ser. B* **71,** 81–83 (1995).

28. Guo, Y.-L. *et al.* Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proc. Natl Acad. Sci. U S A* **106,** 5246–5251 (2009).

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

30. Goring, D. R., Glavin, T. L., Schafer, U. & Rothstein, S. J. An *S* receptor kinase gene in self-compatible *Brassica napus* has a 1-bp deletion. *Plant Cell* **5,** 531–539 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Awadalla, C. Bustamante, A. Caicedo, G. Coop, U. Grossniklaus, T.-h. Kao, C. Mays, R. Moore, K. Olsen, M. Purugganan, J. Reininga, L. Rose, S. Ruzsa and W. Stephan for discussions or technical advice. This work was supported by grants from the University Research Priority Program in Systems Biology/Functional Genomics of the University of Zurich and from the Swiss National Science Foundation (SNF) to K.K.S., and by Grants-in-Aid for Special Research on Priority Areas to S.T., M.W. and K.K.S.; by a grant from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT) and by a Grant-in-Aid for Creative Scientific Research to S.T.; by a Young Scientific Research (S) grant to M.W.; and by a grant from the Japan Society for the Promotion of Science (JSPS). P.P. is the recipient of a grant from the Volkswagen Foundation and a STIBET

scholarship of the Deutscher Akademischer Austauschdienst (DAAD). T.T. and K.S. are recipients of a Research Fellowship for Young Scientists from JSPS.

Author Contributions T.T., K.S., M.W. and K.K.S. conceived and designed the study; T.T., R.S.-I. and K.K.S. performed sequencing, genotyping and crossing experiments; T.T., P.P., T.S. and K.K.S. conducted molecular population genetic analysis; K.S., S.I., S.T. and M.W. conducted transgenic analysis; T.T., K.S., G.S., T.S., M.W. and K.K.S. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Sequence data have been deposited to GenBank under accession numbers GU723782–GU723953. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.K.S. (shimizu@botinst.uzh.ch) (sequence analysis and crossing experiments) or M.W. (nabe@ige.tohoku.ac.jp) (transgenic experiments).

Figure 1 The relationship between population structure and haplogroup frequencies of the *S*-locus and flanking genes, *PUB8* and *ARK3*. In European *A. thaliana*, there are two distinct haplogroups at the *S*-locus (A and C) and two distinct haplogroups for both *ARK3* and *PUB8* (named E and W). Four clusters in Europe are based on ref. 17 (“Western”, “Contact zone”, “Eastern” and “Northern”; Supplementary Note 1 and Supplementary Table 1). The schematic genomic structure depicted along the top of the figure is based on the genomic sequence of Col-0 (GenBank accession number NC_003075). The correlation between haplogroup and population structure is statistically significant for *ARK3* and *PUB8*, but not for the *S*-locus (see text).

Figure 2 Disrupted *SCR-A* in *A. thaliana*. a, Schematic genomic structures of *SCR-A* of *A. halleri* and *A. thaliana* (Col-0). For the Col-0 structure, grey triangles denote the locations of deletions and white triangles denote the locations of insertions. All these indels are located in the intron and there are no disruptive mutations in the exons, except for a 213-bp inversion and a 14-bp duplication. The 14-bp duplication was found in several accessions, including Col-0, but not in others (Nok-0 and Pog-0; see also Supplementary Note 1). **b**, Alignment of predicted amino acid sequences of SCR-A of *A. halleri* and Col-0, SCR-B of Cvi-0, two SCR sequences of *A. lyrata*[7] and SCR-A-Col-0-restored, which is the full-length ORF reconstructed by inverting the 213-bp inversion and deleting the 14-bp duplication. The eight conserved cysteine residues are depicted in red; * denotes stop codons.

Figure 3 Interspecific crosses between *A. halleri* and *A. thaliana*. a, Incompatible and compatible reactions on the stigma. Crosses were carried out between *A. halleri* bearing haplogroup A and *A. thaliana* bearing an intact *SRK-A* of haplogroup A (Wei-1); between *A. halleri* bearing a non-A haplogroup and *A. thaliana* with intact *SRK-A* (Wei-1); and between *A. halleri* bearing haplogroup A and *A. thaliana* with degenerated *SRK-A* (Sha; see text). A bundle of pollen tubes indicate a compatible reaction (arrow), whereas no or few pollen tubes indicate incompatible reactions. Scale bar, 0.1 mm. **b**, Dependence of the incompatible phenotype on developmental stage in Old-1 and Wei-1. We confirmed that pistil length (treated as a continuous variable; Supplementary Note 2) was significantly influenced by developmental stage, that is,

time after the beginning of stage 13 ($P < 2.0 \times 10^{-16}$, by Generalized Linear Model (GLM); Supplementary Fig. 8). At this stage, flowers start to open and their stamens were removed experimentally. Red lines indicate the number of samples plotted at the same position (sunflower-plot). Binomial regressions based on GLM are shown as black lines (Old-1: $P = 0.0001$, Wei-1: $P = 0.001$). The light blue area indicates the mean length of pistils (± s.d.) when anthers would touch the stigma autonomously, which was observed using unmanipulated flowers with intact stamens (Old-1: $n = 6$; Wei-1: $n = 8$).

Figure 4 Restoration of functional *SCR-A* results in self-incompatibility and the prevention of selfing. a, Inhibited growth of pollen tubes on the stigma of a selfed pistil of Wei-1-SI-2. **b**, Growth of pollen tubes in the stigma of a pistil of Wei-1-SI-2 pollinated with pollen from Wei-1 (wild-type, non-functional *SCR-A*). An arrow indicates a bundle of pollen tubes. **a, b,** Scale bar, 0.1 mm. **c**, Inflorescences of the transgenic Wei-1-SI-2 plant and a wild-type Wei-1 plant (WT). Scale bar, 1 cm. **d**, Comparison of number of seeds per silique resulting from selfing of the transgenic Wei-1-SI-2 plant and a wild-type Wei-1 plant. The number of seeds was significantly reduced in the transgenic plant (Mann–Whitney *U* test, $n_{\text{wild-type}} = 30$, $n_{\text{Wei-1-SI-2}} = 30$, $P = 3.57 \times 10^{-11}$). **e**, Comparison of silique length resulting from selfing of the transgenic Wei-1-SI-2 plant and a wild-type Wei-1 plant. Silique length was significantly reduced in the transgenic plant (Mann–Whitney *U* test, $n_{\text{wild-type}} = 30$, $n_{\text{Wei-1-SI-2}} = 30$, $P = 1.91 \times 10^{-11}$). **d**, **e**, Bars represent the median, boxes the interquartile range and whiskers extend out to 1.5 times the interquartile range.

METHODS

Plant material

A. thaliana ecotypes listed in Supplementary Table 3 were obtained from the *Arabidopsis* Biological Resource Center (ABRC; www.biosci.ohio-state.edu/pcmb/Facilities/abrc/abrchome.htm). *A. halleri* was collected at Inagawa, Japan (34.9°N, 135.4°E), Sieber, Germany (51.7°N, 10.3°E) with the help of Maria Clauss, and at Giubiasco, Switzerland (46.2°N, 9.0°E). Plants used in the pollination assay were grown at 22 °C under a 16 h light/8 h dark cycle. *A. halleri* with haplogroup A (named W302) used for genetic crosses were generated by five generations of bud self-pollination to obtain homozygotes at the *S*-locus and to reduce heterogeneity between pollen grains, as often described in *Brassica* studies[31]. W302 plants used for the interspecific crosses were self-incompatible (n = 5) and cross-pollinations with other *A. halleri* plants bearing a haplogroup different from A were successful (n = 5), indicating that pollen of W302 plants was viable. Plants directly taken from wild populations were also used in the pollination assay, but exclusively as pollen donors bearing a haplogroup different from A.

Pollination assay and its statistical analysis

The interspecific crosses between *A. halleri* (male) and *A. thaliana* (female) yielded viable F_1 plants, and no obvious prezygotic isolation in the form of inhibited pollen tube penetrance was observed[32]. Developing floral buds were classified into stages as described in ref. 33. At the beginning of stage 13, anthers were removed from flower

buds and examined for contaminating pollen by light microscopy. At 0, 12, 24 and 36 h later, pollen grains were manually applied to stigmas. Pollination assays were conducted on intact stigmas. Two to three hours after pollination, flowers were fixed in a 9:1 mixture of ethanol and acetic acid, softened for 10 min in 1 M NaOH at 60 °C, stained with aniline blue in a 2% solution of K_3PO_4, and mounted on slides for examining pollen tubes using epifluorescence microscopy. In compatible crosses, more than 100 pollen tubes typically penetrate the stigma[8,31], and <20 pollen tubes were considered as incompatible crosses. We also employed more stringent criteria of <5 and confirmed significant correlations in the same four accessions (see Supplementary Note 2 for details). Because we confirmed that pistil length was significantly influenced by developmental stages (that is, time after the beginning of stage 13; $P < 2.0 \times 10^{-16}$; Supplementary Fig. 8; the Generalized Linear Model (GLM) and Poisson error structure were used), pistil length was measured as an indicator of flowers' developmental stage. GLM was used to assess the significance of the dependence of the incompatible phenotype on developmental stage. We assumed the SI phenotype (compatible = 1 and incompatible = 0) as a dependent variable and pistil length as an independent variable. Binomial error structures and logit link functions were used. By conducting χ^2 tests, likelihoods of the models incorporating pistil length were compared with null models that did not assume any independent variables. All statistical analyses were performed using R 2.8.1 (www.r-project.org).

Isolation of genomic/complementary DNA, genotyping and sequencing

Genomic DNA was isolated from young leaves of plants using Plant DNeasy Mini kits (Qiagen). Total RNA was extracted from floral buds and flower tissues of Ca-0, Co, Ge-1, Wei-1, Di-1, Ws-0, Uk-3, Gie-0, Old-1 and Fi-1. RNeasy kits (Qiagen) were used for extracting total RNA. Complementary DNA (cDNA) was synthesized using RETROscript reverse transcription kits (Ambion). Polymerase chain reactions (PCRs) were performed with Taq DNA polymerase (Roche), Go-Taq polymerase (Promega) or ExTaq (TaKaRa). The number of cycles of reverse transcription (RT)–PCR was 35. When multiple bands were observed, DNA fragments were purified with the GenElute Gel Extraction kit (Sigma-Aldrich). Primers used for amplification and genotyping are shown in Supplementary Table 4. *SCR-A* and *SRK-A* sequences of *A. halleri* subsp. *gemmifera* were obtained using the BD GenomeWalker Universal Kit (BD Biosciences). Direct DNA sequencing was conducted at the Institute of Plant Biology, University of Zurich, using a PRISM 3730 48-capillary automated sequencer (Applied Biosystems) and at the North Carolina State University Genome Research Laboratory with a Prism 3700 96-capillary automated sequencer (Applied Biosystems). All singleton polymorphisms were confirmed visually using BioLign (www2.maizegenetics.net/bioinformatics) and BioEdit (www.mbio.ncsu.edu/BioEdit/BioEdit.html)[34] and any ambiguous polymorphisms were rechecked with PCR re-amplification and sequencing. Nucleotide diversity and divergence were calculated using DnaSP 5.0 (ref. 35). Haplogroups E and W at *ARK3* and *PUB8* are highly divergent and thus inter-haplogroup recombinants are clearly

distinguishable from variation within haplogroups. We detected several low-frequency recombinants in both genes. For assigning genotypes at *ARK3* and *PUB8*, we thus used a partial region of each gene as indicated in Supplementary Fig. 9 (see Supplementary Note 1 for details).

Genotyping and sequence checking for pollination assays

The *S*-locus haplogroup of *A. halleri* (W302) used for the pollination assay was confirmed as haplogroup A by amplification and sequencing of the partial region of its *SRK* with haplogroup A-specific primers (PseSRK1092F1E1 and PseSRK1092R1E1) and with general primers for *SRK* (Aly13F1 and SLGR), as well as amplification and sequencing of the partial region of its *SCR* with haplogroup A-specific primers (SCR3 and SCR5). In addition, we confirmed that W302 was homozygous at the *S*-locus, as amplification with haplogroup A-specific primers (PseSRK1092F1E1 and PseSRK1092R1E1) was successful in eight out of eight F_1 hybrids between W302 and another *A. halleri* accession. Four individuals of *A. halleri* were used as controls ('non-haplogroup A'), for which we confirmed the absence of haplogroup-A alleles by genotyping with haplogroup A-specific primers (PseSRK1092F1E1 and PseSRK1092R1E1).

