Patterns of Polymorphism at the Self-Incompatibility Locus in 1,083 *Arabidopsis thaliana* Genomes

Takashi Tsuchimatsu,[‡,1,2] Pauline M. Goubet,[†,‡,3] Sophie Gallina,[3] Anne-Catherine Holl,[3] Isabelle Fobis-Loisy,[4] Hélène Bergès,[5] William Marande,[5] Elisa Prat,[5] Dazhe Meng,[1] Quan Long,[6] Alexander Platzer,[1] Magnus Nordborg,[1] Xavier Vekemans,[3] and Vincent Castric*,[3]

[1]Gregor Mendel Institute, Austrian Academy of Sciences, Vienna Biocenter (VBC), Vienna, Austria

[2]Department of Biology, Chiba University, Inage-ku, Chiba, Japan

[3]Université de Lille CNRS, UMR 8198-Evo-Eco-Paleo, Lille, France

[4]Reproduction et Développement des Plantes, Univ. Lyon, Centre National de la Recherche Scientifique, Institut National de la Recherche Agronomique, Université Claude Bernard Lyon I, Ecole Normale Supérieure de Lyon, Lyon, France

[5]Centre National des Ressources Génomiques Végétales, INRA UPR 1258, Castanet-Tolosan, France

[6]Department of Biochemistry and Molecular Biology & Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[†]These authors contributed equally to this work.

[†]Deceased.

*Corresponding author: E-mail: vincent.castric@univ-lille1.fr.

Associate editor: Stephen Wright

Abstract

Although the transition to selfing in the model plant *Arabidopsis thaliana* involved the loss of the self-incompatibility (SI) system, it clearly did not occur due to the fixation of a single inactivating mutation at the locus determining the specificities of SI (the *S*-locus). At least three groups of divergent haplotypes (haplogroups), corresponding to ancient functional *S*-alleles, have been maintained at this locus, and extensive functional studies have shown that all three carry distinct inactivating mutations. However, the historical process of loss of SI is not well understood, in particular its relation with the last glaciation. Here, we took advantage of recently published genomic resequencing data in 1,083 *Arabidopsis thaliana* accessions that we combined with BAC sequencing to obtain polymorphism information for the whole *S*-locus region at a species-wide scale. The accessions differed by several major rearrangements including large deletions and interhaplogroup recombinations, forming a set of haplogroups that are widely distributed throughout the native range and largely overlap geographically. "Relict" *A. thaliana* accessions that directly derive from glacial refugia are polymorphic at the *S*-locus, suggesting that the three haplogroups were already present when glacial refugia from the last Ice Age became isolated. Interhaplogroup recombinant haplotypes were highly frequent, and detailed analysis of recombination breakpoints suggested multiple independent origins. These findings suggest that the complete loss of SI in *A. thaliana* involved independent self-compatible mutants that arose prior to the last Ice Age, and experienced further rearrangements during postglacial colonization.

Key words: evolution of selfing, *Arabidopsis*, 1001 genomes project, self-incompatibility, evolutionary transition.

Introduction

The transition from outcrossing to self-fertilization (selfing) is one of the most frequent evolutionary processes in flowering plants (Stebbins 1974, Barrett 2002, Igic et al. 2008, Shimizu and Tsuchimatsu 2015). Although selfing may lead to inbreeding depression, it has advantages *via* reproductive assurance (Darwin 1876), and also has automatic transmission advantage (Fisher 1941). In obligate outcrossers, self-fertilization is often avoided by postpollination genetic self-incompatibility (SI) (Igic et al. 2008, de Nettancourt 2001, Takayama and Isogai 2005); thus the transition to selfing necessarily involves the loss of the SI system.

Given its evolutionary prevalence, the loss of SI has been a major research focus for plant evolutionary biologists (Igic et al. 2008, Charlesworth and Charlesworth 1979, Uyenoyama et al. 2001, Busch and Schoen 2008, Gervais et al. 2014). SI systems generally consist of genes determining self-recognition specificities and other genes involved in signaling pathways that lead to pollen rejection when pollen and pistils express cognate specificities (de Nettancourt 2001, Takayama and Isogai 2005). The male and female specificity genes are generally tightly linked at a single genetic locus (the *S*-locus) with multiple highly diverged alleles segregating (called *S*-alleles, *S*-haplotypes or *S*-haplogroups, de Nettancourt 2001). In the Brassicaceae, where the SI system has been extensively studied, the *S*-locus receptor kinase (SRK) and the *S*-locus cysteine-rich proteins (SCR; also called *S*-locus protein 11, SP11) determine the female and male specificities, respectively. A few other genes involved in the signaling pathway downstream of *SRK* have also been

Open Access

Article

identified, such as the M-locus protein kinase (MLPK), the ARM-repeat Containing1 (ARC1), and the exocyst subunit Exo70A1, and the glutamate receptor-like channel (GLR) mediating Ca^{2+} influx (Stone et al. 2003, Murase et al. 2004, Samuel et al. 2009, Indriolo et al. 2012, Iwano et al. 2015; but see Kitashiba et al. 2011).

Arabidopsis thaliana has been studied extensively as a model for the loss of SI (Kusaba et al. 2001, Nasrallah et al. 2002, Nasrallah et al. 2004, Sherman-Broyles et al. 2007, Liu et al. 2007, Tang et al. 2007, Shimizu et al. 2008, Tsuchimatsu et al. 2010, Dwyer et al. 2013). *A. thaliana* is self-compatible and predominantly selfing (Abbott and Gomes 1989, Platt et al. 2010, Bomblies et al. 2010), whereas closely related species such as *A. lyrata* and *A. halleri* are mainly obligate outcrossers and possess the ancestral Brassicaceae SI system (Mable et al. 2003, Llaurens et al. 2008). Analysis of the *SRK* gene identified three highly divergent but nonfunctional S-haplogroups in *A. thaliana*: A, B, and C (Shimizu et al. 2008). Functional sequences very similar to each of these three S-haplogroups were found in the outcrossing relatives *A. lyrata* and *A. halleri*, demonstrating that the *A. thaliana* alleles derive from distinct ancestral S-alleles (Bechsgaard et al. 2006).

