Molecular Ecology (2008) **17**, 704–714 doi: 10.1111/j.1365-294X.2007.03605.x

Independent origins of self-compatibility in *Arabidopsis thaliana*

KENTARO K. SHIMIZU,*† RIE SHIMIZU-INATSUGI,* TAKASHI TSUCHIMATSU*‡ and MICHAEL D. PURUGGANAN†§

*Institute of Plant Biology, University of Zurich, Zollikerstrasse 107, CH-8008, Zurich, Switzerland, †Department of Genetics, Box 7614, North Carolina State University, Raleigh, NC 27695, USA, ‡Department of General Systems Studies, University of Tokyo, 3-8-1 Komaba, Meguro, Tokyo 153–8902, Japan, §Department of Biology and Center for Genomics and Systems Biology, New York University, 1009 Silver, 100 Washington Square East, New York, NY 10003, USA

Abstract

The evolution from outcrossing based on self-incompatibility (SI) to a selfing system is one of the most prevalent transitions in flowering plants. It has been suggested that the loss of SI in *Arabidopsis thaliana* is associated with pseudogene formation at the *SCR* male component of the *S* locus. Recent work, however, suggests that alternative alleles with large deletions at the *S* locus are also present and may be responsible for the evolution of self-compatibility in this species. We demonstrate that most of these deletion alleles are evolutionarily derived from an *S* haplotype (haplogroups A) that already possessed the *SCR* pseudogene. This haplotype and its deletion variants are nearly fixed in Europe. Together with previous transgenic data, these results suggest that the pseudogenization of Ψ*SCR1* gene changed the SI phenotype in the majority of *A. thaliana* accessions, and was a critical step in the evolution of selfing in this species. Two other haplogroups (B and C) were also identified, the former of which contains a novel and possibly functional *SCR* allele. In contrast to haplogroups A, these two haplogroups are found primarily in Africa and Asia. These results suggest that self-compatibility, which appears to be fixed in this species, arose multiple times with different genetic bases, and indicates that a species-specific trait is associated with parallel evolution at the molecular level.

Keywords: *Arabidopsis thaliana*, evolutionary genomics, molecular adaptation, reproductive strategies, self-incompatibility, *S* locus

Received 31 July 2007; revision received 20 September 2007; accepted 6 October 2007

Introduction

Determining the genetic basis of parallel origins of the same phenotype is critical to understanding models of phenotypic evolution (Shimizu 2002; Mundy *et al*. 2004; Hoekstra & Coyne 2007). Stebbins (1974) reported that 'the evolutionary pathway from obligate outcrossing based upon self-incompatibility to predominant self-fertilization' is one of the most prevalent parallel evolutionary transitions in flowering plants. Charles Darwin proposed the earliest model to explain the prevalence of selfing, hypothesizing that selfing can be evolutionarily advantageous when pollinators or mates are scarce (Darwin 1876). Since then, there have been concerted efforts to understand the genetic basis for the evolution of self-compatibility and the consequences that inbreeding poses on organismal populations (Bustamante *et al*. 2002; Mable *et al*. 2005; Newbigin & Uyenoyama 2005; Shimizu & Purugganan 2005; Glemin *et al*. 2006; Igic *et al*. 2006).

Many Brassicaceae species have a self-incompatibility (SI) system to prevent selfing (80/182 compiled by Bateman 1955; 14/61 compiled by Mable 2004). The studies using *Brassica* and *Arabidopsis* revealed that the sporophytic self-incompatibility system is encoded by the *S* (sterility) locus, which comprises a gene complex containing at least two functional genes (Bateman 1955; Stein *et al*. 1991; Schopfer *et al*. 1999; Suzuki *et al*. 1999; Takasaki *et al*. 2000; Takayama *et al*. 2000, 2001; Kachroo *et al*. 2001; Kusaba *et al*. 2001;

Correspondence: Kentaro K. Shimizu, Fax: +41-44-634-8204; E-mail: shimizu@botinst.uzh.ch

Schierup *et al*. 2001; Nasrallah 2002). The *SRK/Aly13* gene encodes a transmembrane Ser/Thr receptor kinase expressed in the stigma, while the *SCR/SP11* gene encodes a small cysteine-rich protein found in pollen coats that acts as a ligand to the SRK receptor protein. Balancing selection has maintained a number of haplotypes with high sequence divergence. SCR and SRK proteins from the same haplotype bind together specifically, and result in downstream signalling through protein kinases and U-box proteins to inhibit male pollen tube development (Liu *et al*. 2007). Inactivation of at least one of the components of this system is a necessary step in the evolution of selfing.

Arabidopsis thaliana is a predominantly self-fertilizing species with an outcrossing rate estimated at ~1% (Abbott & Gomes 1989). The whole genome sequence of the *A. thaliana* Col-0 accession (or ecotype) showed that the *SCR* and *SRK* genes at the *S* locus are present as the pseudogenes (Kusaba *et al*. 2001). Ψ*SCR1* (*pseudoSCR1*) encodes a truncated open reading frame without five of eight conserved cysteine residues critical for protein structure, and in addition, two highly truncated sequences named Ψ*SCR2* and Ψ*SCR3* were found. Ψ*SRK* has a premature stop mutation (Kusaba *et al*. 2001). Previous analysis indicates that variants of *SRK* at the *S* locus in this species belong to one of three major haplotype groups, referred to as haplogroup A (including Col-0), B and C (Shimizu *et al*. 2004). These variants will be designated here as *SRK* (or *SRK-A*, *SRK-B*, and *SRK-C* for each) instead of Ψ*SRK*, since some of the variants may be intact (Shimizu *et al*. 2004 and see below). Transgenic introduction of functional copies of *SCR* and *SRK* from the sister species *Arabidopsis lyrata* into the *A. thaliana* C24 accession renders it fully self-incompatible. Similar experiments in six more accessions, including Col-0, leads to partial self-incompatibility, with the decay of SI at a later stage of stigma development. These transgenic experiments provide causal evidence that the disruption of *SCR* and/or *SRK* underlies the evolution of self-compatibility of those accessions (Nasrallah *et al*. 2002; Nasrallah *et al*. 2004). In the latter six accessions, a U-box gene was reported to be responsible for the partial rescue and may also have contributed to the evolution of partial self-compatibility (Liu *et al*. 2007).

Previous analysis using 24 *A. thaliana* accessions suggested a single origin of selfing, followed by positive selection on the Ψ*SCR1* pseudogene (Shimizu *et al*. 2004). Since then, population structure in this species has been reported based on genome-wide single nucleotide polymorphism (SNP) data (Nordborg *et al*. 2005; Schmid *et al*. 2006). Moreover, subsequent studies indicated that a deletion of the Ψ*SCR1* sequence was found in ~23% of *A. thaliana* accessions and that part of the genotyping in Shimizu *et al*. (2004) was due to spurious polymerase chain reaction (PCR) amplification (Sherman-Broyles *et al*. 2007). This raised the possibility of independent origins of selfing that could arise from strong selective pressures for self-fertilization (Nasrallah *et al*. 2004; Sherman-Broyles *et al*. 2007).