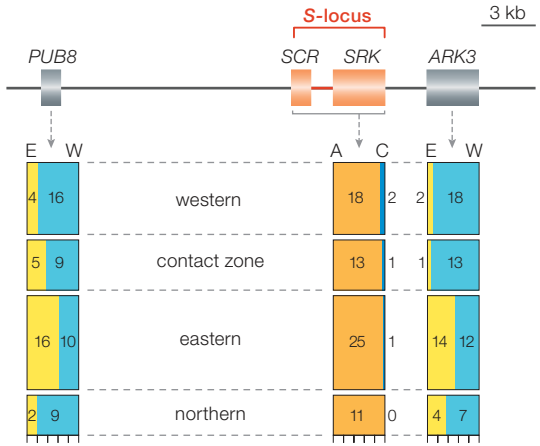
Transformation of restored *SCR*

The restored *SCR* gene was generated as follows: a 956-bp fragment including the first and part of the second exon was amplified by PCR using the primers SCR 5′F and SCR 5′R. The SCR 5′F primer contains a BamHI recognition site at the 5′ end for subcloning,

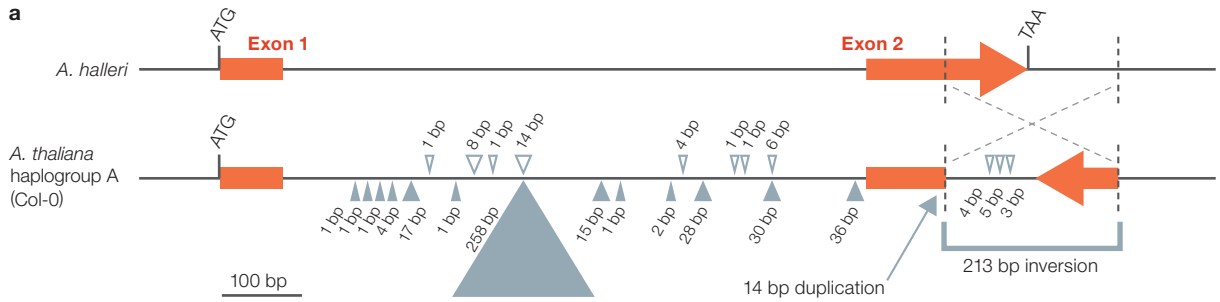
and the SCR 5′R primer includes an attachment sequence for the restored *SCR* ORF at the 5′ end. For amplifying the inverted region including the last half of the second exon, a 213-bp fragment was amplified by PCR using the primers SCR inv. F and SCR inv. R. The SCR inv. F primer contains a SacI recognition site at the 5′ end for subcloning. The resulting fragments were mixed and joined by PCR for the restored *SCR* ORF, using the primers SCR 5′F and SCR inv. F. The promoter region of the *ATA7* gene was amplified by PCR using the primers ATA7 pro 5′ and ATA7 pro 3′. The *ATA7* promoter and the restored *SCR* gene fragments were subcloned into the pBI121 vector and introduced into *Agrobacterium tumefaciens* strain EHA105 by electroporation. The *ATA7 promoter::GUS* construct was also introduced into EHA105 by the same procedure. The two constructs were transformed into *A. thaliana* plants (Wei-1) using the floral dip method[36]. Histochemical and microscopic GUS assays were carried out according to ref. 37.

31. Nou, I. S., Watanabe, M., Isogai, A. & Hinata, K. Comparison of *S*-alleles and *S*-glycoproteins between two wild populations of *Brassica campestris* in Turkey and Japan. *Sex. Plant Reprod.* **6,** 79–86 (1993).
32. Shimizu, K. K. Ecology meets molecular genetics in *Arabidopsis*. *Popul. Ecol.* **44,** 221–233 (2002).
33. Smyth, D. R., Bowman, J. L. & Meyerowitz, E. M. Early flower development in *Arabidopsis*. *Plant Cell* **2,** 755–767 (1990).
34. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symp. Ser.* **41,** 95–98 (1999).
35. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25,** 1451–1452 (2009).
36. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana. Plant J.* **16,** 735–743 (1998).

37. Park, J. I. *et al*. Molecular characterization of two anther-specific genes encoding putative RNA-binding proteins, AtRBP45s, in *Arabidopsis thaliana. Genes Genet. Syst.* **81,** 355–359 (2006).

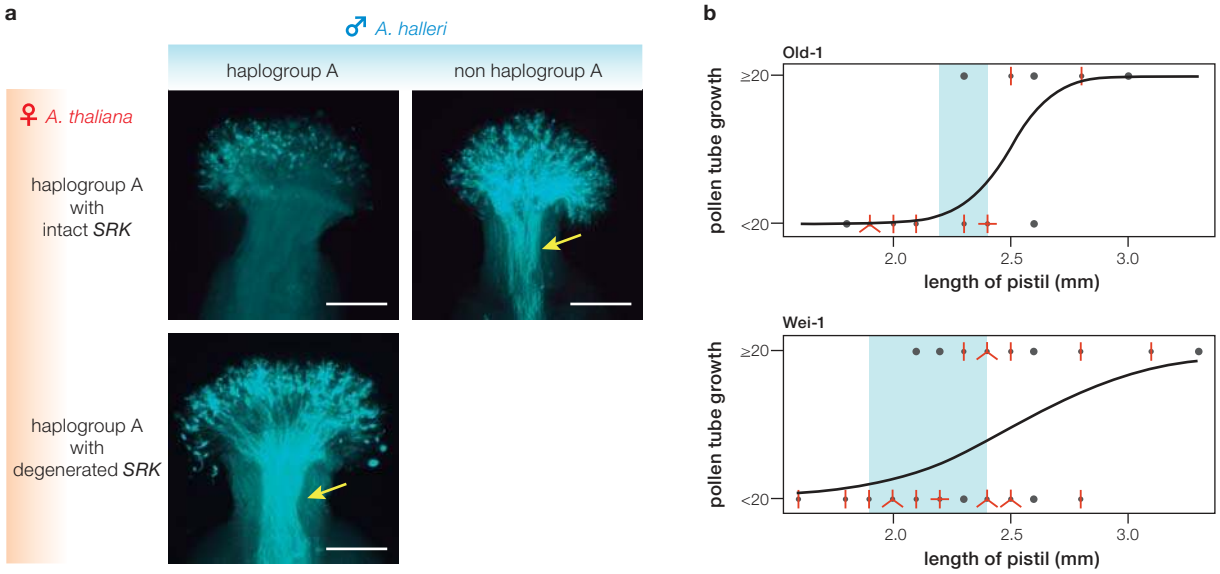


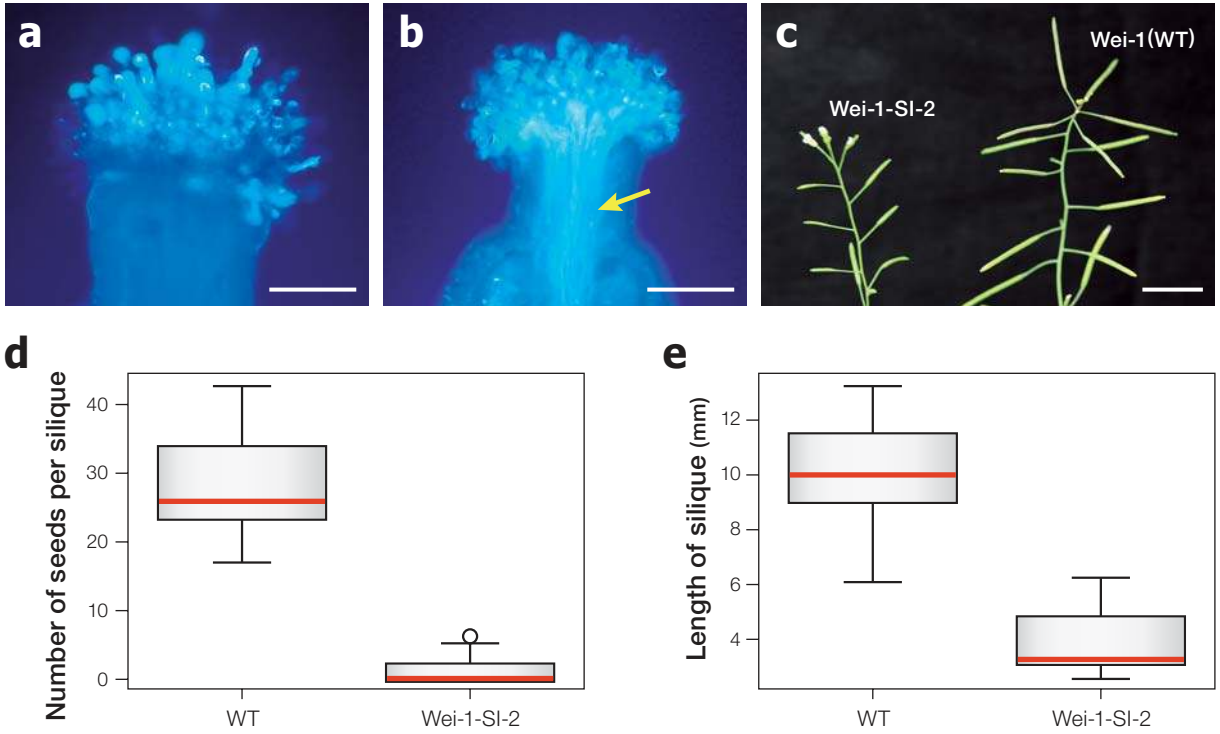
Frequency of haplogroups



b

```
                    10        20        30        40        50        60        70        80        90
                    .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
SCR-A-Ahalleri      MRCVVLFMVSCLLIVLLINHFEEVEAQKWKE-CNLRDIFP-GKCEHDANAKLRCKEDIAKNFRPSRPFECDCQTFDQG-RICYCKKCLV*
SCR-A-Col-0         MRCVVLFMVSCLLIVLLINHFEEVEAQKWNK-CFLRDIFP-GKCEHDANAKLRCKEDDAKKTLA*
SCR-A-Col-0-restored MRCVVLFMVSCLLIVLLINHFEEVEAQKWNK-CFLRDIFP-GKCEHDANAKLRCKEDIAKNFRPSRPFECNCQTFDKG-GICYCKKCLV*
SCR-B-Cvi-0         MKCAVSFMVSCFLIVFFTRHIKELEAQKWKA-CLIKQICP-GSCRTDGY--IRCKNDITKNGKH-RPFECKCKDVD-GDRLCFCYKCLVLRASSDLTT*
SCRa-Alyrata        MRCSVLFVVSYVIMSLLISHVQGMEDQKWKKVCNLEGNFP-GRCVGNGDEQ--CKRDLTEDGNN--PSKCRC-RFRAGRRHCRCIYCEVFGM*
SCRb-Alyrata        MRNATFFIVFYVFISLVLSNVQDVTAQK-NK-CMRSEMFPTGPCGNNGEET--CKKDFKNIYRT--PIQCKCLDKYDFARLCDCRFC*
```


a
♂ A. halleri
haplogroup A
non haplogroup A
♀ A. thaliana
haplogroup A with intact SRK
haplogroup A with degenerated SRK
b
Old-1
≥20
<20
pollen tube growth
2.0
2.5
3.0
length of pistil (mm)
Wei-1
≥20
<20
pollen tube growth
2.0
2.5
3.0
length of pistil (mm)


a
b
c
Wei-1(WT)
Wei-1-SI-2
d
Number of seeds per silique
40
30
20
10
0
WT
Wei-1-SI-2
e
Length of silique (mm)
12
10
8
6
4
WT
Wei-1-SI-2

Supplementary Note 1: Molecular population genetic analyses of the *S*-locus region

Genotypes of *ARK3*, *SRK*, and *PUB8* were surveyed in the same 76 accessions of *A. thaliana* that were used by François *et al*.[17], which are a subset of those examined by Nordborg *et al*.[38] (accession numbers: CS22564–CS22659). In addition to Francois *et al.*[17] and Sharbel *et al.*[16], other studies also support the pattern of population structure, such as the east–west gradient or clustering of accessions and the divergence of northern European *A. thaliana* accessions[38–40].