A major challenge in the study of the loss of SI has been determining the sequence of events, and in particular distinguishing primary inactivating mutations from secondary decay after initial loss of function (Igic et al. 2008, Busch and Schoen 2008, Tsuchimatsu et al. 2010, Boggs et al. 2009). Sequence analysis of the S-locus region identified several gene-disruptive mutations in both *SCR* and *SRK* (Kusaba et al. 2001, Nasrallah et al. 2004, Tang et al. 2007, Shimizu et al. 2008, Tsuchimatsu et al. 2010). When functional copies of these genes from *A. lyrata* were transformed into *A. thaliana*, SI reactions were restored in several accessions (Nasrallah et al. 2002, Liu et al. 2007, Dwyer et al. 2013, Boggs et al. 2009), suggesting that the mutations in the S-locus genes may have been responsible for the loss of SI in *A. thaliana*, although the contribution of other loci such as *ARC1* and *PUB8* remains controversial (Indriolo et al. 2012, Indriolo et al. 2014, Nasrallah and Nasrallah 2014, Goring et al. 2014). One study described a 213-bp inversion of the *SCR* gene in haplogroup A and reported that restoring the ancestral arrangement was sufficient to restore functional SI (Tsuchimatsu et al. 2010). This result suggested that degradation of *SCR* was one of the primary mutations responsible for the loss of SI. However, the inversion is not fixed in the species: it is restricted to accessions carrying haplogroup A (Tsuchimatsu et al. 2010). In parallel, several structural rearrangements of the S-locus have been reported, including interhaplogroup recombination between haplogroups A and C as well as large deletions encompassing S-locus genes (Sherman-Broyles et al. 2007). Although these rearrangements could obviously disrupt the gene functions and therefore may have contributed to the evolutionary loss of SI, they may simply represent secondary decay after the loss of SI. Shimizu et al. (2008) indeed suggested that a ~20-kb deletion occurred after the pseudogenization of *SCR-A* based on the sequence of the deletion border, but the historical order of major

rearrangements, particularly the interhaplogroup recombination, remains largely unknown.

To chart the historical degradation of the S-locus, it is essential to have a detailed view of polymorphism of the whole region from geographically widely distributed samples. However, this is challenging because the S-locus region has extremely high nucleotide diversity, structural rearrangements and repetitive sequences that make traditional sequence analysis impracticable (Guo et al. 2011, Goubet et al. 2012). Previous studies in *A. thaliana* were based on small numbers of short fragments obtained by PCR-based genotyping, Sanger sequencing, or DNA gel blotting for a limited set of accessions (Nasrallah et al. 2004, Sherman-Broyles et al. 2007, Tang et al. 2007, Shimizu et al. 2008, Boggs et al. 2009).

Recently, whole-genome information has become publicly available for more than one thousand worldwide accessions in *A. thaliana* (Cao et al. 2011, Gan et al. 2011, Long et al. 2013, The 1001 Genomes Consortium 2016). We took advantage of these data to explore species-wide polymorphism of the whole S-locus region. We disentangled the complicated mutational history of this region by making use of the raw read sequences from natural accessions. First, as a reference for the short-read mapping, we obtained the full S-locus sequence of haplogroup C, to complement the existing sequences for A and B (Tang et al. 2007, Arabidopsis Genome Initiative 2000; but see also Dwyer et al. 2013) and we functionally characterized the self-incompatibility genes. Second, we mapped the short reads of each accession to the reference sequences of each of the A, B, and C S-haplogroups. Based on these data, we consider how ancestral S-haplotypes have lost functionality independently. *A. thaliana* survived the last glacial age as a set of isolated lineages in separate glacial refugia (The 1001 Genomes Consortium 2016). Although the extent of postglacial recolonization of most lineages occupying distinct refugia remained limited with no secondary contact among them, descendants of a population occupying a single refugium drastically expanded and invaded most of the European continent, admixing with local lineages. We refer to this lineage as the "invasive lineage." Interestingly, several accessions show no evidence of admixture with the invaders, that is, they appear to be "relicts" of their respective glacial refugia (The 1001 Genomes Consortium 2016, Lee et al. 2017) . By including these accessions, we show that selfing is likely to have originated before the last glacial period. We find a high frequency of haplotypes originating from interhaplogroup recombination, and show that these were formed by many independent recombination events in the population that successfully recolonized most of Eurasia.

Results

Polymorphism and Divergence among Copies of Haplogroup C

Only a partial sequence was available for the S-locus of haplogroup C (from accession Lz-0, Dwyer et al. 2013), so we first constructed a bacterial artificial chromosome (BAC) library

and obtained the complete sequence of the S-locus region from the accession Ita-0, which also carries haplogroup C (Shimizu et al. 2008). We recovered an 11-kb sequence fragment with 97% overall identity to the previous data from haplogroup C sequence, including in the intergenic regions (from ATREP20 to ARK3; fig. 1). However, there are major structural differences between the two sequences, including a large 25-kb fragment present only in Ita-0 (from TAT1_ATH to ARNOLDY1c, fig. 1). Notably, Ita-0 also lacks two peculiar features of the S-locus sequence of Lz-0, namely the insertion of a truncated duplicated copy of the flanking gene ARK3 (called ΔARK3) as well as the five fragmented duplicates of SRK described in Dwyer et al. (2013), demonstrating that these structural changes are not fixed within haplogroup C. Overall, this direct comparison demonstrates that several types of structural polymorphisms, including large indels and local rearrangements involving gene fragments, exist within haplogroup C. To polarize the observed polymorphisms, we analyzed the full sequence of the S-locus region for a copy of the functional ortholog of haplogroup C in A. halleri (Ah36), also obtained from a dedicated BAC library (Bechsgaard et al. 2006, Durand et al. 2014). The coding sequence of SCR and SRK showed high conservation between Ah36 and haplogroup C (fig. 2A and B), but most intergenic regions within the S-locus showed too little sequence conservation for reliable polarization of the polymorphisms segregating in A. thaliana (fig. 2A).

Lack of SCR Expression Associated with a Large Deletion in the Promoter in Haplogroup C

Dwyer et al. (2013) showed that SRK is expressed normally in Lz-0 but has inactivating mutations in its coding sequence, whereas the coding sequence of SCR is functional but poorly expressed. We confirmed a similar pattern in Ita-0: we found substantial expression of SRK in stigmatic tissues but no detectable SCR transcripts in anthers (fig. 2C). Based on our complete sequences of the Ita-0 and Ah36 S-locus, we further

observed high conservation of the promoter region of SRK over about 440 bp (fig. 2B), whereas the 5′ flanking region of SCR was extremely diverged from that of the functional AhSCR36 (fig. 2B), possibly accounting for the reduction of SCR expression. SRK amino-acid sequence comparison shows that Ita-0 is an outgroup to all other members of haplogroup C (bootstrap support = 96%, supplementary fig. S1, Supplementary Material online), suggesting that the rearrangements (as well as any other putative inactivating mutations) occurred before the divergence between Ita-0 and other accessions of haplogroup C.

Identification of Diverse Sequence Types by Mapping to Multiple References

We then took advantage of recently published genomic resequencing data in 1,083 A. thaliana accessions from throughout the species range (The 1001 Genomes Consortium 2016) to document the species-wide molecular diversity of the S-locus genomic region. The high density of transposable elements which characterizes the S-locus region (Goubet et al. 2012) prevents accurate de novo assembly of the region. Furthermore because S-locus haplogroups are highly diverged from one another, alignment to the A. thaliana reference genome is impracticable. Hence, we aligned the raw short reads to available references for each of the three haplogroups: haplogroup A from Col-0 (Arabidopsis Genome Initiative 2000), haplogroup B from Cvi-0 (Tang et al. 2007), and haplogroup C from Ita-0 (described above).