The lack of Ψ*SCR1*, however, may be explained as a secondary loss rather than an independent origin of self-compatibility. To determine the origins of these deletions, we have examined those accessions that appear to have a deletion as well as the variant *SRK* haplotypes. Here, we show that there is a significant fraction of *S* haplotypes that have a large deletion, and these appear to be secondarily derived from the predominant haplogroup A. Moreover, we do find a novel full-length *SCR* allele in the offshore African Islands, and provide strong evidence of independent origins of self-compatibility in this species.

Materials and methods

Amplification and sequencing of alleles

Arabidopsis thaliana accessions were obtained from the Arabidopsis Biological Resource Center (ABRC). The accessions used for sequencing are shown in Table S1, Supplementary material. Genomic DNA was isolated from young leaves by using the Plant DNeasy Mini kit (QIAGEN). PCR was performed with *Taq* DNA polymerase (Roche), Go-*Taq* polymerase (Promega) or Ex*Taq* (TaKaRa). DNA fragments were purified with the QIAquick PCR Purification and Gel Extraction kits (QIAGEN) if required and sequenced directly. TAIL-PCR (Liu & Whittier 1995; Terauchi & Kahl 2000) and BD GenomeWalker Universal Kit (BD) were used to isolate flanking sequences. Primers used for amplification and genotyping are shown in Table S2, Supplementary material. DNA sequencing was conducted at the Institute of Plant Biology, University of Zurich, with 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 visually confirmed, and ambiguous polymorphisms were rechecked with PCR re-amplification and sequencing. GenBank Accession nos for these genes are EF692396–EF692486, EU083397–EU083398, EU162138 and AY772638.

Molecular population genetics analysis

Sequences were aligned manually against the whole genome *A. thaliana* Col-0 sequence (Arabidopsis Genome Initiative 2000). CS6622 (Bla-10) and CS1044 (Bu-21) used in Shimizu *et al*. (2004) were removed from the sequence analysis, since heterogeneity and/or heterozygosity were suggested in genotyping. Indeed, both Bla and Bu populations are highly polymorphic in the genotypes of S locus (Table S3, Supplementary material). Instead, CS6652 (Bu-21), which is a single-plant descent from the bulk stock of CS1044 (Bu-21), was sequenced and analysed (Table S1). Note that

nucleotide diversity of *ARK3* may be overestimated because of closely linked paralogues (Hagenblad *et al*. 2006; Sherman-Broyles *et al*. 2007). Polymorphism analyses were conducted by using DNASP 4.10.7 (Rozas *et al*. 2003). Levels of silent or synonymous site nucleotide diversity were estimated as π (Tajima 1983). Recombination analysis was conducted using four-gamete test as implemented in DNASP 4.10.7 to estimate all R_M (minimum number of recombination events) pairs of sites where it is possible to assign at least one recombination event (Hudson & Kaplan 1985). The *SCR* sequences of Cvi-0 and Col-0 were aligned with known *SCR* alleles of *Arabidopsis lyrata* and *Arabidopsis halleri* (*AlSCR37*, GenBank Accession no. DQ520280, *AhSCR04*, GenBank DQ520277, *SCRa*, AB052703, *SCRb*, AB052754) (Kusaba *et al*. 2001; Bechsgaard *et al*. 2006). Phylogenetic analysis of *SCR* was conducted using the neighbour-joining method with the Kimura 2-parameter model and implemented in MEGA 3.1 (Kumar *et al*. 2004). Parsimony analysis of long terminal repeat (LTR) phylogeny was conduced using PAUP* version 4.0 (Swofford 1999). Node confidence was determined by 1000 bootstrap replicates of the data, and bootstrap values supported in more than 50% are indicated.

Expression analysis and the isolation of complementary DNAs

Total RNA was extracted from floral bud and flower tissues of Col-0, Cvi-0, Nok-0 and Pog-0 plants with the RNeasy kit (QIAGEN). The complementary DNA (cDNA) was synthesized with the RETROscript reverse transcription kit (Ambion). Reverse transcription (RT)-PCR with 25 cycles was conducted for each gene pair by using primers anchored in exons but designed to amplify across introns. The constitutively expressed gene *ACT8* was used as a control for cDNA quality and to equalize loadings of RT-PCR products (shown in fig. 4 A of Shimizu *et al*. 2004 using the same cDNA). Negative control reaction with no reverse transcriptase yielded no PCR amplification, which confirmed that the amplified bands were derived from RNA (data not shown). The 3′ ends of the Ψ*SRK* cDNA of Col-0, Pog-0 and Cvi-0 were isolated with the 5′/3′ RACE Kit, 2nd Generation (Roche). GenBank Accession nos for the longest cDNA sequences are EF692487–EF692490.

DNA gel blotting

The protocol based on Hebbar and Curtis (Hebbar & Curtis 2000) was used for DNA gel botting. Genomic DNA of each accession (1.4 μg) was digested with *Dra*I, and run on 1% agarose gels and blotted to Nytran membranes. A probe DNA of 1.5 kb including the putative start and stop codons of Ψ*SCR1* was amplified by primers PseSCRF1 and PseSCRR3. The Alkphos Direct kit (Amersham) was used to label the probe according to the manufacturer's instructions. Another gel blotting experiment using *Hin*cII yielded a similar result (data not shown).

Genotyping

We re-genotyped the S locus in 297 *A. thaliana* accessions that represent a nonredundant geographical collection. In this collection, accessions from the same locality with redundant genotype based on SNP data (Schmid *et al*. 2006) were removed as described by Korves *et al*. (2007). In addition to the stock centre accessions, an accession from Ketumbeine Mountain in northern Tanzania, collected and donated by William Kindeketa and James Campanella, was included for the broader coverage of the distribution range. Primers were designed to assay presence/absence polymorphisms of eight regions at *S* locus of haplogroup A: two large deletions (A-t2, A-t3), the 5′ portion of *SRK-A* (encompassing start codon), the middle of *SRK-A* (*S* domain coding region) and the 3′ portion of *SRK-A* (characterizing full-length alleles of A-t4 and A-t5), the 5-bp deletion of A-t5, Ψ*SCR1*, and the U-box gene. We also genotyped the haplogroup B and C alleles using primers to amplify the *S* domain of *SRK*, as previously described in Shimizu *et al*. (2004). Positive control *ACT8* was successfully amplified from all 297 accessions. In addition, PCR of the A-t3 deletion worked as a negative control in that it was not amplified except for one accession. However, it should be noted that genotyping of presence/absence polymorphism may not yield clear-cut results compared with length polymorphisms. Faint amplification was sometimes observed (Table S3), which may suggest PCR failure, subtle contamination, mutations or rearrangements including inverted repeats.