For the analysis of associations between genotypes and population structure (Fig. 1), three accessions, Bur-0, Fei-0, and Mr-0, were removed from our analyses because François *et al*.[17] found that they seem to have experienced recent complicated admixture events. Pro-0 and Ren-1 were also excluded because of signatures of presumable gene conversion of haplogroup E and haplogroup W or a parallel mutation in *ARK3* and *PUB8* of Pro-0, and because of heterogeneity in *ARK3* and *PUB8* of Ren-1. Among four individuals of Ren-1, we found one haplogroup E, two haplogroup W and one heterozygote at *PUB8*; similarly, we found two haplogroup W, one recombinant of E and W, and one heterozygote at *ARK3*. In other population genetic studies using the same sample set in *A. thaliana*, a low frequency of heterogeneous or heterozygous accessions has also been found and was eliminated from the analyses (e.g., Nordborg *et al.* 2005)[38].

Cramer's coefficient *V* (also called ϕ)[41,42] was calculated for each gene. Cramer's coefficient *V* measures the strength of association or interdependence between two categorical variables. Its statistical significance was assessed using Fisher's exact test. Genotyping data are summarised in Supplementary Table 1. High levels of sequence divergence between two haplogroups were observed in *PUB8*, *SRK* and *ARK3* (Supplementary Fig. 1 for *PUB8*, Supplementary Fig. 2 for *ARK3*, and see Bechsgaard *et al*.[23] for *SRK*).

(1) *ARK3*: The downstream part of the 5′–UTR, the entire first exon encoding the S-domain, and the upstream part of intron 1 were amplified and sequenced with the primer pair ARK3-1F and ARK3-4aR. Because we found several recombinants within *ARK3*, we used the partial region in exon 1 and the upstream part of intron 1 to assign the genotype of *ARK3* (11391631–11392010; described in Supplementary Fig. 9).

Gene duplication and recombination between distinct sequences of *ARK3* have been reported by Sherman-Broyles *et al*.[9]. In C24, the duplicated gene fragment (Δ*ARK3*) of *ARK3* is known to be present, and Δ*ARK3* lacks the region we sequenced[9]. Thus, this region is considered a single-copy fragment. All 76 accessions, except Ren-1 with possible heterozygosity described above, yielded amplification products with no evidence of heterozygosity or paralogous sequences, indicating that the region of *ARK3* that was amplified with these primers represents a single-copy region in our samples. The correlation between population structure

("clustering") and the proportions of haplogroups at *ARK3* (see main text) also suggests that possible effects of rearrangements and duplications at *ARK3* on haplogroup frequency of exon 1 have been minor.

(2) *PUB8* coding region: Most of the single exon of the *PUB8* gene was sequenced with two primers, 4g21350f(2) and 4g21350r(2). Because several recombinants were found within the *PUB8* gene, we used a partial region of the exon to assign the *PUB8* genotype (11356518–11357095; Supplementary Fig. 9).

(3) *PUB8* promoter: We used the data previously reported by Liu *et al.*[13]. Liu *et al.* reported polymorphism in the promoter region of *PUB8* that might be responsible for the cryptic natural variation in the strength of SI in *A. thaliana*[13]. Polymorphism was assigned as "C" and "R" for a worldwide collection of 85 accessions. We examined the correlation with population structure ("clustering") for the European accessions.

(4) *S*-locus: Genotyping primers were designed to amplify exon 1 (encoding the divergent S-domain of haplogroups A and C) of *SRK* based on the following reasons: (1) There has been no report of inter-haplogroup recombination between *SRK* and *SCR*, nor within *SRK* or *SCR*[10,11]. We note that a few genomic rearrangements or non-homologous recombination events reported at the *S*-locus rather involve tandem duplicates of haplogroups A and C in the genomic region of *SRK* and *ARK3* (in the European accession C24, a fragment of haplogroup A is flanked by a fragment of haplogroup C, and thus its *S*-locus has part of exon 1 of *SRK-A* (A-t2-type deletion), chimeric and partial *ARK3*, part of exon 7 of *SRK-C* and a full-length *ARK3*. The Asian accession Kas-2 has *SRK-C* and part of exon 1 of *SRK-A*[12]. There has been no report suggesting that these duplications have reached intermediate or high frequency). The tight linkage between *SRK* and *SCR* indicates that either one of them could be used to represent the *S*-locus genotype. (2) *SCR* of haplogroup C has not been found in previous studies, maybe due to deletion of *SCR*[11]. Thus *SCR* would not be ideal for genotyping, since haplogroup C should be represented as the lack of amplification. We used the following three primer pairs for *SRK-A* that were designed by Shimizu *et al*.[11] to encompass major known rearrangements: PseSRK1092F1E1 and PseSRK1092R1E1 designed within exon 1 of *SRK-A*, PseSRKFU1 and PseSRKR1E1 to amplify the promoter and exon 1 of *SRK-A*, LTRF1 and PseSRKnon6799R1E2 to amplify exon 1 of *SRK-A* to the LTR region including the A-t2-type deletion of *SCR-A*. Because Shimizu *et al*.[11] reported that the A-t2 deletion encompassing *SCR-A* and most of *SRK* was evolutionarily derived from haplogroup A that already possessed the *SCR-A* pseudogene (based on an analysis of the deletion border sequence), the At-2 type was regarded as haplogroup A. The frequency of the A-t2 deletion type was 30%. To amplify exon 1 of *SRK-C*, the primer pair PseSRK6751F1E1 and PseSRK6751R1E1 was used. We

note that these PCR-based genotyping methods give minimum estimates of the frequency of each sequence, because small duplicated fragments that would not be amplified with these primers may exist in some accessions as chimeric sequences. However, all 76 accessions showed amplification of either haplogroup A or C and no accession had both haplogroups, suggesting that the region of *SRK* exon 1 that is amplified with these primers represents a single-copy region in our samples. We also note that *SCR-A* was not found in accessions that possess exon 1 of *SRK-C* in our samples, consistent with the paucity of inter-haplogroup recombination that might otherwise yield novel combinations of haplogroups at *SRK* and *SCR*. In this paper, we genotyped additional accessions that were not genotyped by Shimizu *et al.*[11].

One of the primers used to amplify the *SCR-A* pseudogene in Shimizu *et al*.[11] turned out to be located within the 213-bp inversion mutation, and thus successful amplification with the primer pair described by Shimizu *et al*.[11] was interpreted to indicate the presence of the 213-bp inversion. The 14-bp duplication was found in several accessions including Col-0 but not in others (Nok-0 and Pog-0), and the 5-bp deletion was found in Nok-0 and Pog-0 but not in other surveyed accessions[11], which strongly suggests that these mutations occurred after the 213-bp inversion mutation and did not affect the self-incompatibility phenotype.

As inter-haplogroup divergence is high both in *SRK* (A and C) and in *ARK3* (E and W), we confirmed their linkage by genetic experiments. We obtained 128 lines of Col *gl1-1* × Kas-1 recombinant inbred lines from the Arabidopsis Biological Resource Center (stock numbers CS84871–84998). These lines were constructed by repeated selfing of the F_2 population and thus most of the genome became homozygous. One of the parental accessions (Col) had *SRK-A* and *ARK3-E*, and the other parental accession (Kas-1) had *SRK-C* and *ARK3-W*. *SRK* alleles were genotyped using haplogroup-specific primer pairs (haplogroup A: PseSRK1092F1E1 and PseSRK1092R1E1, haplogroup C: PseSRK6751F1E1 and PseSRK6751R1E1). *ARK3* was amplified using primers ARKIII3 and ARKIII5, and allelic differences were assayed by digestion with DdeI. Among these 128 lines, heterozygosity was found in only three lines. In the remaining 125 lines, no recombinants between *ARK3* and *SRK* were found. These data strongly indicate that *ARK3* and *SRK* are physically linked both in haplogroup A and haplogroup C, and that despite high sequence divergence a major relocation in the genome did not occur.

Putative major disruptive mutation events for haplogroup A at the *S*-locus are summarised in Supplementary Fig. 10.

Supplementary Note 2: Interspecific crossing experiments with *A. halleri*

To examine the incompatibility phenotype due to the SI system from the interspecific crossing

experiments, not only crossings between *A. halleri* bearing haplogroup A and *A. thaliana* bearing an intact *SRK-A* of haplogroup A are necessary, but two other types of interspecific crosses need to be performed: (1) crosses between *A. halleri* bearing a haplogroup different from haplogroup A and *A. thaliana* bearing an intact *SRK-A* of haplogroup A, and (2) crosses between *A. halleri* bearing haplogroup A and *A. thaliana* bearing pseudogenized *SRK-A* of haplogroup A. The first type of cross was conducted to exclude the possibility that intrinsic reproductive barriers between *A. thaliana* and *A. halleri* prevent elongation of pollen tubes and/or fertilisation. The second type of cross aimed at establishing that the *A. halleri* plant bearing haplogroup A (W302) produces viable pollen with normal pollen tube elongation, while it also functions to test for the absence of intrinsic interspecific reproductive barriers.

We found incompatible phenotypes in seven accessions. In four of these, a constitutive incompatible reaction was observed up to the flower developmental stage when autopollination occurs, and pollen grains were significantly inclined to be accepted in later stages. In the other three accessions, weak incompatible phenotypes were found; the number of growing pollen tubes was highly variable and no clear dependency on floral developmental stage was found (Fig. 3; Supplementary Fig. 6; Supplementary Table 2). In addition, four accessions invariably accepted pollen grains of any *S*-haplogroup (Supplementary Table 2), suggesting that one or more female SI components are degenerated.

While we considered <20 elongated pollen tubes as a threshold for an incompatible reaction (see Methods), we confirmed that the tendency we found was robust to more stringent criteria: we analysed the data using a threshold of <5 pollen tubes, and found significant dependencies of incompatibility on development in Wei-1 ($p = 0.001$), Di-1 ($p = 0.046$), Co ($p = 0.0007$) and Old-1 ($p = 0.0001$).

We used pistil length as a continuous variable to measure developmental stage because pistil length was significantly influenced by developmental stage, i.e., time after the beginning of stage 13, when flowers start to open (Supplementary Fig. 8). We found that the incompatible phenotype was also significantly influenced when time after the beginning of stage 13 was used directly to measure developmental stage (four discrete values: 0, 12, 24 and 36 hours; Old-1: $p = 0.0093$, Wei-1: $p = 0.001$).

Supplementary Note 3: Pseudo-self-compatibility (PSC)

The loss of incompatibility in older flowers found in our experiments might correspond to pseudo-self-compatibility (PSC), which is a common phenomenon in the wild[43,44] and in cultivated populations[45]. Nasrallah *et al.*[46] found cryptic natural variation in the strength of SI by transforming different *A. thaliana* accessions with *SCR* and *SRK* genes from

self-incompatible *A. lyrata*. Liu *et al*.[13] considered these phenomena as PSC and reported that the promoter region of the *PUB8* gene is responsible for this variation in *A. thaliana*. This raises the possibility that polymorphism at *PUB8* might also be responsible for the phenotypic variation in our interspecific crossing experiment (Supplementary Table 2). To test this possibility, we sequenced the promoter region of *PUB8* of the accessions with full-length *SRK*. We found no clear correlation between the genotype of the promoter region of *PUB8* and the incompatible phenotype (Supplementary Table 5), suggesting that other loci are responsible for variation in the strength of the incompatibility reaction. The loss of *SRK* expression is not likely to explain the variation either, because amplification and sequencing of *SRK* cDNA was successful in all of those accessions. Boggs *et al.*[12] recently reported several quantitative trait loci (QTLs) responsible for SC in *A. thaliana*. Phenotypic variation in our interspecific experiments might be explained by polymorphisms at these loci. However, we note that the results of our experiment and those of Liu *et al.*[13] are not directly comparable because our experiment involved interspecific crosses while the results reported by Liu *et al*.[13] were based on transgenic experiments with heterologous *SCR-SRK*.