We first generated coverage plots against each haplogroup reference for every accession (fig. 3). Strikingly, the vast majority of accessions could be categorized into a small number of groups with similar patterns of quantitative coverage to the three reference sequences (fig. 4). Closer examination of the coverage plots allowed us to identify nine distinctive contiguous fragments of the references (Cov_1, 3, and 4 on haplogroup A and Cov1, 2, 3, 4, 6, and 7 on haplogroup C, supplementary fig. S2, Supplementary Material online) that
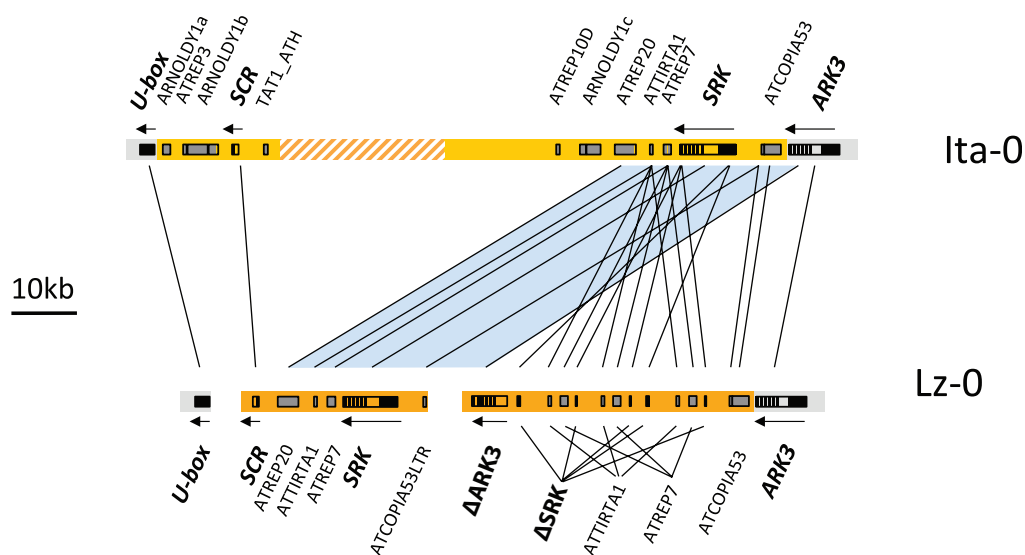

Fig. 1. The S-locus region of haplogroup C from Ita-0 compared with Lz-0. Annotations were performed independently and lines join identical annotations in the two sequences.
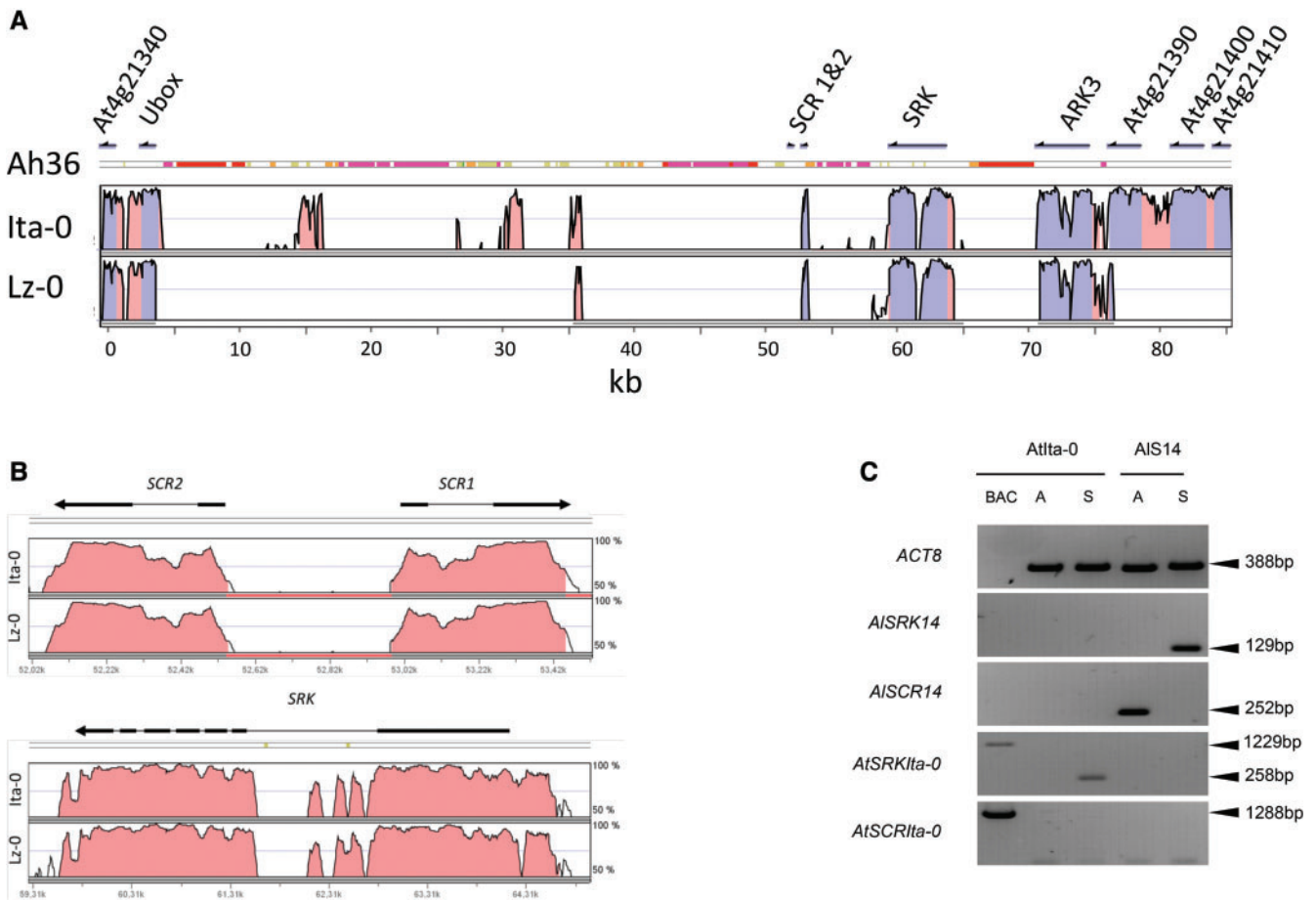
Fig. 2. Polymorphism and divergence within haplogroup C. (*A*) Vista plot comparing the *S*-locus region of Ah36 from *Arabidopsis halleri* (used as a reference) with those of Lz-0 and Ita-0 accessions. The grey bars below the Lz-0 sequence represent the three noncontiguous fragments obtained by Dwyer et al. (2013). (*B*) Zoom on the *SCR* and *SRK* regions. Ah36 is used as a reference. (*C*) Expression analysis of *SCR* and *SRK* in anthers (A) and stigmas (S) of the Ita-0 accession. Transcripts of AlSCR14 and AlSRK14 were readily detected in anthers and stigmas, respectively, of an S14 *A. lyrata* plant. In Ita-0 in contrast, while transcripts of AtSRKIta-0 were detected in stigmas, we did not detect any PCR product for *AtSCRIta-0* even after 35 PCR cycles. BAC DNA was used as a positive PCR control for *AtSCRIta-0* and *AtSRKIta-0* amplification and *ACT8* was used as a loading control.

varied qualitatively among accessions and showed a clear bimodal distribution of coverage (supplementary fig. S3, Supplementary Material online). We scored these fragments as either present or absent, leading us to distinguish a total of twelve sets of accessions, plus one singleton sequence, based on the composition of the *S*-locus (fig. 4, supplementary table S1, Supplementary Material online). These sets included three different groups of "pure A" (A1, A2, A3), one limited group of only two "pure B" accessions and three groups of "pure C" accessions (C1, C2, and C3). In addition, five groups of accessions showed substantial coverage on both A and C references, and presumably correspond to recombinant accessions (see below). Sanger sequencing confirmed that the border of the deletion of the Cov-2 fragment (distinguishing A2 and R2 haplotypes from the rest of A-carrying accessions) was the same as that previously identified in C24 (Sherman-Broyles et al. 2007). A single accession (Db-1) did not appear to match any of the three reference sequences (fig. 4), and may represent either a fourth ancestral haplogroup (present at extremely low frequency), or a rare complete deletion of the *S*-locus specific to this accession.