Results

Deletion derivatives of the S locus

We analysed Ag-0, Sha and Mt-0 accessions (Table S1), which were removed from previous population genetic analyses because of lack of PCR amplification in *SRK* (supporting table S1 of Shimizu *et al*. 2004). A Southern blotting experiment suggested a rearrangement of the genomic structure of Sha and Mt-0 at this locus (Nasrallah *et al*. 2004).

We confirm a large deletion removing ~23 kb of the *S* locus that includes Ψ*SCR1* as well as most of *SRK-A* (Fig. 1, Table S2), which was reported from the C24 accession (Sherman-Broyles *et al*. 2007). We refer to the original Col-0 haplotype as A-t1 (type 1), and the deletion haplotype as A-t2 (see Fig. 1). The 5′ breakpoint of this deletion is within the LTR of an upstream *Ty1/copia*-like retrotransposon sequence (At4g21360) found in this genomic region, while the 3′–deletion breakpoint is within the first exon of *SRK*. Between them is an inverted repeat of ~1 kb which includes *SRK* sequence found in haplogroup A; however, we and

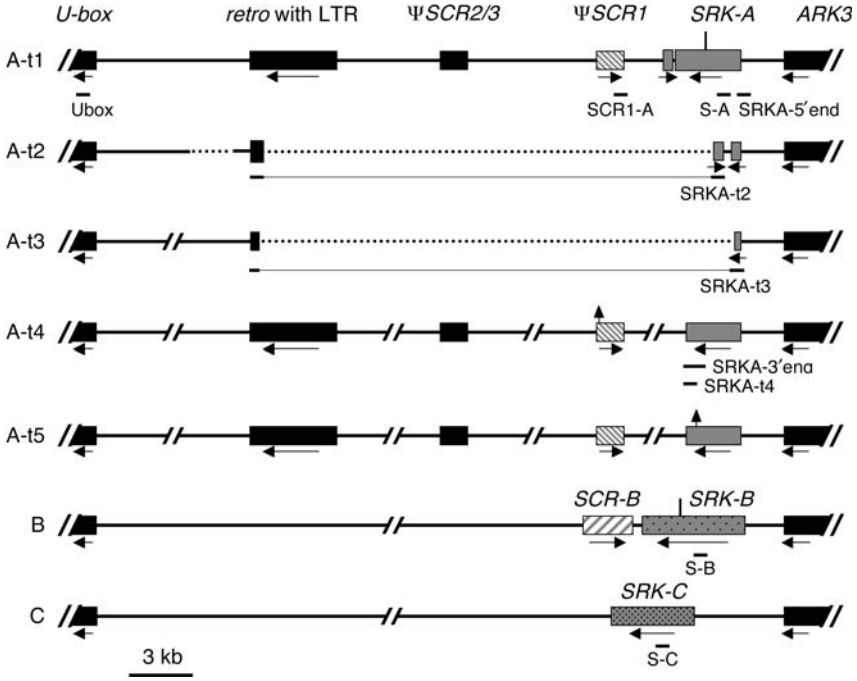

Fig. 1 Schematic diagram of the structure of the predominant haplotype (A-t1) at the *Arabidopsis thaliana S* locus and related variant alleles. The relative locations of genes in the genomic region are indicated, and the arrows give gene orientation. Horizontal bars indicate the relative positions of the genotyping PCR primers. Note that the length of genotyping PCR products and the sequences flanking deletions and inversions do not strictly reflect the original size, in order to show the relative positions clearly. Broken lines show two large deletions in the A-t2 and another in the A-t3 variants alleles. Breaks in the lines indicate that the lengths and arrangements between the genes are unclear. Ψ*SCR1* pseudogene in A-t1, A-t4 and A-t5 encodes a truncated open reading frame without five of eight conserved cysteine residues. The triangle in variant A-t4 shows the relative position of a 5-bp frameshift deletion, which further degraded the Ψ*SCR1* pseudogene. The triangle in variant A-t5 shows the relative position of a 5-bp frameshift deletion at the *SRK* gene. Vertical lines on *SRK* of A-t1 and B indicate a premature stop mutation and a splicing mutation, respectively. The region shown here includes *SRK*, *SCR*, LTR and part of the flanking genes U-box and *ARK3*, and it is reported that a duplicated *ARK3* and a fragment of *SRK-C* sequences are located in the 3′ side of the depicted region of A-t2. A-t1 and A-t2 are based on GenBank Accession nos NC_003075 of Col-0 and EF182720 of C24, respectively, and A-t3, A-t4, A-t5, B and C are based on Mir-0, Pog-0, Chi-1, Cvi-0 and Kr-0, respectively.

others (Sherman-Broyles *et al*. 2007) could not sequence this region.

We also observe a different type of deletion in one accession, Mir-0, which we designate as A-t3. The 5′–deletion breakpoint for this haplotype is also located in the retrotransposon LTR, but starts 124 bp distal from where the A-t2 deletion begins; the 3′ border is also in the middle of *SRK* exon 1, leaving 914 bp of coding sequence.

In our previous analysis (Shimizu *et al*. 2004), no apparent disruptive mutations were found in exon 1–5 of *SRK* in the Pog-0 and Nok-0 accessions. We refer to this haplogroup A variant as the A-t4 haplotype. We isolated the cDNA of *SRK* containing all seven exons of this gene using rapid amplification of cDNA ends (RACE) PCR and RT-PCR. Predicted amino acids of A-t4 was aligned with *SRK* haplogroup B and C sequence of *Arabidopsis thaliana* (Shimizu *et al*. 2004), and with *SRKa* and *SRKb* of *Arabidopsis lyrata* (Kusaba *et al*. 2001) (Fig. S1, Supplementary material). The position of the stop codon is conserved among all *Arabidopsis* haplotypes in which the stop position is known, suggesting that A-t4 has a full-length allele of the *SRK* gene. Moreover, the *SRK* A-t4 allele has no nonsense or frameshift mutations, nor the inverted repeat found in A-t1 (Fig. S1). Together, these sequence data suggest that haplogroup A-t4 contains an intact allele of the *SRK* gene.

We also identified a derivative of this A-t4 haplotype, which has the identical structure but contains a 5-bp frameshift deletion in exon 5 (see Fig. 1) and 27-bp deletion in exon 7 compared with the *SRK* gene of A-t4; we refer to this haplogroup A deletion variant as A-t5. Ψ*SCR1* was found both in A-t4 and in A-t5 as well as in A-t1 as reported (Shimizu *et al*. 2004). Its truncated open reading frame lacks five of eight conserved Cys residues (Fig. 2a). In addition, sequencing of the entire exon 1 revealed that Ψ*SCR1* of A-t4 was further degraded by the 5-bp frameshift deletion.