Supplementary Note 4: Transformation of restored *SCR-A*

We introduced *ATA7::restored SCR* into the Wei-1 accession. Among 35 independent transformants, we confirmed the expression of a restored *SCR* sequence by reverse transcription (RT)–PCR in two lines (Wei-1-SI-2 and Wei-1-SI-10), which were selected for the following SI-phenotype and seed-set monitoring assays. As a positive control, we confirmed that pollen grains of Col-0 were invariably accepted by transformants (Supplementary Fig. 11). Segregation ratios in generations T_2 and T_3 did not follow a simple one-locus biallelic Mendelian system: SI:SC = 1:5 in the T_2 generation and SI:SC = 1:11 in the T_3 generation. Phenotypic instability and low heritability arising from epigenetic silencing are also suggested in the mapping population of *AlSRKb-SCRb* transformants of the C24 and Col-0 accessions[12].

Additional References in Supplementary Information

38. Nordborg, M. *et al*. The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol*. **3,** e196 (2005).
39. Schmid, K. J. *et al*. Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *Theor. Appl. Genet.* **112,** 1104–1114 (2006).
40. Beck, J. B., Schmuths, H. & Schaal, B. A. Native range genetic variation in *Arabidopsis*

thaliana is strongly geographically structured and reflects Pleistocene glacial dynamics. *Mol. Ecol.* **17,** 902–915 (2008).

41. Cramer, H. *Mathematical Method of Statistics* (Princeton University Press, Princeton, 1946).
42. Sokal, R. R. & Rohlf, F. J. *Biometry: the Principles and Practice of Statistics in Biological Research. 3rd ed.* (W. H. Freeman, New York, 1995).
43. Levin, D. A. The evolutionary significance of pseudo-self-fertility. *Amer. Natur.* **148,** 321–332 (1996).
44. Good-Avila, S.V. & Stephenson, A. G. The inheritance of modifiers conferring self-fertility in the partially self-incompatible perennial *Campanula rapunculoides* L. (Campanulaceae). *Evolution* **56,** 263–272 (2002).
45. Tatebe, T. Studies on the behavior of incompatible pollen in the Japanese radish. *J. Japan. Soc. Hort. Sci.* **11,** 207–234 (1942).
46. Nasrallah, M. E., Liu, P., Sherman-Broyles, S., Boggs, N. A. and Nasrallah, J. B. Natural variation in expression of self-incompatibility in *Arabidopsis thaliana*: implications for the evolution of selfing. *Proc. Natl Acad. Sci. USA* **101,** 16070–16074 (2004).
47. Charlesworth, D., Awadalla, P., Mable, B. K. & Schierup, M. H. Population-level studies of multiallelic self-incompatibility loci, with particular reference to Brassicaceae. *Ann. Bot.* **85,** 227–239 (2000).
48. Korves, T. M. *et al.* Fitness effects associated with the major flowering time gene *FRIGIDA* in *Arabidopsis thaliana* in the field. *Amer. Natur.* **169,** E141–E157 (2007).

Supplementary Table 1. Genotyping data of *ARK3*, *SRK*, *PUB8,* and *SCR-A* with population structure ("clustering") reported by François *et al*.[17]

Accession	clustering	*PUB8*	*S*-locus	*ARK3*	*SCR-A*§
Ag-0	Western	W	A	W	no
An-1	Western	W	A	W	no
C24	Western	W	A	W	no
CIBC-17	Western	W	A	W	no
CIBC-5	Western	E	A	E	yes
Edi-0	Western	E	A	W	yes
Gy-0	Western	W	A	W	no
HR-10	Western	W	A	W	no
HR-5	Western	E	A	E	yes
LL-0	Western	W	A	W	no
Lz-0	Western	W	C	W	no
NFA-10	Western	W	A	W	no
NFA-8	Western	W	A	W	no
Pro-0	Western	NA†	C	NA†	no
Ra-0	Western	W	C	W	no
Ren-1	Western	NA†	A	NA†	yes
Ren-11	Western	W	A	W	yes
Se-0	Western	W	A	W	no
Sq-1	Western	W	A	W	no
Sq-8	Western	E	A	W	yes
Ts-1	Western	W	A	W	no
Ts-5	Western	W	A	W	no
Bay-0	Contact zone	E	A	W	no
Ct-1	Contact zone	W	A	W	no
Ei-0 (Ga-0)‡	Contact zone	E	A	W	no
Ei-2	Contact zone	E	A	E	yes
Got-22	Contact zone	W	A	W	no
Got-7	Contact zone	W	A	W	no
Gu-0	Contact zone	W	A	W	yes
Mrk-0	Contact zone	E	A	W	yes
Mt-0	Contact zone	W	A	W	no
Mz-0	Contact zone	W	A	W	yes
Nd-1	Contact zone	E	A	W	yes
Nok-3	Contact zone	W	A	W	yes
Wei-0	Contact zone	W	A	W	no
Wt-5	Contact zone	W	C	W	no
Bur-0	NA*	W	C	W	no
Fei-0	NA*	W	A	W	no
Mr-0	NA*	W	C	W	no
Bor-1	Eastern	E	A	W	yes
Bor-4	Eastern	E	A	E	yes
Br-0	Eastern	W	C	W	no
CS22491	Eastern	W	A	E	yes
Est-1	Eastern	E	A	E	yes
Ler-1	Eastern	E	A	E	yes
Lp2-2	Eastern	E	A	E	yes
Lp2-6	Eastern	E	A	E	yes
Ms-0	Eastern	W	A	E	yes
Omo2-1	Eastern	E	A	W	yes
Omo2-3	Eastern	E	A	W	yes
Pu2-23	Eastern	E	A	W	yes
Pu2-7	Eastern	E	A	W	yes
Spr1-2	Eastern	W	A	E	yes
Spr1-6	Eastern	W	A	E	yes
Ull2-3	Eastern	W	A	W	no
Ull2-5	Eastern	W	A	E	yes
Uod-1	Eastern	E	A	W	yes
Uod-7	Eastern	W	A	W	no
Var2-1	Eastern	W	A	E	yes
Var2-6	Eastern	W	A	E	yes
Wa-1	Eastern	E	A	W	yes
Ws-0	Eastern	E	A	W	yes
Ws-2	Eastern	E	A	E	yes
Zdr-1	Eastern	E	A	W	yes
Zdr-6	Eastern	E	A	E	yes
Bil-5	Northern	W	A	W	no
Bil-7	Northern	W	A	W	no
Eden-1	Northern	W	A	W	no
Eden-2	Northern	W	A	E	yes
Fab-2	Northern	W	A	E	yes
Fab-4	Northern	W	A	W	no
Lov-1	Northern	W	A	W	yes
Lov-5	Northern	W	A	W	yes
Oy-0	Northern	W	A	W	yes
Tamm-2	Northern	E	A	E	yes
Tamm-27	Northern	E	A	E	yes

§ Amplification of *SCR-A* by PCR.

* Not clearly assigned to any clusters[17].

† heterozygote or duplicate (*PUB8* and *ARK3* in Ren-1) and presumable gene conversion or a parallel mutation (*PUB8* and *ARK3* in Pro-0). See Supplementary Note 1 for details.

‡ Ga-0 and Ei-0 are synonymous (O. François, personal communication).

Supplementary Table 2. Summary of results of interspecific crosses

			Pollen donor (*A. halleri*)			
			haplogroup A		Non haplogroup A	
			<20 (incompatible)	>=20 (compatible)	<20 (incompatible)	>=20 (compatible)
Pistil donor (*A. thaliana*)	SRK pseudogenized accessions	Bu-21	0	5		
		Co-3	0	21		
		Chi-1	0	11		
		Col-0	1	6		
		Ler-0	0	2		
		Sha	0	13 (15)*		
		Da-0	1	13		
	SRK intact accessions	Co‡	16 (18)*	2 (8)*†	0	6
		Wei-1‡	25 (29)*	9 (17)*†	0	6
		Old-1‡	18	2 (7)*†	0	2
		Di-1‡	23 (24)*	12 (18)*†	0	4
		Uk-3‡	12	5	0	4
		Pog-0‡	21 (24)*	23 (32)*	0	6
		Ca-0‡	11	8 (13)*	0	4
		Fi-1	1	11		
		Ge-1	0	8		
		Gie-0	0	7		
		Nok-0	1	3 (4)*		
		Ws-0	0	15	0	1

*Numbers outside parentheses denote the results of crosses using relatively young flowers (0, 12, and 24 h after the beginning of stage 13). Numbers in parentheses include the results for very old flowers (36 h after the beginning of stage 13).

‡The seven accessions in which incompatible reactions were found.

†The four accessions in which pollen grains were significantly inclined to be accepted in later stages.

Supplementary Table 3. Accessions used as female parents for interspecific crosses, with additional information based on the Arabidopsis Biological Resource Center

Number	Accession Name	Country	Intact *SRK*	*SCR-A*‡
CS1602	Ws-0	Belarus‡‡	yes	no §
CS3180	Co	Portugal	yes	yes
CS6658	Ca-0	Germany	yes	yes
CS6681	Di-1	France	yes	yes
CS6705	Fi-1	Germany	yes	yes
CS6718	Ge-1	Switzerland	yes	yes
CS6720	Gie-0	Germany	yes	yes
CS6820	Old-1	Germany	yes	yes
CS1476/CS6842	Pog-0*	Canada	yes	yes
CS6880	Uk-3	Germany	yes	yes
CS6925	Wei-1	Switzerland	yes	yes
CS6807	Nok-0	Netherlands	yes	yes
CS1044/CS6652	Bu-21	Germany	no	yes
CS6671	Co-3	Portugal	no†	yes
CS6665	Chi-1	Russia	no	yes
CS1092	Col-0	USA	no	yes
CS20	Ler-0	Poland	no	yes
CS6180	Sha	Tajikistan	no	yes
CS6676	Da-0	Germany	no†	yes

‡ Amplification of *SCR-A* by PCR.

‡‡ While the TAIR website (http://arabidopsis.org) lists Ws-0 as being from Russia, its longitude and latitude information indicate it is from Belarus.

§ *SCR-A* was amplified when a synonymous accession (Ws-0, CS22623; see Supplementary Table 1) was used.

* Note that Pog-0 is not from Europe but from Canada. However, the North American population is considered a subpopulation of the European population arising from recent invasion[38].

† Co-3 and Da-0 were reported as A-t4 type (intact *SRK* type) by Shimizu *et al.*[11] based on genotyping screening, but subsequent sequencing results revealed disruptive mutations that could not be detected previously by genotyping PCR.