Geographical Distribution of S-Haplogroups

The data resulting from the classification described above provide a species-level distribution of *S*-locus genotypes (fig. 5). Although previous studies also reported the species-wide frequency and distribution of *S*-haplogroups in *A. thaliana*, here we present the most comprehensive view based on the largest sample size and the whole *S*-locus region. Globally, haplogroup A is common (43.21%), whereas haplogroup C is rare (5.38%), and haplogroup B is extremely rare (two accessions, 0.18%). Surprisingly, the remaining accessions carry recombinant haplotypes (51.5%, see below). Haplogroup A was widely distributed across Europe but was particularly common in the east. In contrast, haplogroup B was only found in the two accessions in which it had already been reported, that is, in the Canary and Cape Verde Islands. Haplogroup C was more common in western Europe, and it always occurred together with haplogroup A and recombinant haplotypes. Altogether, both A and C are found across Europe, with a tendency for higher frequency of A in the east and C in the west (supplementary fig. S4, Supplementary Material online). These findings generally coincide with the
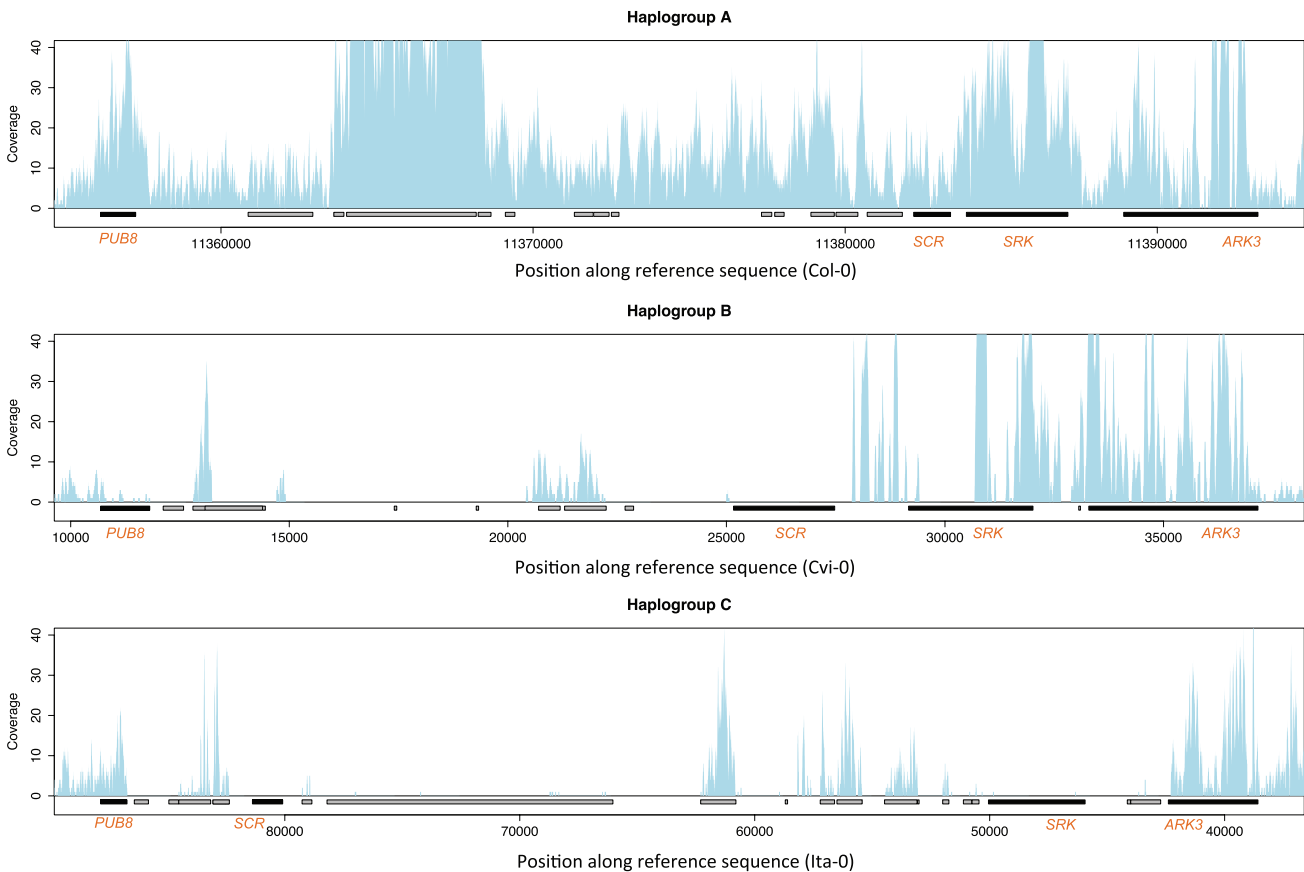
FIG. 3. Example coverage plots for one of the 1083 accessions (accession Moran-1, haplotype A1). Coverage represents the number of sequencing reads for each position of the three reference sequences from haplogroups A, B, and C. Regions between two flanking genes (*ARK3* and *PUB8*) are shown. Transposable elements (TE) and genes are represented by grey and black horizontal bars, respectively. Note that even though this accession is assigned as haplotype A1, some coverage is apparent on the *SRK* gene of the B reference, an artifact due to some sequence similarity between A and B *SRK* alleles (Bechsgaard et al. 2006). Note the important coverage on the TE sequences due to their repetitive nature, but which is not informative with regard to haplotype assignment.

previous reports (Sherman-Broyles et al. 2007; Shimizu et al. 2008; Tsuchimatsu et al. 2010), except the high frequency of recombinants, which was previously estimated to be much lower (23%; Sherman-Broyles et al. 2007; 25,6% in Shimizu et al. 2008).