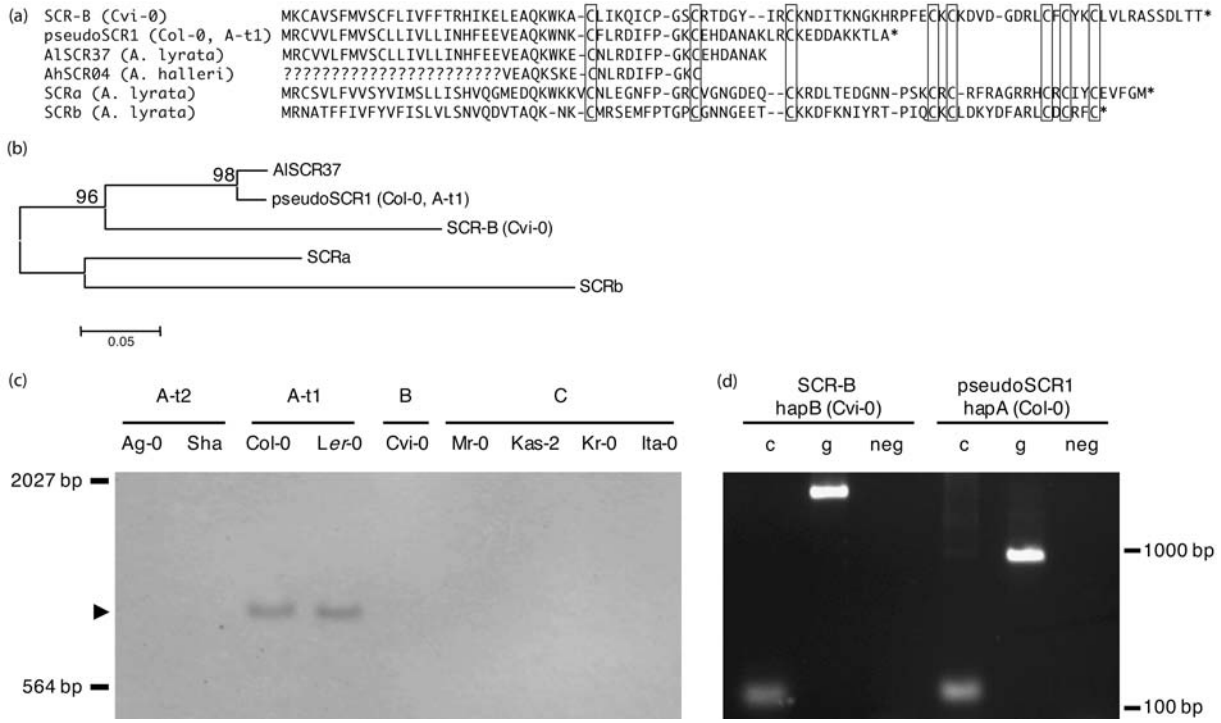

Fig. 2 Novel *SCR* sequence from *Arabidopsis thaliana* haplogroup B. (a) Partial alignment of predicted amino acid sequences of *SCR* haplogroups. *SCR-B* of haplogroup B Cvi-0 and *pseudoSCR1* (Ψ*SCR1*) of A-t1 Col-0 were aligned with *SCR* sequences of *Arabidopsis lyrata* and *Arabidopsis halleri*. *denotes stop codon. Boxes denote eight conserved cysteine residues. (b) Unrooted, nucleotide-based NJ tree showing the relationship of the *SCR* in haplogroup B to other alleles/genes. (c) Southern blotting of Ψ*SCR1* of haplogroup A. A band of the expected length 1.3 kb was found in accessions of haplogroup A-t1, while it was not observed in haplogroups A-t2, B, or C. (d) RT-PCR of *SCR*. Expected lengths of spliced bands were confirmed from the sequencing of cDNA of both haplogroups A and B. In addition, a faint band with the same length as the genomic control was repeatedly found in haplogroup A, suggesting that a small portion of the transcript of this pseudoallele was not spliced. c, cDNA; g, genomic control including intron 1; neg, negative control using no reverse transcriptase.

The isolation of a full-length *SRK* gene of A-t4 also revealed that the *SRK* gene of Col-0 (A-t1) was disrupted not only by the nonsense mutation in exon 4 but also by the deletion of the end of exon 7 and the skipping of exons 6 and 7, both of which might have been caused by a single rearrangement involving the inverted repeat (Fig. S1).

The deletion haplotypes at S *locus are derived in haplotype A*

We genotyped the 297 *A. thaliana* accession set, containing a nonredundant geographical collection, for the *S* locus (Table 1 and Table S3). The haplogroup A (with all its variants including A-t1 to A-t5) is found in 93.9% in our sample. The canonical A-t1 haplotype is still the most frequent haplotype in *A. thaliana* (54% of accessions, see Table 1). The A-t2 deletion haplotype is also common and is found in 26% of the accessions. In contrast, the deletion derivative A-t3 is found in only one accession. Ψ*SCR1* was not amplified from 91 accessions with A-t2 or from one accession with A-t3, confirming that this pseudogene was deleted. In contrast, Ψ*SCR1* was amplified from almost all other accessions which have haplogroups A-t1, t4 and t5 (185 of 187 accessions), indicating that A-t2 is the only major subgroup with the deletion of Ψ*SCR1* in haplogroup A. One of the two accessions where Ψ*SCR1* could not be amplified, Sah-0, was shown by Southern blot analysis to have a rearranged Ψ*SCR1* rather than deletion (Sherman-Broyles *et al*. 2007).

Both the A-t4 and A-t5 haplotypes together are found in only 7% of the accessions. Further rearrangements appear to have occurred in the *S* locus to create additional possible haplotypes, although their frequencies are relatively minor. For example, 15 accessions had the A-t2 deletion, but either no or faint amplification of the 5′ portion of *SRK* was observed. This type is designated A-t2-var, indicating that further rearrangement(s) may have occurred in the A-t2 haplotype. Recent work also suggests that the A-t2 deletion variant has a chimeric structure, in which a rearrangement has incorporated another *SRK* allele at the distal part of the *S* locus (Sherman-Broyles *et al*. 2007).

Table 1 Frequency of allele types

Allele type	Number	Frequency	Europe	Africa & Asia
A-t1	161	0.542	160	1
A-t2	76	0.256	72	4
A-t2-var	15	0.051	15	0
A-t3	1	0.003	1	0
A-t4	16	0.054	16	0
A-t5	4	0.013	4	0
A-het	1	0.003	1	0
A-others	5	0.017	5	0
B	2	0.007	0	2
C	12	0.040	9	3
A + C (het or duplicate)	2	0.007	1	1
unknown	2	0.007	2	0
Sum	297		286	11

See Table S3 for details of each accession.