Supplementary Table 4. List of primers, with sequences shown from 5′ to 3′

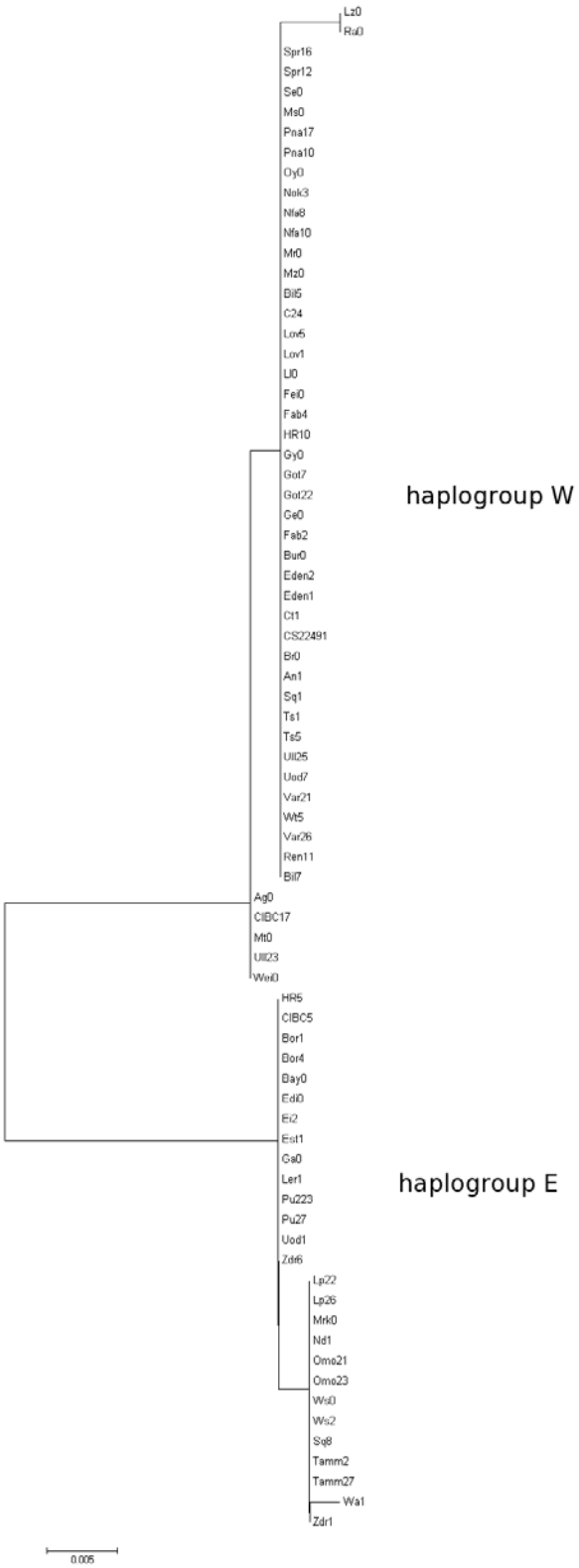
	Primer name	Primer sequence	Reference
SRK coding region sequencing	PseSRKFU1	CGTGGAATAACACTACTCTTTGTCGTG	11
	PseSRKR3E2	CAAGCAAGCGTCTCCCACTTGAC	this study
	SRK3	ACGCTAATGTCTTCTTTGCATCG	this study
	SRKAintactRD1	TTAGCCATTGAGGTTTACTTCTTTCA	11
S-locus genotyping	PseSRK1092F1E1	GATTGACAATAAATACCGTAGGACG	11
	PseSRK1092R1E1	CCTTCGACATCTACCTGTCACG	11
	Aly13F1	CCGACGGTAACCTTGTCATCCTC	47
	SLGR	ATCTGACATAAAGATCTTGACC	47
	SCR3	ACGCTAATGTCTTCTTTGCATCG	11
	SCR5	GGTTTCTTGTCTTCTCATAGTTCTCC	11
	PseSRK6751F1E1	TGACTGTGAGTTTCTCTGGGTC	11
	PseSRK6751R1E1	CCTCACACACCCACTTGAACCA	11
	PseSRKFU1	see above	
	PseSRKR1E1	GATCCCGACCCAATTTCATCTG	11
	LTRF1	ATTCGGCAATTCACTAACAACAAG	11
	PseSRKnon6799R1E2	AAAGTGTAAGCGACCTCCTCTCTATTC	11
ARK3 sequencing	ARK3-1F	CTTGATTGGGCAGTTTTAGC	10
	ARK3-4aR	CCAAATGCGAACAAGAAGACC	10
PUB8 exon sequencing	4g21350f(2)	GTTATCCTCCAATCGGGTCA	11
	4g21350r(2)	GTCCCAACACTTCCACTGCT	11
PUB8 genotyping	Ubox_C_F	TGCAGGAACAGAGCAGAGAGAAC	13
	Ubox_C_R	CATTGCTGCAAAAAGGTTCAC	13
cDNA sequencing of SRK	SRKAt-4RTF5	TCATGGGTTAGATAAACCTGAAA	this study
	SRKR3E2	see above	
	SRK3	see above	
	SRKRTR2	AATAACTGGGTCGGTAGCTTTTACA	this study
transgenic experiment	SCR5'F	GGATCCATGAGATGTGTTGTTTTGTTTATGGTTTC	this study
	SCR5'R	GAGAGGGTCTGAAATTCTTAGCAATGTCTTCTTTG CATCGTAGTTTTGCG	this study
	SCR inv. F	GAGCTCATTAGCGTAAAATACAGCTTTAAAATGAAT ATTTTG	this study
	SCR inv. R	ATTGCTAAGAATTTCAGACCCTCTCGCCC	this study
	ATA7 pro 5’	ATAGTCTTCTTGTACACGTCGACCA	this study
	ATA7 pro 3’	GGCTTAGTTTAATGAACACATGTAGTTC	this study
mapping of SRK/ARK3	ARKIII3	GAGTATTGGAGTGAGCGTTTTGC	this study
	ARKIII5	CCAACCACCTTCCTCGTATTAGC	this study

Supplementary Table 5. Relationship between the genotype of the promoter region of *PUB8* and phenotype of pollen tube behaviour in interspecific crosses

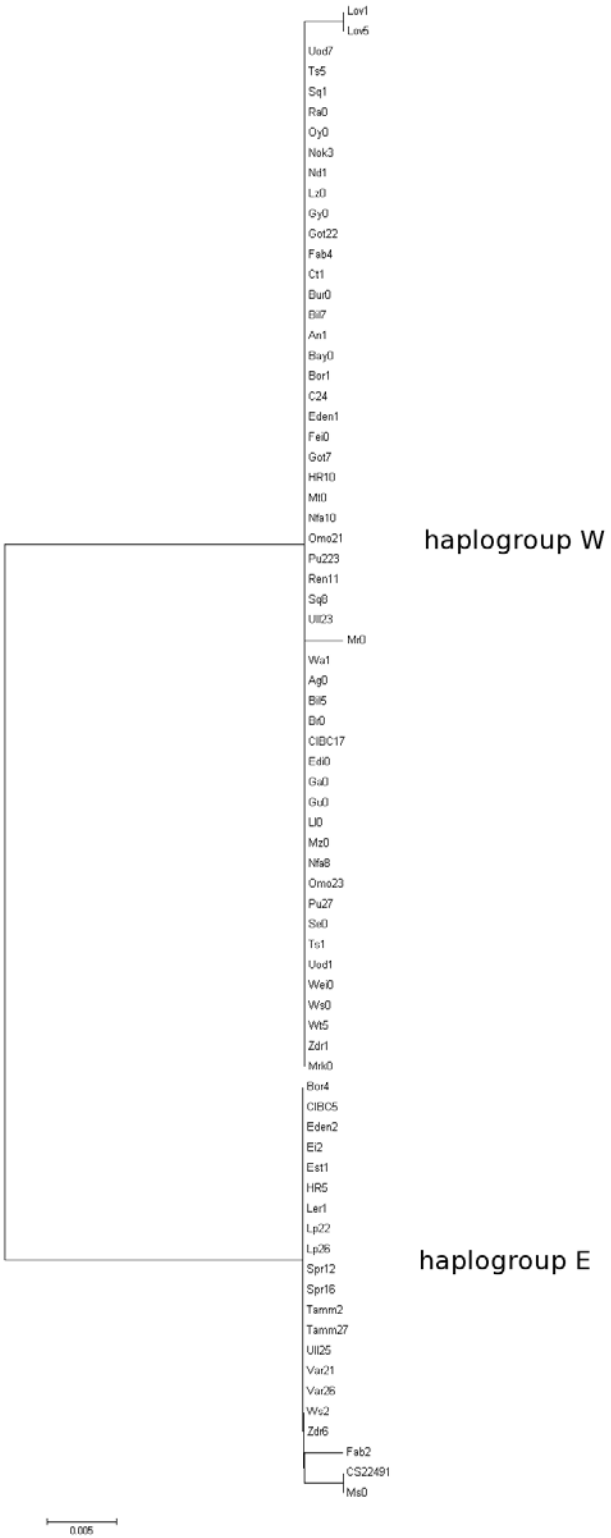
Accession		PUB8 promoter genotype*	Phenotypes found in inter-specific cross
CS1602	Ws-0	(GA)x8	Compatible
CS3180	Co	(GA)x8	Incompatible
CS6658	Ca-0	(GA)x10	Unstably incompatible
CS6681	Di-1	(GA)x9	Incompatible
CS6705	Fi-1	(GA)x8	Compatible
CS6718	Ge-1	(GA)x8	Compatible
CS6720	Gie-0	(GA)x8	Compatible
CS6807	Nok-0	(GA)x8	Compatible
CS6820	Old-1	(GA)x9	Incompatible
CS1476/CS6842	Pog-0	(GA)x8	Unstably incompatible
CS6880	Uk-3	(GA)x9	Unstably incompatible
CS6925	Wei-1	(GA)x10	Incompatible

*The polymorphism of the number of GA-repeats in the simple-sequence-repeat (SSR) region. Although Liu *et al*.[13] reported only two alleles of the *PUB8* promoter region based on electrophoresis (named "C" and "R"), our sequencing results using accessions with full-length *SRK* identified more alleles in this SSR region.

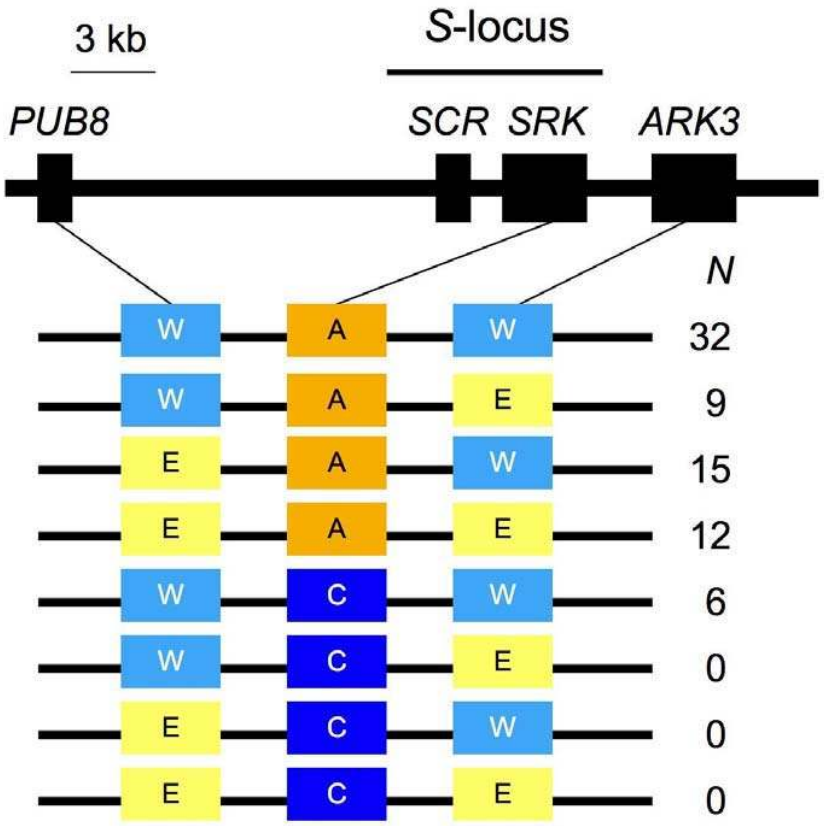
*Based on the information for the BAC sequence, the genotype of *PUB8* in C24 was (GA)x11 (ref. 9).