Genome-wide patterns of divergence identified several groups of "relict" accessions that directly derive from glacial refugia (The 1001 Genomes Consortium, 2016). The analysis using MSMC (Schiffels and Durbin, 2014) revealed that the coalescence rates between relicts and nonrelicts were essentially zero since the last glaciation, indicating that they were isolated from each other during this period (The 1001 Genomes Consortium, 2016). If the breakdown of SI had occurred within these refugia, then *S*-haplotypes would have become differentially fixed locally and the "relict" accessions would be monomorphic, each with a different haplogroup at the *S*-locus. This prediction may hold true for the glacial refugia in offshore Africa where haplogroup B was detected, but since the relict accessions from Iberia contained both haplogroups A and C, it seems likely that the breakdown of SI in *A. thaliana* occurred prior to the last Ice Age, and that the dramatic reduction of the number of *S*-haplogroups (as compared with self-incompatible species)

associated with the breakdown of SI had already occurred when the glacial refugia were established. Importantly, gene-disruptive mutations, including the 213-bp inversion on *SCR-A*, were shared between relicts and nonrelicts accession, further supporting the loss of SI prior to the last Ice Age.

Frequent Recombinant Haplotypes and Their Origins

Recombination between *S*-locus haplotypes is believed to be suppressed because of extreme sequence divergence, and to be selectively disfavored because of functional constraints (Goubet et al. 2012). Thus, the high frequency of interhaplogroup recombinants, with as many as 51.5% of accessions aligning substantially to both references of haplogroups A and C, is very striking.

To understand the origin of the recombinants, we inspected the recombination breakpoints in detail. A single recombinant haplotype sequence had previously been reported (C24; Sherman-Broyles et al. 2007), and an independent recombination event has also been suggested (Kas-2, Boggs et al. 2009). These studies suggested that the C24 recombinant haplotype could have been formed via a partially duplicated flanking gene, *ARK3* (called Δ*ARK3*), which could have acted as the recombination breakpoint
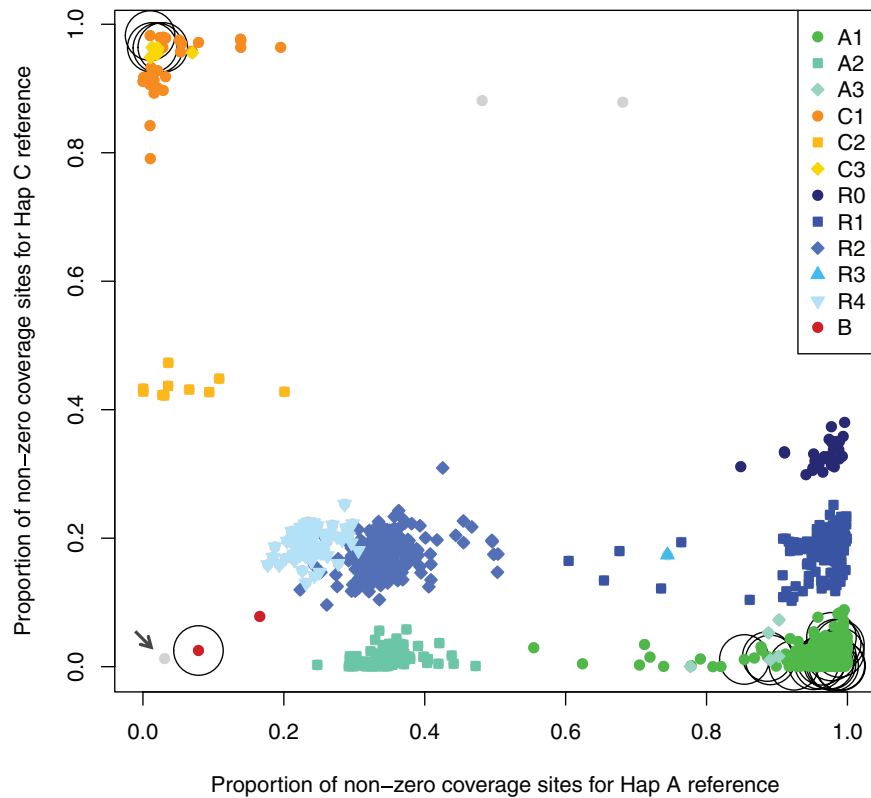
FIG. 4. Proportion of nucleotide sites with nonzero coverage on the A and C references. Accessions are color-coded according to their pattern of presence/absence of the fragments defined in Fig. S2 (see supplementary table S1, Supplementary Material Online). "Relict" accessions are circled with a black circle in the background. Coverage on the B reference was consistently low except for two accessions (Cvi-0 and Can-0) and is not shown here. The Db-1 accession is indicated by an arrow.
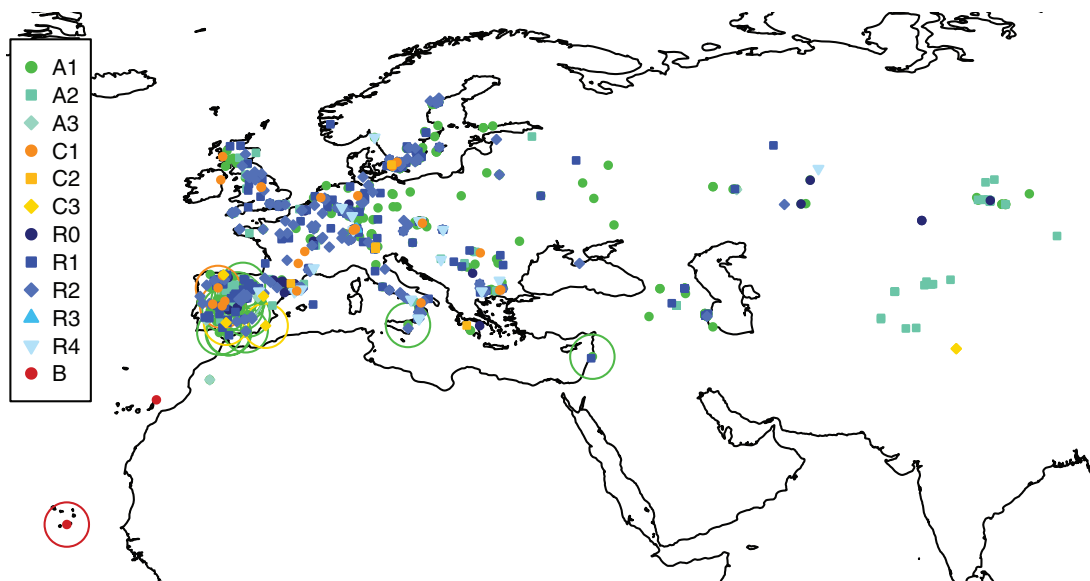

FIG. 5. Geographical distribution of the main haplogroups identified. Each accession is represented by a color-coded dot on the map. "Relict" accessions are circled with a circle in the background, colored according to the *S*-haplotype assignment.