To examine the origin of the A-t2 that has a deletion of the Ψ*SCR1* pseudogene, we sequenced a 507-bp portion of the LTR of the upstream *Ty1/copia*-like retrotransposon At4g21360 found at the *S* locus using 33 accessions. This LTR is located right at the 5′ deletion boundary of the large deletion haplotype A-t2 (Fig. 1). The sequence of the LTR is unique and can be easily discriminated from other LTR sequences in the genome. Moreover, this region is easily aligned among all the 33 accessions of haplogroup A, although most of the sequence between *SRK* and U-box were reported to be unalignable because of retro-elements based on the genomic sequence of Col-0 (A-t1) and C24 (A-t2) (Sherman-Broyles *et al*. 2007). As an outgroup, we use the LTR of At4g08108, which is the only sequence in the Col-0 whole genome sequence with more than 90% homology to the At4g21360 sequence in the *S* locus.

The phylogenetic analysis (Fig. 3) suggests that the moderate-frequency haplogroup A-t2 deletion is evolutionarily derived from the haplogroup A-t4 haplotype, which contains the Ψ*SCR1* pseudogene. The LTR sequences of A-t2 are identical to that of Nok-0 (A-t4) in this region except for length variation at a simple repetitive sequence. By PCR and sequencing, we also confirmed that the 20 accessions in this phylogenetic analysis that had haplotypes A-t1, A-t4 and A-t5 contain nonfunctional Ψ*SCR1.* The A-t3 variant has a shorter LTR sequence because of a bigger deletion, and was removed from the phylogenetic analysis. However, the alignable portion (385 bp among 507 bp used in the phylogeny of Fig. 3) is identical to A-t2, suggesting the deletion occurred in the clade with A-t2, A-t4 and A-t5. These results suggest that the pseudogenization of the *SCR* gene, and the accompanying loss of self-incompatibility, had already occurred in the common ancestor of haplogroup A, preceding the deletion that led to the A-t2 haplotype as well as other rearrangements found in haplogroup A.

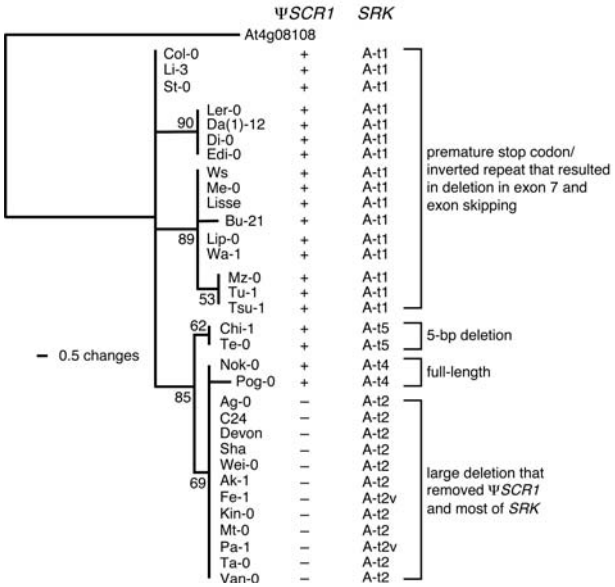

Fig. 3 Phylogeny of the LTR of the retrotransposon At4g21360 found upstream of the Ψ*SCR1* gene (or its deletion border) at the haplogroup A of *Arabidopsis thaliana S* locus. This LTR is found in all variant haplogroup A alleles. The tree is based on a parsimony analysis, and the presence or absence of Ψ*SCR1* and the type of *SRK* genes are mapped onto the phylogeny. The sequence alignment of 507 bp included a gap of 1 bp as a fifth base but excluded an $(AGA)_n$ repeat. LTR sequence of another retrotransposon (At4g08108) was used as an outgroup sequence.

Recombination is known to occur in the *S* locus of haplogroup A of *A. thaliana* (supporting table S2 of Shimizu *et al*. 2004), which may confound any phylogenetic inference of SI haplotypes. We analysed recombination in this genomic region using 15 accessions in which all of U-box, LTR, Ψ*SCR1*, *SRK* and *ARK3* were sequenced. The position of the recombination was estimated using the four-gamete

Table 2 Nucleotide variation at the *Arabidopsis thaliana S* locus of haplogroup A

Gene	U-box	LTR	Ψ*SCR1*	*SRK-A*	*ARK3*
Alignment length with gaps in bps	617	1005	950	2625	833
Number of silent sites (total sites)	161	860	820 (943)	1062 (2590)	385
			A-t1, A-t4 and A-t5 alleles		
π (π of total sites)	0.0506	0.0059	0.0005 (0.0010)	0.0010 (0.0013)	0.0276
Number of alleles	20	20	20	16	18
			All haplogroup A alleles		
π	0.0520	0.0051	NA	NA	0.0267
Number of alleles	33	33	NA	NA	31

π is the estimate of nucleotide diversity at silent sites. Numbers of alleles used for each estimate are shown. The Edi-0 and Bu-21 accessions were removed from the *ARK3* analysis, since double peaks in sequencing (possibly by heterozygosity or tandem duplication) were observed. NA, not applicable.

test (Hudson & Kaplan 1985). Three recombination events were estimated: one between U-box and LTR, one between exon 4 and exon 1 of *SRK*, and one between exon 1 of *SRK* and exon 4 of *ARK3*. In turn, no recombination was detected in the region spanning the LTR, Ψ*SCR1* to exon 4 of *SRK*. This block of linkage disequilibrium includes the major gene disruptive mutations of Ψ*SCR1* and *SRK*, including the premature stop mutation and inverted repeat of A-t1, the 5-bp deletion of A-t5, and the deletion of Ψ*SCR1* and *SRK* of A-t2 and A-t3. This suggests that we can rely on the close linkage between these loci to map evolutionary changes in the latter two genes onto an SI haplotype phylogeny based on the LTR sequence.

We mapped the Ψ*SCR1* deletions and the mutations disrupting the *SRK* gene on the phylogeny of LTR (Fig. 3). The phylogeny indicates that the multiple mutations that resulted in the disruption of the *SRK* gene in *A. thaliana* occurred independently after the pseudogenization of Ψ*SCR1*. Sequences associated with the canonical A-t1 haplotype are found in three separate groups in the tree, but all share the *SRK* inverted repeat and premature stop codon, suggesting that they should belong to one clade. Another clade with 85% bootstrap support contains the intact A-t4 and the 5-bp deletion haplotype A-t5, as well as the A-t2 large deletion haplotypes disrupting *SRK*. Both the phylogeny as well as the structure of these haplotypes (i.e. the implausibility of reversion to an intact sequence from a deletion haplotype variant), suggests that the 5-bp deletion occurred at the branch leading to the Chi-1 and Te-0 accessions (Fig. 3).