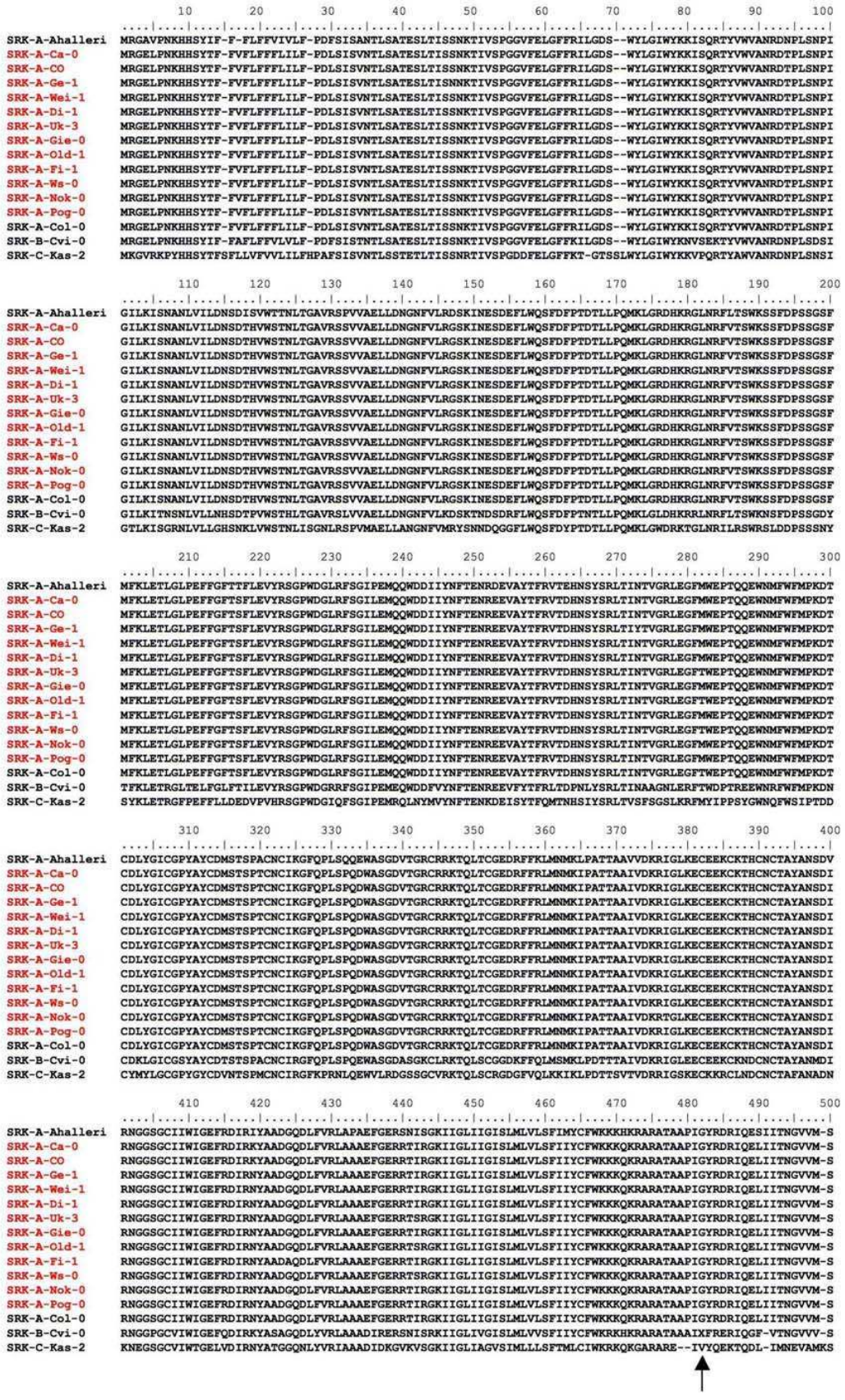
Supplementary Figure 1. Neighbour-joining tree of the partial region of *PUB8* used for genotyping (see Supplementary Note 1). Nucleotide distances were estimated using the Kimura two-parameter model.

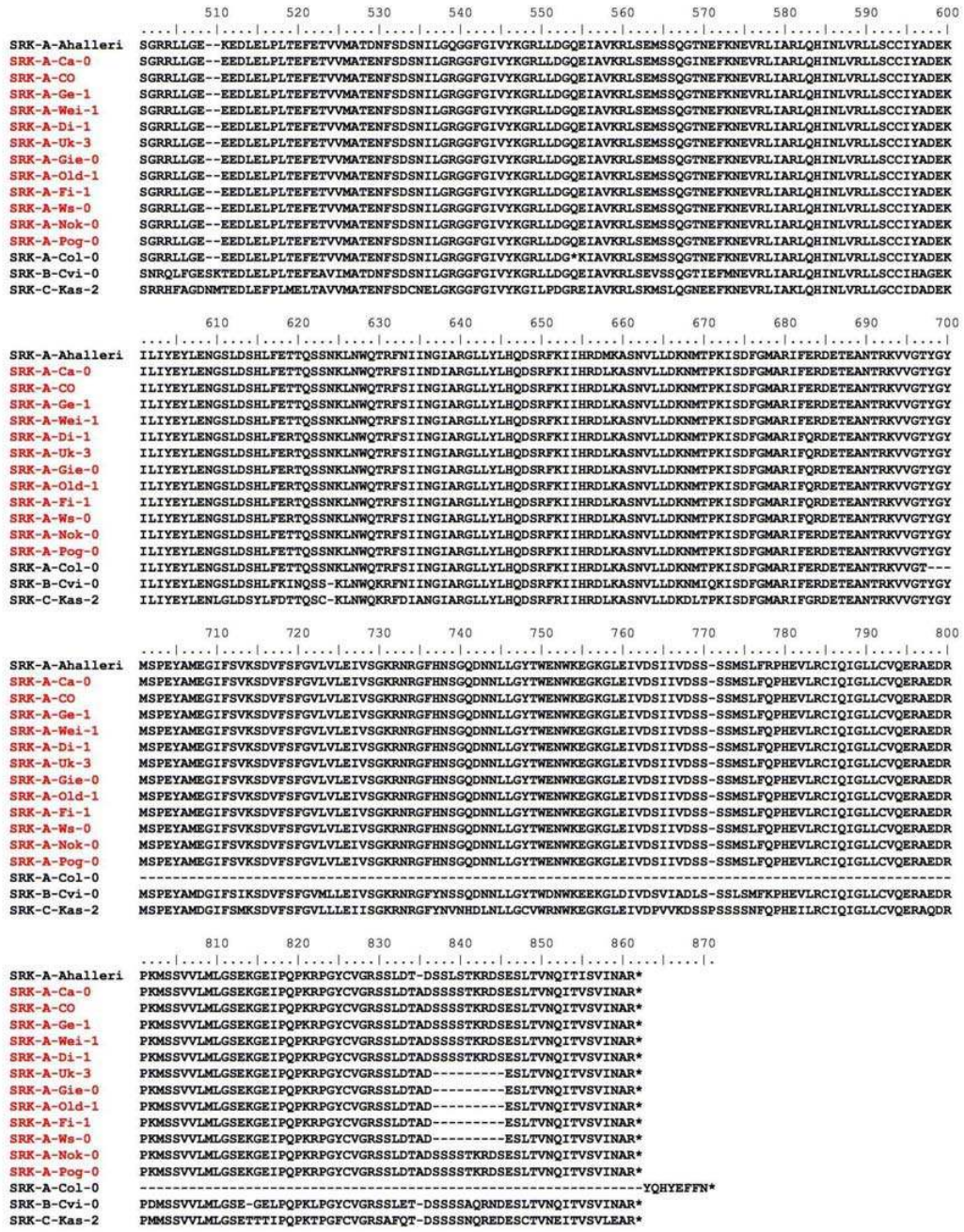


Supplementary Figure 2. Neighbour-joining tree of the partial region of *ARK3* used for genotyping (see Supplementary Note 1). Nucleotide distances were estimated using the Kimura two-parameter model.

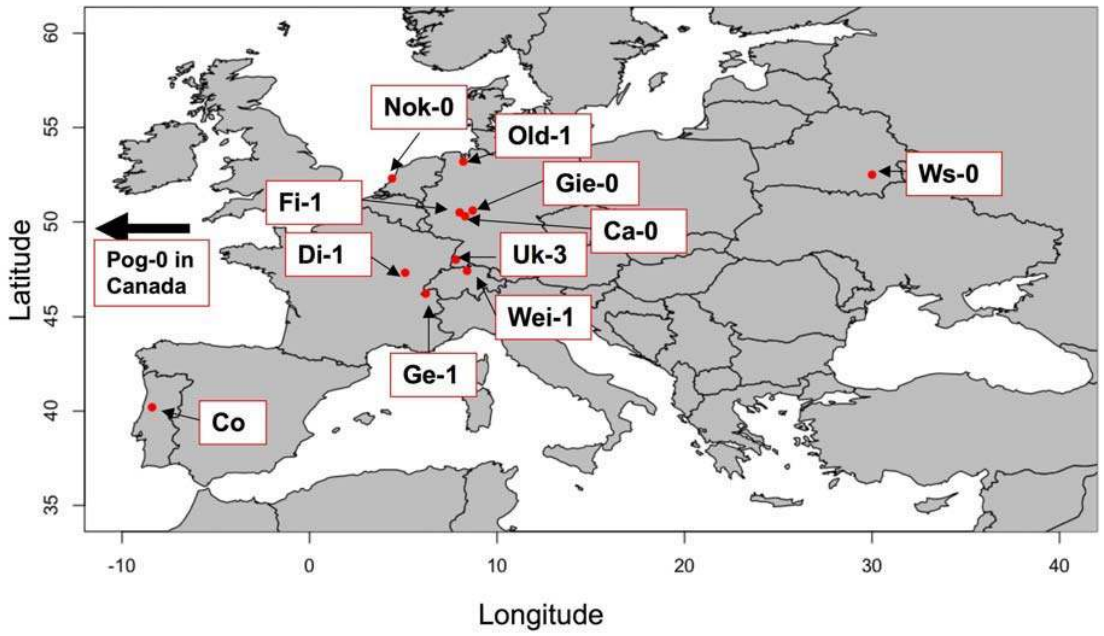


Supplementary Figure 3. Frequencies of haplotype patterns at the *S*-locus and flanking genes. Two distinct haplogroups at the *S*-locus (A and C) were found in European *A. thaliana*, and two distinct haplogroups (E and W) were also found at *ARK3* and *PUB8*, which flank the *S*-locus. The frequency of each of these eight multilocus haplotypes is not significantly different from the expected frequency given the haplogroup frequencies at each of the three genes (Fisher's exact test, $p = 0.469$).