(Sherman-Broyles et al. 2007, Dwyer et al. 2013). Our mapping data are also consistent with the scenario that *ΔARK3* in one parental C haplotype recombined heterologously with the original *ARK3* from an A-haplotype in heterozygous A/C

plants (supplementary fig. S5, Supplementary Material online), as recombinants have substantial coverage that extends up to this point. We used long-range PCR to compare the flanking sequences of *ΔARK3* in a set of eleven R1, R2, R3
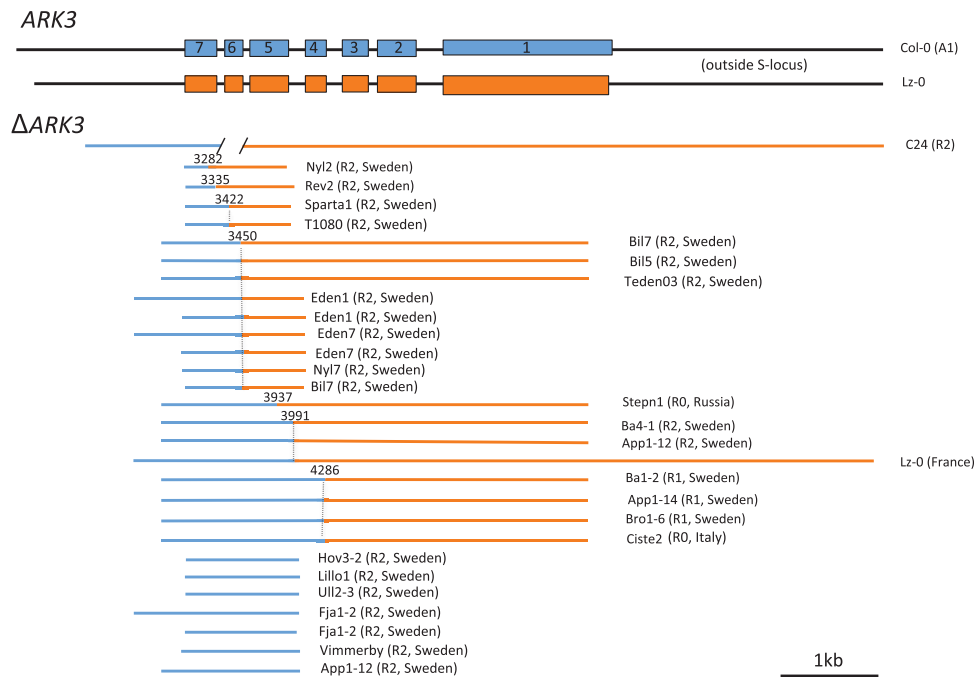
Fig. 6. Sanger resequencing identifies several recombination breakpoints at the ΔARK3 region. Blue and orange colors identify the divergent *ARK3* haplotypes. Some accessions were sequenced from multiple pairs PCR primers and in such case they were represented several times. We note that the deletion encompassing exon 6 of ΔARK3, which was reported previously in the C24 accession, was not found in other accessions, suggesting its uniqueness to C24. We confirmed that it indeed exists in the C24 accession by Sanger sequencing.

or R4 haplotypes (hence spanning across the A–C boundary) and found that they were almost identical to those in C24, confirming that their ΔARK3 were inserted at orthologous positions and therefore derive from a single duplication event.

To identify the precise location of the breakpoints, we performed Sanger sequencing of the recombination breakpoint proposed in C24 (Sherman-Broyles et al. 2007). In line with Tsuchimatsu et al. (2010), we found that the vast majority of the observed polymorphisms could be classified into one of two divergent haplogroups of *ARK3*, which enabled us very precisely to pinpoint the position of the recombination breakpoint, to within a few nucleotides. Strikingly, we found at least seven distinct breakpoints (fig. 6). Although gene conversion from the *ARK3* gene cannot be excluded, these results suggest that the distinct breakpoints represent as many independent recombination events, and that once the ΔARK3 sequence was in place, haplogroups A and C repeatedly recombined.

Lack of Recombinants in Relict Accessions

Despite the overall high frequency of recombinants, we note that not a single of the 22 "relict" accessions carried a recombinant haplogroup, a significant underrepresentation (chi square = 12.65, df = 1, $P = 0.0005$). Hence, recombinants are specific to the accessions derived from the "invasive" lineage that recolonized most of Eurasia and recombination seems to have taken place on the specific genetic background carrying ΔARK3 sequence. It is therefore tempting to speculate that recombination between distinct *S*-haplogroups played a direct role in the loss of SI,

as suggested by Sherman-Broyles et al. (2007). To test this, we examined mapping patterns of the genomic reads from the recombinant haplotypes (R0 and R1) in the region containing the *SCR* sequence. Twenty-six accessions showed no or low coverage in the *SCR-A* region (<10 reads mapped), possibly suggesting a small secondary deletion. All other haplotypes, however, had substantial coverage and yet a low proportion of sequencing reads indicating the presence of an inversion (mean = 0.74% of such reads). Thus, they clearly harbored the 213-bp inversion in *SCR-A*, which was previously suggested to be the mutation responsible for the loss of SI in haplogroup A (Tsuchimatsu et al. 2010), and also share most of the derived mutations that potentially deactivate *SRK*. This strongly suggests that the haplotypes that gave rise to the recombinants were already nonfunctional.

Discussion

Geographic Distribution of *S*-Haplogroups, Post-Glacial Colonization, and Multiple Origins of Self-Compatibility

Although the shift to selfing is common in flowering plants, the scenario under which this occurs varies across species (Vekemans et al. 2014). *Capsella rubella*, for example, has undergone a global fixation of a single nonfunctional *S*-allele associated with speciation (Guo et al. 2009). In contrast, in *A. lyrata* nonfunctional alleles segregate in some populations along with functional alleles, and in *Leavenworthia alabamica* different selfing populations have fixed different

nonfunctional alleles, so that no species-wide fixation of the nonfunctional alleles is observed in either species (Mable et al. 2005, 2017; Busch et al 2011).

In *A. thaliana*, we find yet another pattern, combining species-wide loss of SI with the maintenance of at least three nonfunctional S-haplogroups (given the large sample analyzed here, from a broadly distributed set of populations, it is unlikely that additional haplotypes, if they exist, have substantial frequencies). Given that the number of S-haplotypes that segregated in the ancestral outcrossing species was probably very large, this entails a major reduction in the number of SI alleles (fig. 7). We note that *Arabidopsis kamchatica* also represents an example of maintenance of multiple nonfunctional S-haplogroups despite the species-wide loss of SI, but its S-haplogroup distribution is suggested to be associated with multiple origins of the species by allopolyploidization events (Tsuchimatsu et al. 2012), which would therefore not be directly comparable to the case of *A. thaliana*. The existence of more than one nonfunctional haplotype in *A. thaliana* is puzzling, and raises the question why the first disrupted haplotype did not become fixed species-wide if the conditions for invasion were met. A first possibility is that different selfing genotypes became fixed in separate glacial refugia. This might be true for haplotype B, which we found to be restricted to offshore Africa. However, the widespread geographical distribution of both haplogroups A and C shows that their putative fixation did not occur within specific refugia from the last glaciation. We even found that both haplotypes segregate among relict accessions from the single Iberian refugium, demonstrating that they were already cooccurring when the refugium was formed (fig. 7). Hence, local fixations of alternative S-haplotypes in refugia, if they happened, are likely to have taken place during an earlier glacial cycle, followed by secondary admixture during a previous interglacial period.