Lowest nucleotide diversity at ΨSCR1 *in the S locus of haplogroup A*

We analysed the pattern of nucleotide diversity using 20 accessions that contain Ψ*SCR1* (A-t1, A-t4 and A-t5). This analysis considered silent site positions, since *SRK* as well as Ψ*SCR1* is expressed (Fig. 2b and Shimizu *et al*. 2004). The results show that nucleotide diversity π was lowest at Ψ*SCR1* (π = 0.0005) and second lowest at *SRK* (silent π = 0.0010) among neighbouring loci (Table 2). We also calculated π of total sites of Ψ*SCR1* and *SRK*, since most alleles are pseudoalleles and may be treated as silent. Still, the same pattern is observed (Table 2). Moreover, if we examine all accessions that have haplogroup A variants (including the deletion haplotypes A-t2 and A-t3), we find that the nucleotide diversity levels for the adjacent (undeleted) retrotransposon LTR, U-box and *ARK3* genes are similar to those when considering only the nondeletion variants (see Table 2). These data suggest that a selective sweep occurred in the *S* locus of haplogroup A (see discussion).

Re-evaluation of haplogroups B and C: population structure and a novel SCR allele

The haplotype group A-t1 and its derived deletion haplotypes are found at very high frequency (~94%) in our sample of 297 accessions. In contrast, haplotype groups B and C are found at higher frequency in the limited number of non-European accessions we genotyped (see Table 1). The haplogroup B alleles are present in only two accessions that are found in offshore African island groups: the Cape Verde Island (Cvi-0) and Canary Island (Can-0) accessions. The haplotype group C alleles are found at 4% frequency in the global sample; in Europe they are at 3% frequency, but are at 45% frequency in Africa and Asia. This geographical differentiation in distribution is significant (9/286 in Europe, 5/11 in Africa and Asia, Fisher's Exact test, $P = 0.00004$) and is consistent with recent reports of some population structure of *A. thaliana*, which is possibly associated with the expansion from multiple glacial refugia

in Europe and Asia (Sharbel *et al*. 2000; Nordborg *et al*. 2005; Schmid *et al*. 2006). In these previously published analyses, Cvi-0 and Mr-0 accessions differ significantly from other accessions (Nordborg *et al*. 2005), and consistently have the *SRK* of haplogroup B and C (*SRK-B* and *SRK-C*), respectively (supporting tables S1 and S3 of Shimizu *et al*. 2004).

As opposed to our previous report, our re-examination showed that Ψ*SCR1* of haplogroup A were not detected from accessions that had *SRK* sequence of the haplogroup B and C by both PCR and Southern blot analysis (Sherman-Broyles *et al*. 2007) (Fig. 2c). Our previous report of the Ψ*SCR1* and Ψ*SCR2/3* gene from haplogroups B and C (Shimizu *et al*. 2004) was due to either spurious PCR amplification by the primers (Sherman-Broyles *et al*. 2007) and/or by DNA contaminations (correction submitted). Instead, we do find a novel haplotype of *SCR* in the haplogroup B Cvi-0 accession, which we name *SCR-B* (Fig 2a, b). PCR genotyping with sequencing revealed that the Can-0 accession also has *SCR-B* as well as *SRK-B* (supporting table S3 of Shimizu *et al*. 2004).

SCR-B is tightly linked to an *SRK* gene (*SRK-B*) located on the opposite strand, with only 106 bp between the stop codons (Fig. 1). This allele, as well as the Ψ*SCR1* of haplogroup A, is transcriptionally expressed in floral tissue (Fig. 2d), and is predicted to encode a protein with 92 amino acids. The alignment shows that eight cysteine residues known to be important for SCR protein structure are conserved in SCR-B (Watanabe *et al*. 2000; Sato *et al*. 2002) (Fig. 2a), suggesting that the gene encodes a full-length product with no obvious gene-disruptive mutations.

Using the Mr-0 accession containing haplogroup C, we could not identify any *SCR* allele either by isolating sequences flanking *SRK*, by designing conserved primers in SCR, or using degenerate RT-PCR (Sato *et al*. 2002). This, together with the Southern blot analysis, suggests that there is indeed a deletion of *SCR* in Mr-0 or that it contains a highly divergent *SCR* allele.

Discussion

Previous work suggested that Ψ*SCR1* has the lowest nucleotide diversity in the *S* locus and that a selective sweep at this pseudogene accompanied the rise of selfing (Shimizu *et al*. 2004). We now find that Ψ*SCR1* was not fixed throughout the species. Nevertheless, the high frequency of haplogroup A variants (including the deletion haplotypes A-t2 and A-t3), suggests that selection has driven members of this haplotype group to high frequencies, particularly in Europe. The predominance of one haplotype group in *Arabidopsis thaliana* is all the more impressive given that diversifying selection at SI loci is expected to result in a large number of highly divergent allelic types, as observed in the sister species *Arabidopsis lyrata* (Charlesworth *et al*. 2003).

We also analysed the pattern of nucleotide diversity using 20 accessions that contain Ψ*SCR1* (A-t1, A-t4 and A-t5). The results show that nucleotide diversity π was lowest at Ψ*SCR1* and second lowest at *SRK-A* among neighbouring loci (Table 2), which is consistent with a selective sweep on the *S* locus in haplogroup A. Modelling selection at this locus, however, requires the development of methods to examine partial selective sweeps that arise in a background of balancing selection and population structure. Currently, this is not feasible.

Nevertheless, sequence analysis also implies that the pseudogene Ψ*SCR1*, rather than *SRK-A*, is the most likely target of selection associated with the evolution of self-compatibility in *A. thaliana* in European accessions. Although both *SCR* and *SRK* have mutations that disrupt gene function in the haplogroup A (A-t1 to A-t5), there are multiple independent mutations in *SRK-A* that leads to putative loss-of function alleles, and several accessions also possess what appears to be an intact, expressed SRK locus. Both observations are inconsistent with the hypothesis that *SRK* is the target of recent positive selection that may have led to the high-frequency self-compatible haplotypes in *A. thaliana*. This result, however, is consistent with a model of multiple disruptive mutations of *SRK* occurring during or after the formation of Ψ*SCR1*.

The evolution of selfing involves a number of evolutionary changes such as the loss of self-incompatibility and often changes in floral morphology to allow for self-pollination. Selfing is prevented by self-incompatibility in many Brassicaceae species including *A. lyrata* and *Arabidopsis halleri*, which are close relatives of *A. thaliana*. Thus, the loss of self-incompatibility was a necessary step in the evolution of selfing in *A. thaliana*. So far, *SCR* and *SRK* are the only genes that can cause complete self-compatibility when mutated. Indeed, self-incompatibility can be restored fully or at least partially by introducing functional *SCR* and *SRK* (Nasrallah *et al*. 2004), indicating that the disruption of *SCR* and/or *SRK* caused the transition from full- or partial-self-incompatibility to complete self-compatibility. Here, we showed that the nonfunctional Ψ*SCR1* pseudogene and its derivatives spread to 94% of all the worldwide accessions tested. These data strongly suggest that the pseudogenization of Ψ*SCR1* gene changed the SI phenotype of most individuals, and was a critical step in the evolution of selfing in *A. thaliana*.