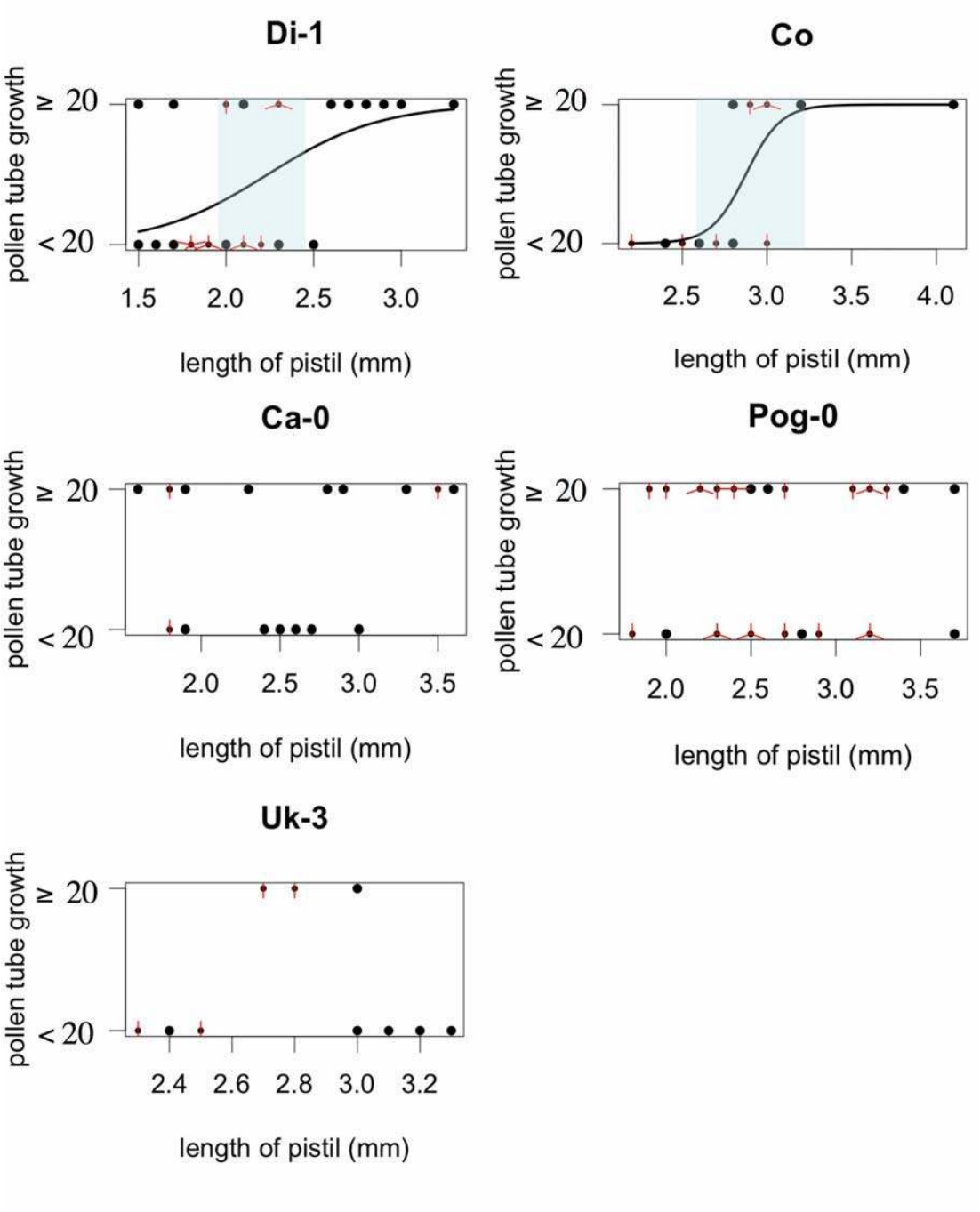

10 20 30 40 50 60 70 80 90 100
SRK-A-Ahalleri MRGAVPNKHHSYIF-F-FLFFVIVLF-PDFSISANTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDNPLSNPI
SRK-A-Ca-0 MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-A-CO MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-A-Ge-1 MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISNNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-A-Wei-1 MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-A-Di-1 MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-A-Uk-3 MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-A-Gie-0 MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-A-Old-1 MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-A-Fi-1 MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-A-Ws-0 MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-A-Nok-0 MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-A-Pog-0 MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-A-Col-0 MRGELPNKHHSYTF-FVFLFFFLILF-PDLSISVNTLSATESLTISSNKTIVSPGGVFELGFFRILGDS--WYLGIWYKKISQRTYVWVANRDTPLSNPI
SRK-B-Cvi-0 MRGELPNKHHSYIF-FAFLFFVLVLF-PDFSISTNTLSATESLTISSNKTIVSPGGVFELGFFKILGDS--WYLGIWYKNVSEKTYVWVANRDNPLSDSI
SRK-C-Kas-2 MKGVRKPYHHSYTFSFLLVFVVLILFHPAFSISVNTLSSTETLTISSNRTIVSPGDDFELGFFKT-GTSSLWYLGIWYKKVPQRTYAWVANRDNPLSNSI
110 120 130 140 150 160 170 180 190 200
SRK-A-Ahalleri GILKISNANLVILDNSDISVWTTNLTGAVRSPVVAELLDNGNFVLRDSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFLTSWKSSFDPSSGSF
SRK-A-Ca-0 GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-A-CO GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-A-Ge-1 GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-A-Wei-1 GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-A-Di-1 GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-A-Uk-3 GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-A-Gie-0 GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-A-Old-1 GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-A-Fi-1 GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-A-Ws-0 GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-A-Nok-0 GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-A-Pog-0 GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-A-Col-0 GILKISNANLVILDNSDTHVWSTNLTGAVRSSVVAELLDNGNFVLRGSKINESDEFLWQSFDFPTDTLLPQMKLGRDHKRGLNRFVTSWKSSFDPSSGSF
SRK-B-Cvi-0 GILKITNSNLVLLNHSDTPVWSTHLTGAVRSLVVAELLDNGNFVLKDSKTNDSDRFLWQSFDFPTNTLLPQMKLGLDHKRRLNRFLTSWKNSFDPSSGDY
SRK-C-Kas-2 GTLKISGRNLVLLGHSNKLVWSTNLISGNLRSPVMAELLANGNFVMRYSNNDQGGFLWQSFDYPTDTLLPQMKLGWDRKTGLNRILRSWRSLDDPSSSNY
210 220 230 240 250 260 270 280 290 300
SRK-A-Ahalleri MFKLETLGLPEFFGFTTFLEVYRSGPWDGLRFSGIPEMQQWDDIIYNFTENRDEVAYTFRVTEHNSYSRLTINTVGRLEGFMWEPTQQEWNMFWFMPKDT
SRK-A-Ca-0 MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTINTVGRLEGFMWEPTQQEWNMFWFMPKDT
SRK-A-CO MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTINTVGRLEGFMWEPTQQEWNMFWFMPKDT
SRK-A-Ge-1 MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTIYTVGRLEGFMWEPTQQEWNMFWFMPKDT
SRK-A-Wei-1 MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTINTVGRLEGFMWEPTQQEWNMFWFMPKDT
SRK-A-Di-1 MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTINTVGRLEGFMWEPTQQEWNMFWFMPKDT
SRK-A-Uk-3 MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTINTVGRLEGFTWEPTQQEWNMFWFMPKDT
SRK-A-Gie-0 MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTINTVGRLEGFMWEPTQQEWNMFWFMPKDT
SRK-A-Old-1 MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTINTVGRLEGFTWEPTQQEWNMFWFMPKDT
SRK-A-Fi-1 MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTINTVGRLEGFMWEPTQQEWNMFWFMPKDT
SRK-A-Ws-0 MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTINTVGRLEGFTWEPTQQEWNMFWFMPKDT
SRK-A-Nok-0 MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTINTVGRLEGFMWEPTQQEWNMFWFMPKDT
SRK-A-Pog-0 MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTINTVGRLEGFMWEPTQQEWNMFWFMPKDT
SRK-A-Col-0 MFKLETLGLPEFFGFTSFLEVYRSGPWDGLRFSGILEMQQWDDIIYNFTENREEVAYTFRVTDHNSYSRLTINTVGRLEGFTWEPTQQEWNMFWFMPKDT
SRK-B-Cvi-0 TFKLETRGLTELFGLFTILEVYRSGPWDGRRFSGIPEMEQWDDFVYNFTENREEVFYTFRLTDPNLYSRLTINAAGNLERFTWDPTREEWNRFWFMPKDN
SRK-C-Kas-2 SYKLETRGFPEFFLLDEDVPVHRSGPWDGIQFSGIPEMRQLNYMVYNFTENKDEISYTFQMTNHSIYSRLTVSFSGSLKRFMYIPPSYGWNQFWSIPTDD
310 320 330 340 350 360 370 380 390 400
SRK-A-Ahalleri CDLYGICGPYAYCDMSTSPACNCIKGFQPLSQQEWASGDVTGRCRRKTQLTCGEDRFFKLMNMKLPATTAAVVDKRIGLKECEEKCKTHCNCTAYANSDV
SRK-A-Ca-0 CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRIGLKECEEKCKTHCNCTAYANSDI
SRK-A-CO CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRIGLKECEEKCKTHCNCTAYANSDI
SRK-A-Ge-1 CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRIGLKECEEKCKTHCNCTAYANSDI
SRK-A-Wei-1 CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRIGLKECEEKCKTHCNCTAYANSDI
SRK-A-Di-1 CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRIGLKECEEKCKTHCNCTAYANSDI
SRK-A-Uk-3 CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRIGLKECEEKCKTHCNCTAYANSDI
SRK-A-Gie-0 CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRIGLKECEEKCKTHCNCTAYANSDI
SRK-A-Old-1 CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRIGLKECEEKCKTHCNCTAYANSDI
SRK-A-Fi-1 CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRIGLKECEEKCKTHCNCTAYANSDI
SRK-A-Ws-0 CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRIGLKECEEKCKTHCNCTAYANSDI
SRK-A-Nok-0 CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRTGLKECEEKCKTHCNCTAYANSDI
SRK-A-Pog-0 CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRIGLKECEEKCKTHCNCTAYANSDI
SRK-A-Col-0 CDLYGICGPYAYCDMSTSPTCNCIKGFQPLSPQDWASGDVTGRCRRKTQLTCGEDRFFRLMNMKIPATTAAIVDKRIGLKECEEKCKTHCNCTAYANSDI
SRK-B-Cvi-0 CDKLGICGSYAYCDTSTSPACNCIRGFQPLSPQEWASGDASGKCLRKTQLSCGGDKFFQLMSMKLPDTTTAIVDKRIGLEECEEKCKNDCNCTAYANMDI
SRK-C-Kas-2 CYMYLGCGPYGYCDVNTSPMCNCIRGFKPRNLQEWVLRDGSSGCVRKTQLSCRGDGFVQLKKIKLPDTTSVTVDRRIGSKECKKRCLNDCNCTAFANADN
410 420 430 440 450 460 470 480 490 500
SRK-A-Ahalleri RNGGSGCIIWIGEFRDIRIYAADGQDLFVRLAPAEFGERSNISGKIIGLIIGISLMLVLSFIMYCFWKKKHKRARATAAPIGYRDRIQESIITNGVVM-S
SRK-A-Ca-0 RNGGSGCIIWIGEFRDIRKYAADGQDLFVRLAAAEFGERRTIRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-A-CO RNGGSGCIIWIGEFRDIRKYAADGQDLFVRLAAAEFGERRTIRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-A-Ge-1 RNGGSGCIIWIGEFRDIRNYAADGQDLFVRLAAAEFGERRTIRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-A-Wei-1 RNGGSGCIIWIGEFRDIRNYAADGQDLFVRLAAAEFGERRTIRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-A-Di-1 RNGGSGCIIWIGEFRDIRNYAADGQDLFVRLAAAEFGERRTIRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-A-Uk-3 RNGGSGCIIWIGEFRDIRNYAADGQDLFVRLAAAEFGERRTSRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-A-Gie-0 RNGGSGCIIWIGEFRDIRNYAADGQDLFVRLAAAEFGERRTIRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-A-Old-1 RNGGSGCIIWIGEFRDIRNYAADGQDLFVRLAAAEFGERRTSRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-A-Fi-1 RNGGSGCIIWIGEFRDIRNYAADAQDLFVRLAAAEFGERRTIRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-A-Ws-0 RNGGSGCIIWIGEFRDIRNYAADGQDLFVRLAAAEFGERRTSRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-A-Nok-0 RNGGSGCIIWIGEFRDIRNYAADGQDLFVRLAAAEFGERRTIRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-A-Pog-0 RNGGSGCIIWIGEFRDIRNYAADGQDLFVRLAAAEFGERRTIRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-A-Col-0 RNGGSGCIIWIGEFRDIRNYAADGQDLFVRLAAAEFGERRTIRGKIIGLIIGISLMLVLSFIIYCFWKKKQKRARATAAPIGYRDRIQELIITNGVVM-S
SRK-B-Cvi-0 RNGGPGCVIWIGEFQDIRKYASAGQDLYVRLAAADIRERSNISRKIIGLIVGISLMLVVSFIIYCFWKRKHKRARATAAAIXFRERIQGF-VTNGVVV-S
SRK-C-Kas-2 KNEGSGCVIWTGELVDIRNYATGGQNLYVRIAAADIDKGVKVSGKIIGLIAGVSIMLLLSFTMLCIWKRKQKGARARE--IVYQEKTQDL-IMNEVAMKS



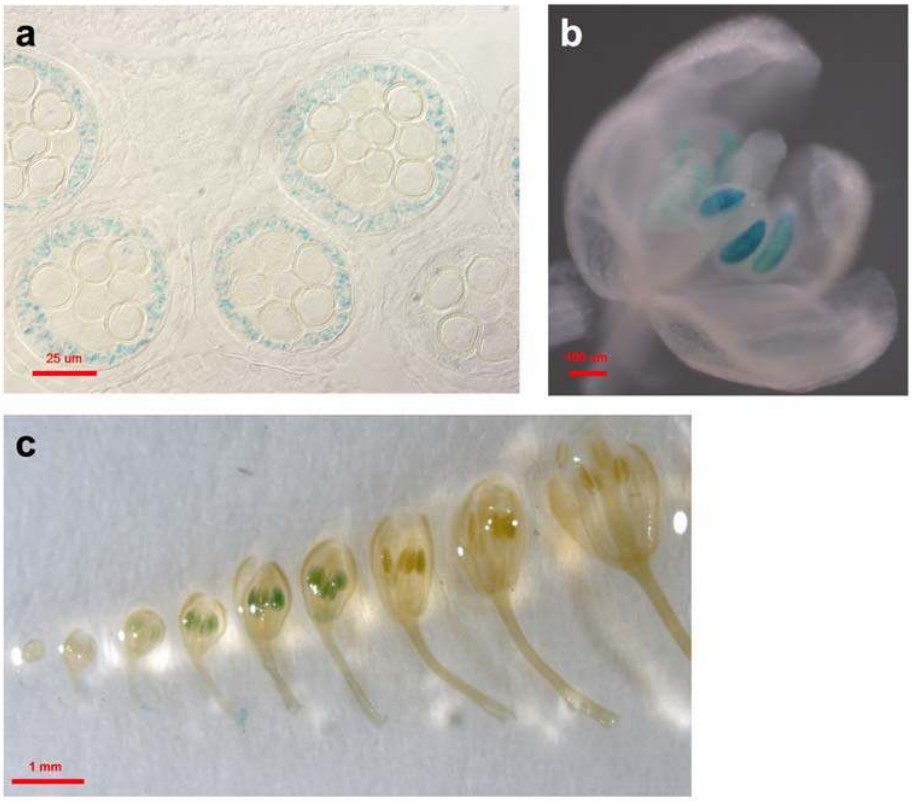
Supplementary Figure 4. Alignment of predicted amino acid sequences of SRK-A from *A. halleri* and several *A. thaliana* accessions. The amino acid sequences were deduced from the cDNA sequences. The cDNA sequences were also confirmed by genomic sequences. SRK-B of Cvi-0 and SRK-C of Kas-2 are also included[11]. Twelve *A. thaliana* accessions with full-length SRK-A are indicated in red. A splicing mutation causing a frameshift in SRK-B is shown as X indicated by an arrow (position 482), and subsequent amino acids are shown as if the frameshift did not happen. * denotes stop codons.