A second possibility is that the species-wide fixation process was slow, providing enough time for the emergence of several nonfunctional haplotypes. Theory predicts that nonfunctional haplotypes may stably segregate in a self-incompatible population along with functional haplotypes, specifically when inbreeding depression is high and the proportion of self-pollen deposited on pistils is low to intermediate (Charlesworth and Charlesworth 1979; Uyenoyama et al. 2001; Gervais et al. 2011). If so, multiple nonfunctional haplotypes could have sequentially appeared and been maintained within the species. Because all three haplotypes carry inactivating mutations, this polymorphism could have been effectively neutral and may be on its way to slow fixation by genetic drift. Population structure, which is currently strong in *A. thaliana* (Nordborg et al. 2005), would further slow down the species-wide fixation of a particular nonfunctional haplotype. In addition, the discovery of a fixed deletion in *A. thaliana* of a putative gene of the signaling cascade of SI (Indriolo et al. 2012, but see Nasrallah & Nasrallah 2014) suggests the possibility that the ancestral *A. thaliana* lineage experienced partial self-compatibility. This could have contributed to the reduction of the selective forces promoting nonfunctional S-haplotypes, hence slowing down their fixation.

Extensive Recombination at the S-Locus in a Selfing Species

Recombination between functional SI haplotypes is apparently suppressed in outcrossing species, but it is currently unclear whether recombination suppression is mainly due to selective constraints (to preserve linkage of cognate functional elements within the S-locus) or to the major structural polymorphism preventing proper chromosomal pairing during meiosis (Castric et al. 2010). The overall frequency of recombinant haplotypes we observed is much higher than previous estimates (22.86%, Sherman-Broyles et al. 2007), possibly because previous genotyping was based on detecting small gene fragments that we found to have been subsequently deleted in specific haplotypes (supplementary fig. S2, Supplementary Material online). With the whole S-locus sequence, we have much greater power to detect recombination events even after subsequent deletions have occurred.

Our data show that in self-compatible *A. thaliana*, where functional constraints are absent, recombination has occurred repeatedly between a pair of highly diverged haplogroups. Hence, the extreme level of structural polymorphism among haplogroups does not seem to prevent recombination. We note, however, that the signatures of interhaplogroup recombination indicate that it is caused by unequal crossing-over involving a duplicated copy of the flanking gene *ARK3* (supplementary fig. S5, Supplementary Material online). Interestingly, *ARK3* seems to be prone to duplications even in functional SI haplotypes, as distinct ΔARK3-like sequences have been found in multiple haplotypes in *A. lyrata* and *A. halleri* (Guo et al. 2011, Goubet et al. 2012), and a polymorphism survey at *ARK3* in *A. lyrata* has reported the frequent occurrence of multiple copies (Hagenblad et al. 2006), a property that could facilitate rearrangements of the S-locus. Although *A. thaliana* is a predominant selfer, outcrossing rate is still significantly higher than zero (~3%; Platt et al. 2010). Occasional outcrossing events would provide chances for recombination, albeit rare. Indeed, genome-wide linkage disequilibrium is known to decay relatively quickly (e.g., Kim et al. 2007).

Structural Rearrangements as Secondary Decay

An important challenge in the studies on the loss of SI has been to distinguish primary inactivating mutations causing loss of SI from subsequent decay by further gene-disruptive mutations and structural rearrangements (Igic et al. 2008, Busch and Schoen 2008, Tsuchimatsu et al. 2010, Boggs et al. 2009). Sherman-Broyles et al. (2007), who first reported the extensive structural rearrangements at the S-locus region of C24, proposed two scenarios for the loss of SI in the recombinant haplotypes: 1) recombination between the haplogroups A and C in a heterozygous self-incompatible individual causing loss of SI, followed by subsequent restructuring, or 2) inactivation of the haplogroups A and C by independent rearrangements and deletions followed by recombination in a self-fertile heterozygous individual. We found that all recombining haplotypes containing the *SCR-A* region had the 213-bp inversion in *SCR-A*, which was reported to be responsible for the loss of SI within haplogroup A (Tsuchimatsu et al.
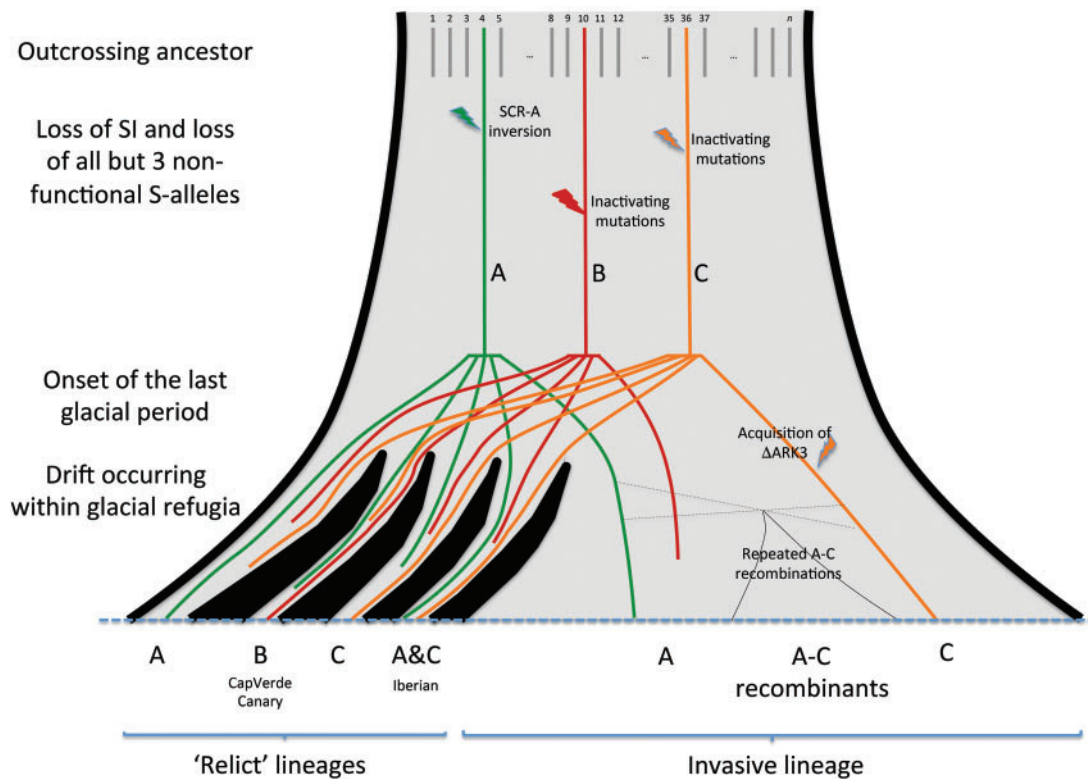
Fig. 7. Scenario for the multiple loss of self-incompatibility, subsequent decay of the *S*-locus and distribution of *S*-haplotypes in different glacial refugia. The number of functional *S*-haplotypes was initially very large in the outcrossing ancestor, whereas only three *S*-haplogroups currently segregate, each containing specific inactivating mutations. As glacial refugia became isolated, genetic drift led to differential fixation of *S*-haplogroups in the different refugia, whereas others remained polymorphic (as in the Iberian refugium). The lineage from the refugia that became invasive thoughout Eurasia specifically acquired a local duplication of the flanking gene *ARK3* on haplogroup C that allowed for repeated recombination events among copies of the different *S*-haplogroups that were locally segregating.