It is possible that downstream genes such as U-box changed at an earlier step to result in a mixed mating system based on partial self-compatibility (Liu *et al*. 2007), although these systems might have degraded during or after the evolution of complete self-compatibility. The stability of an intermediate selfing rate is one of the most controversial questions in plant reproductive ecology (Lande & Schemske 1985; Schemske & Lande 1985; Goodwillie *et al*. 2005). Further studies on the relative evolutionary timing

and contributions of *S* genes vs. downstream genes may give new insights into this classic question.

In contrast to the European accessions, the Cvi-0 accession from the Cape Verde Islands and Can-0 accessions from Canary Islands possess *S* haplogroup B, which contains a novel *SCR* sequence expressed in flower tissues and with no obvious gene-disruptive mutations. The *SRK* gene of haplogroup B has a splicing mutation causing a frameshift (Shimizu *et al*. 2004), while its counterparts in self-incompatible *A. lyrata* and *A. halleri* seem to have functional alleles without this mutation (Bechsgaard *et al*. 2006). These results suggest that the ancestor of Cvi-0 and Can-0 had a self-incompatibility system with functional *SRK-B* and *SCR-B*, and that these accessions do not have a haplogroup A Ψ*SCR1* pseudogene. In addition, Cvi-0 has other divergent gene sequences at the *S* locus genomic region, including the U-box gene (Shimizu *et al*. 2004) and *ARK3* (Sherman-Broyles *et al*. 2007). Since Cvi-0 and Can-0 are selfers, these results indicate that the evolution of self-compatibility occurred independently from the pseudogenization of Ψ*SCR1* of haplogroup A.

It appears that the evolution of self-compatibility occurred at least twice in *A. thaliana* — once in an ancestor that gave rise to most of the European accessions and which has carried to near-fixation by selection (~94% of all accessions tested, ~96% in Europe), and the other in the progenitor of the Cape Verde Islands and Canary Islands accessions. In the former case, *SCR* (Ψ*SCR1*) was already pseudogenized in their common ancestor. In the latter case, an obvious candidate mutation that caused the self-compatibility phenotype associated with haplogroup B is the splicing mutation of *SRK-B*.

Molecular genetic studies of phenotypic evolution reveal that parallel evolution is much more prevalent than is apparent from mere consideration of phenotypic states, and independent mutations in the same genes are often observed (Shimizu 2002; Mundy *et al*. 2004; Hoekstra & Coyne 2007). For example, melanism in two bird species occurs as a result of separate mutations in the melanocortin-1 receptor (Mundy *et al*. 2004). In a recent speciation event in *Petunia*, five independent loss-of-function mutations in *AN2* were suggested to have resulted in fixation of white flowers *Petunia axillaries* (Hoballah *et al*. 2007). Within *A. thaliana*, multiple mutations at the flowering time gene *FRIGIDA* result in variation in the early flowering phenotype (Shindo *et al*. 2005). Self-compatibility is fixed within the species *A. thaliana*. It diverged from its self-incompatible relatives an estimated 5–6 million years ago (Koch *et al*. 2000), and so phenotypic observation cannot suggest parallel evolution of self-compatibility. We show here that even a single fixed trait, such as self-compatibility in *A. thaliana*, may be derived from parallel evolutionary changes. As more ecologically important traits are dissected at the molecular level, we may begin to assess to what degree evolutionary diversification is driven by selection on independent mutations that lead to similar phenotypes, and the extent to which this may lead to parallel evolution in organisms.

Acknowledgements

The authors would like to acknowledge Rebecca Thayer and Stephanie Curtis for experimental support, Detlef Weigel, Stephen Schauer and Christina Richards for critical comments on the manuscript. This work was funded in part by grants from the NSF FIBR program and the Department of Defense to M.D.P., and University Research Priority Program of Systems Biology/Functional Genomics of University of Zurich to K.K.S. and to R.S.-I., Grants-in-Aid for Special Research on Priority Areas (No. 119043010) from MEXT, Japan to K.K.S., Forschungskredit of University of Zurich to R.S.-I. and the research fellowship for Young Scientists of the Japan Society for the Promotion of Science to T.T.

References

Abbott RJ, Gomes M (1989) Population genetic structure and the outcrossing rate of *Arabidopsis thaliana*. *Heredity*, **62**, 411–418.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Bateman AJ (1955) Self-incompatibility systems in angiosperms. III. Cruciferae. *Heredity*, **9**, 53–68.

Bechsgaard JS, Castric V, Charlesworth D, Vekemans X, Schierup MH (2006) The transition to self-compatibility in *Arabidopsis thaliana* and evolution within S-haplotypes over 10 Myr. *Molecular Biology and Evolution*, **23**, 1741–1750.

Bustamante CD, Nielsen R, Sawyer SA *et al*. (2002) The cost of inbreeding in *Arabidopsis*. *Nature*, **416**, 531–534.

Charlesworth D, Mable BK, Schierup MH, Bartolome C, Awadalla P (2003) Diversity and linkage of genes in the self-incompatibility gene family in *Arabidopsis lyrata*. *Genetics*, **164**, 1519–1535.

Darwin C (1876) *The Effects of Cross and Self Fertilization in the Vegetable Kingdom*. John Murray, London, UK.

Glemin S, Bazin E, Charlesworth D (2006) Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 3011–3019.

Goodwillie C, Kalisz S, Eckert CG (2005) The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annual Review of Ecology, Evolution and Systematics*, **36**, 47–79.

Hagenblad J, Bechsgaard J, Charlesworth D (2006) Linkage disequilibrium between incompatibility locus region genes in the plant *Arabidopsis lyrata*. *Genetics*, **173**, 1057–1073.

Hebbar PB, Curtis SE (2000) Characterization of devH, a gene encoding a putative DNA binding protein required for heterocyst function in *Anabaena* sp. strain PCC 7120. *Journal of Bacteriology*, **182**, 3572–3581.

Hoballah ME, Gubitz T, Stuurman J *et al*. (2007) Single gene-mediated shift in pollinator attraction in *Petunia*. *Plant Cell*, **19**, 779–790.

Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution*, **61**, 995–1016.

Hudson RR, Kaplan NL (1985) Statistical properties of the number

of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Igic B, Bohs L, Kohn JR (2006) Ancient polymorphism reveals unidirectional breeding system shifts. *Proceedings of the National Academy of Sciences, USA*, **103**, 1359–1363.