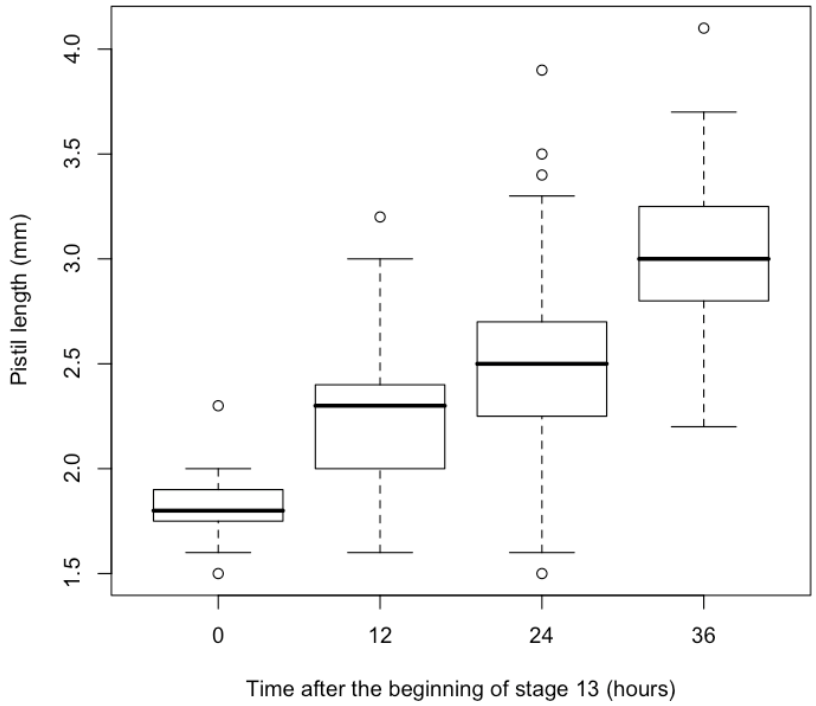
Supplementary Figure 5. Geographical distribution of *A. thaliana* accessions with intact *SRK*. Information from Korves *et al.*[48] was used. This map was drawn by R 2.8.1 with the package `maptools`.



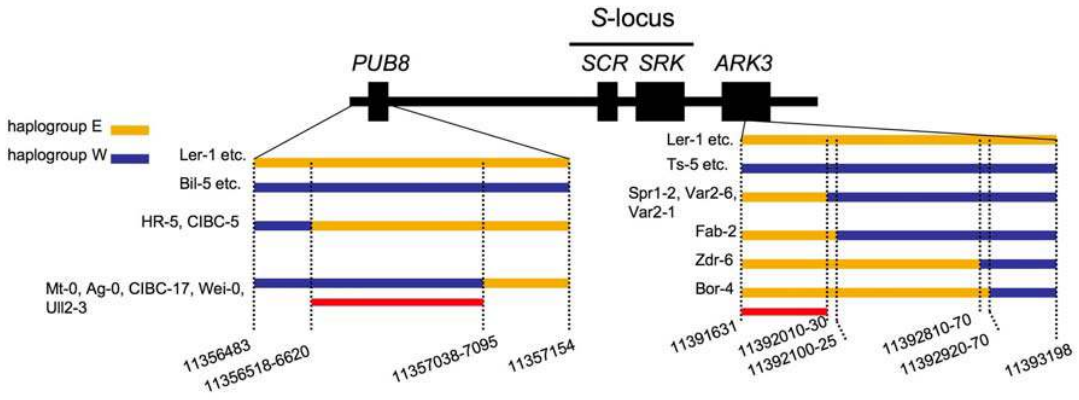
Supplementary Figure 6. Relationships between developmental stages and incompatible phenotypes in five accessions. Red lines indicate the number of samples plotted at the same position (sunflower plots). If significant, binomial regressions based on GLM are shown as black lines (Di-1: $p = 0.002$, Co: $p = 0.0005$, Ca-0: $p = 0.31$, Pog-0: $p = 0.95$, Uk-3: $p = 0.71$). Light blue areas indicate the mean length of pistils (± s. d.) when anthers touch the stigma of the same flowers (Di-1: *n* = 7, Co: *n* = 6).



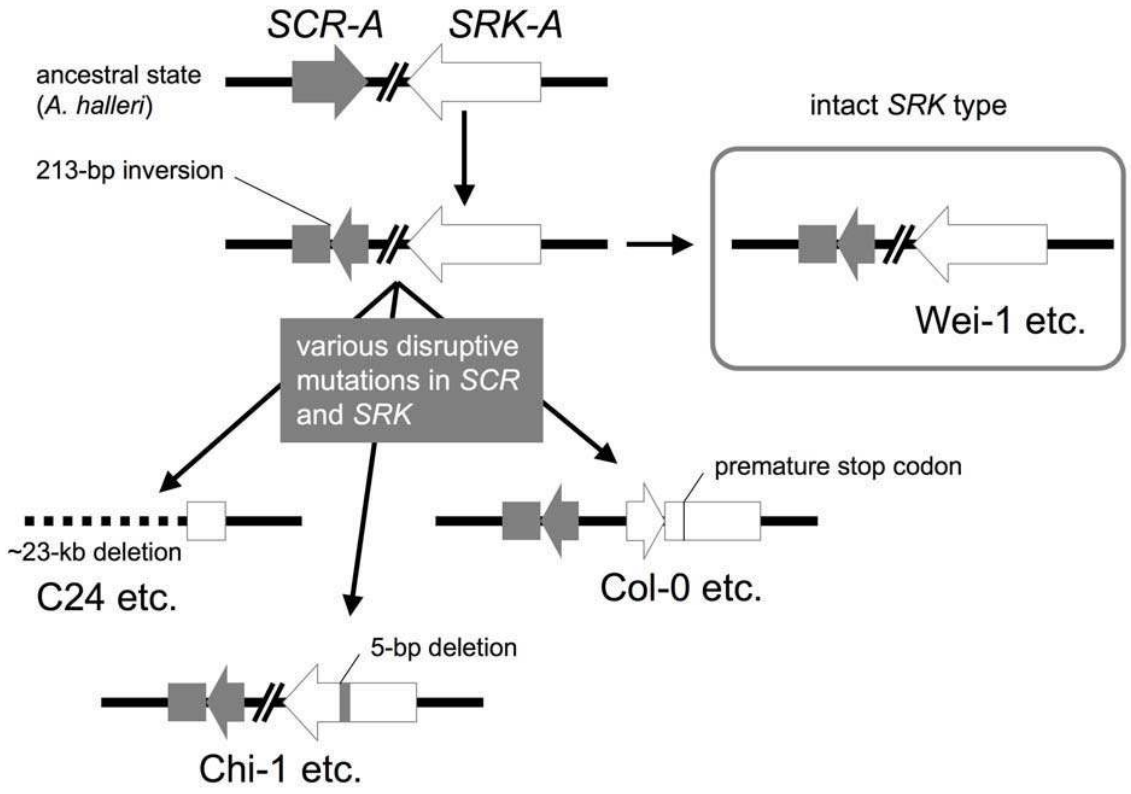
Supplementary Figure 7. Expression pattern of the *ATA7* promoter with GUS reporter in Wei-1. a, A cross-section of an anther; the GUS staining was detected in tapetum cells. Scale bar = 25 μm. **b,** Expression in anthers. Scale bar = 100 μm. **c,** Expression changes during development (left to right). Scale bar = 1 mm.



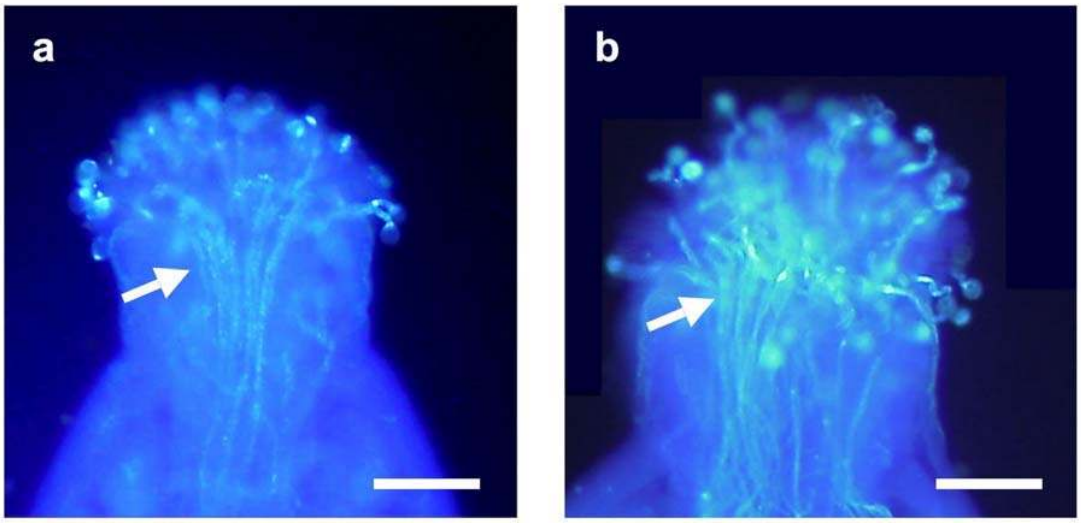
Supplementary Figure 8. Relationship between pistil length and the time after the beginning of stage 13. Bars represent the median, boxes the interquartile range, and whiskers extend out to 1.5 times the interquartile range. Pistil length was significantly influenced by developmental stage, i.e., time after the beginning of stage 13 in the accessions listed in Supplementary Table 2 ($p < 2.0 \times 10^{-16}$, *N* = 315, GLM and Poisson error structure were used).



Supplementary Figure 9. Recombinant haplotypes within *ARK3* and *PUB8*. Six independent recombination points were found within *PUB8* (two) and *ARK3* (four). All recombinants had haplogroup E (either *ARK3-E* or *PUB8-E*) proximal to *SRK* and haplogroup W distal to it. Red regions were used for genotyping (results shown in Fig. 1). Physical positions based on the Col-0 genome sequence are shown.



Supplementary Figure 10. Putative disruptive mutation events in haplogroup A. Several representative mutations are shown (not exhaustive). Arrows indicate the 5′- to 3′-orientation of the genes. The hatchmarks between genes or gene fragments indicate that the distance and orientation are not known. Genes are not drawn to scale.



Supplementary Figure 11. Compatible reactions at the pistils of transformants, Wei-1-SI-2 (a) and Wei-1-SI-10 (b). Col-0 was used as the pollen donor. Arrows indicate bundles of pollen tubes. Scale bar = 0.1 mm.