Kachroo A, Schopfer CR, Nasrallah ME, Nasrallah JB (2001) Allele-specific receptor–ligand interactions in *Brassica* self-incompatibility. *Science*, **293**, 1824–1826.

Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis, Arabis*, and related genera (Brassicaceae). *Molecular Biology and Evolution*, **17**, 1483–1498.

Korves TM, Schmid KJ, Caicedo AL *et al*. (2007) Fitness effects associated with the major flowering time gene *FRIGIDA* in *Arabidopsis thaliana* in the field. *American Naturalist*, **169**, E141–157.

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, **5**, 150–163.

Kusaba M, Dwyer K, Hendershot J *et al*. (2001) Self-incompatibility in the genus *Arabidopsis*: characterization of the *S* locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *Plant Cell*, **13**, 627–643.

Lande R, Schemske DW (1985) The evolution of self-fertilization and inbreeding depression in plants. 1. genetic models. *Evolution*, **39**, 24–40.

Liu P, Sherman-Broyles S, Nasrallah ME, Nasrallah JB (2007) A cryptic modifier causing transient self-incompatibility in *Arabidopsis thaliana*. *Current Biology*, **17**, 1–7.

Liu YG, Whittier RF (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics*, **25**, 674–681.

Mable BK (2004) Polyploidy and self-compatibility: is there an association? *New Phytologist*, **162**, 803–811.

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

Mundy NI, Badcock NS, Hart T *et al*. (2004) Conserved genetic basis of a quantitative plumage trait involved in mate choice. *Science*, **303**, 1870–1873.

Nasrallah JB (2002) Recognition and rejection of self in plant reproduction. *Science*, **296**, 305–308.

Nasrallah ME, Liu P, Nasrallah JB (2002) Generation of self-incompatible *Arabidopsis thaliana* by transfer of two S locus genes from *A. lyrata*. *Science*, **297**, 247–249.

Nasrallah ME, Liu P, Sherman-Broyles S, Boggs NA, Nasrallah JB (2004) Natural variation in expression of self-incompatibility in *Arabidopsis thaliana*: implications for the evolution of selfing. *Proceedings of the National Academy of Sciences, USA*, **101**, 16070–16074.

Newbigin E, Uyenoyama MK (2005) The evolutionary dynamics of self-incompatibility systems. *Trends in Genetics*, **21**, 500–505.

Nordborg M, Hu TT, Ishino Y *et al*. (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biology*, **3**, e196.

Rozas J, Sanchez-De I, Barrio JC, Messeguer X, Rozas R (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Sato K, Nishio T, Kimura R *et al*. (2002) Coevolution of the *S*-locus genes *SRK, SLG* and *SP11/SCR* in *Brassica oleracea* and *B. rapa*. *Genetics*, **162**, 931–940.

Schemske DW, Lande R (1985) The evolution of self-fertilization and inbreeding depression in plants. 2. empirical observations. *Evolution*, **39**, 41–52.

Schierup MH, Mable BK, Awadalla P, Charlesworth D (2001) Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*. *Genetics*, **158**, 387–399.

Schmid KJ, Torjek O, Meyer R *et al*. (2006) Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *Theoretical and Applied Genetics*, **112**, 1104–1114.

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

Sharbel TF, Haubold B, Mitchell-Olds T (2000) Genetic isolation by distance in *Arabidopsis thaliana:* biogeography and postglacial colonization of Europe. *Molecular Ecology*, **9**, 2109–2118.

Sherman-Broyles S, Boggs N, Farkas A *et al*. (2007) *S* locus genes and the evolution of self-fertility in *Arabidopsis thaliana*. *Plant Cell*, **19**, 94–106.

Shimizu KK (2002) Ecology meets molecular genetics in *Arabidopsis*. *Population Ecology*, **44**, 221–233.

Shimizu KK, Purugganan MD (2005) Evolutionary and ecological genomics of *Arabidopsis*. *Plant Physiology*, **138**, 578–584.

Shimizu KK, Cork JM, Caicedo AL *et al*. (2004) Darwinian selection on a selfing locus. *Science*, **306**, 2081–2084.

Shindo C, Aranzana MJ, Lister C *et al*. (2005) Role of *FRIGIDA and FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. *Plant Physiology*, **138**, 1163–1173.

Stebbins GL (1974) *Flowering Plants: Evolution Above the Species Level*. Harvard University Press, Cambridge, Massachusetts.

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

Suzuki G, Kai N, Hirose T *et al*. (1999) Genomic organization of the *S* locus: identification and characterization of genes in *SLG/SRK* region of S (*9*) haplotype of *Brassica campestris* (syn. *rapa*). *Genetics*, **153**, 391–400.

Swofford DL (1999) PAUP* *Phylogenetic Analysis Using Parsimony (and Other Methods)*, Version 4. Sinauer & Associates, Sunderland, Massachusetts.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Takasaki T, Hatakeyama K, Suzuki G *et al*. (2000) The *S* receptor kinase determines self-incompatibility in *Brassica* stigma. *Nature*, **403**, 913–916.

Takayama S, Shiba H, Iwano M *et al*. (2000) The pollen determinant of self-incompatibility in *Brassica campestris*. *Proceedings of the National Academy of Sciences, USA*, **97**, 1920–1925.

Takayama S, Shimosato H, Shiba H *et al*. (2001) Direct ligand–receptor complex interaction controls *Brassica* self-incompatibility. *Nature*, **413**, 534–538.

Terauchi R, Kahl G (2000) Rapid isolation of promoter sequences by TAIL-PCR: the 5′-flanking regions of Pal and Pgi genes from yams (*Dioscorea*). *Molecular and General Genetics*, **263**, 554–560.

Watanabe M, Ito A, Takada Y *et al*. (2000) Highly divergent sequences of the pollen self-incompatibility (S) gene in class-I S haplotypes of *Brassica campestris* (syn. *rapa*) L. *FEBS Letters*, **473**, 139–144.

Prof. Kentaro K. Shimizu is at the Institute of Plant Biology and a member of the University Research Priority Program of Systems Biology/Functional Genomics, University of Zurich. He studies the evolutionary and ecological genomics of *Arabidopsis* relatives. Dr Rie Inatsugi-Shimizu is at the Institute of Plant Biology, University of Zurich, and studies the molecular and physiological basis of plant reproductive systems. Takashi Tsuchimatsu is a Ph.D. student at the University of Tokyo and at the University of Zurich, and studies the evolutionary origins and consequences of self-fertilization. Prof Michael D. Purugganan is in the Department of Biology, New York University, and studies evolutionary and ecological genomics.

Supplementary material

The following supplementary material is available for this article:

Fig. S1 Alignment of predicted amino acids of *SRK* haplogroups.

Table S1 Accessions used for sequence analysis

Table S2 Primers and PCR

Table S3 Genotyping

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2007.03605.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.