2010). This result suggests that pseudogenization of *SCR-A* preceded the recombination events as well as the large deletions, consistent with the previous report that the large deletion of C24 represents secondary decay (Shimizu et al. 2008). In addition, our results suggest that the parental haplotype carrying haplogroup C was also nonfunctional due to the loss of the *SCR* promoter region, such that we have high confidence that the SI system of the A/C heterozygote was already deactivated. Finally, the fact that none of the relict accessions had any of the large-scale rearrangements we identified, strongly suggests that those rearrangements happened in the very recent past, possibly postglacially, and were specific to the invasive "weedy" lineage that recolonized most of Eurasia (fig. 7). At any rate, it is now clear that the independent recombination events and their associated structural rearrangements represent truly secondary decay that occurred after the loss of SI of both parental haplotypes.

Materials and Methods

Isolation of a Genomic Fragment Containing the *A. thaliana* C Haplogroup
Isolation of a genomic fragment containing the *A. thaliana* C haplogroup was performed according to Goubet et al. (2012). Briefly, high molecular weight DNA was prepared from young leaves of the Ita-0 accession and a BAC library was constructed.

The library was screened on nylon filters using radiolabelled probes designed from the flanking genes *ARK3* and *PUB8*. Positive BAC clones detected by hybridization were validated by PCR amplification using the primer pairs used for probes synthesis, and visualisation of PCR products after agarose gel electrophoresis. The BAC clone was sequenced at Genoscope using a 454 multiplexing technology on the Titanium sequencer version (www.roche.com). *De novo* assembly was performed by Newbler (www.roche.com) and the sequence was obtained in five contigs. Only contigs representing the extremities of the BAC were initially oriented. Long-range PCR was then employed to order and orientate the rest of the contigs (supplementary table S2, Supplementary Material Online). Genes and transposable elements annotation was performed according to Goubet et al. (2012). Because the Ah36 sequence published in Durand et al. (2014, Genbank accession number KM592804-KM592809) contains an unusual duplicated copy of *SCR*, we resequenced this BAC clone using the PACBIO platform and confirmed this structural feature. The two copies are nearly identical and differ only by one nonsynonymous and one synonymous nucleotide differences in exon 1 and 2, respectively.

Expression Analysis
Total RNAs were extracted from stigma (S) and anthers (A) (stage 12 according to Smyth et al. 1990) of *A. lyrata* *S*14

haplotype and *A. thaliana* Ita-0 plants following the manufacturer's recommendations (Picopure kit; Arcturus) except that a second DNase treatment (RNase-free Dnase, Qiagen) was added. The complementary DNA (cDNA) was synthesized with the Revert Aid M-MuLV reverse transcription (Fermentas) using $(N)_6$ random primers and 260 ng of total RNAs. Reverse transcription (RT)-PCR with 25 cycles for *SRK* amplification or 35 cycles for *SCR* amplification was performed with primer pairs located within the exon 1 for *SRK14*, surrounding the intron 1 for *SRKIta-0* and on the ATG and the stop codons for *SCR14 and SCRIta-0*. The constitutively expressed gene *ACT8* was used as a control to normalize loading of RT-PCR products. Amplification using primers located within a nonexpressed region (intergenic region between At1g49230 and At1g49240 (*ACT8*)) confirmed the absence of DNA contamination (data not shown). Primer pairs for *SRKIta-0* and *SCRIta-0* were tested using, as template, BAC DNA (BAC) containing the S-locus of Ita-0. The PCR products amplified correspond to the genomic fragments including introns (1,229 and 1,288 bp, respectively). PCRs were performed with the GoTaq® enzyme (Promega) at 50 °C for the hybridization step. Primer sequences used are listed in supplementary table S2, Supplementary Material Online.

Analysis of the S-Locus Region in Accessions from the 1,001 Genomes Project

We used the paired-end reads from whole genome sequencing of 1,083 accessions from the 1,001 genomes project. Briefly, each accession was sequenced using Illumina sequencing with paired-end reads of 76 bp. Because of the large sequence divergence among haplotypes, the S-locus can typically not be assembled using standard procedures (i.e., mapping on a single genomic reference, Col-0). Hence, for each of these accessions, three new alignments of the S-locus were performed using the Burrows-Wheeler Aligner (Li and Durbin 2009), taking in turn each known sequence (Col-0 for haplogroup A, Cvi-0 for haplogroup B, Ita-0 for haplogroup C) as a reference.

We then plotted for each accession the proportion of nucleotide positions showing nonzero coverage (>1 reads) on references of haplogroups A, B, and C. We masked transposable elements (TE) for this analysis, as the extremely high coverage on TEs would be mainly due to nonspecific reads and thus confound the results. This analysis revealed a discrete set of coverage patterns, suggesting a limited number of S-locus types among accessions. Based on a representative subset of accessions for each main type, we then searched manually for fragments of the reference sequences that distinguished among these main types. Although most fragments varied in a highly coordinated manner, a 12-kb fragment of Ita-0 (represented by a dashed motif on fig. 1 and supplementary fig. S2, Supplementary Material online) was not specific to accessions defined as haplotype C or R by all other fragments. This fragment also varied in a presence/absence manner (data not shown) and had substantial coverage in about 45% of all accessions but we could not confirm contiguity between this fragment and its flanking sequences

on the C-reference by any paired-end sequencing read straddling these positions. To further examine the origin of this fragment, we extracted paired end reads for which one read mapped at one extremity of the fragment and the other was unmapped on the references. We then blasted the unmapped reads on the complete *A. thaliana* genome and found for the two extremities a strong homology with one particular position on chromosome 2 (around position 1,825,600). This observation suggests that this 12-kb fragment segregates on chromosome 2 of 45% of all accessions and has been specifically inserted in the S-locus of the Ita-0 accession. Consequently, we decided to ignore its pattern of presence/absence in the analyzed accessions. Overall, this analysis allowed us to distinguish a total of 12 different types that varied by their pattern of presence/absence of the different fragments. We used this set of fragments to automatically classify the rest of the accessions. A fragment was considered to be present if the proportion of sites with nonzero coverage was more than the threshold value defined for each fragment, and otherwise considered to be absent.

Based on the distribution of the median pairwise distance to all other accessions computed from all SNPs (The 1001 Genomes Consortium 2016), we chose to conservatively consider as relict accessions those with a median distance above 0.00356, corresponding to a gap in the distribution. Application of this strict threshold resulted in 22 relict accessions. According to The 1001 Genomes Consortium (2016), the majority of these relicts were from the Iberian peninsula ($n = 19$) and are considered as descendants from a single glacial refugium, with one accession each representing relicts from Sicily ($n = 1$), Lebanon ($n = 1$), and the Cape-Verde island ($n = 1$).

To investigate the occurrence of the 213-bp inversion of *SCR-A* in the accessions having A1, R0, and R1, we looked at coverage in the inverted region (Chr4:11383163–11383376). We considered that the region was present when at least ten reads were mapped in this interval. We then specifically looked for sequencing reads whose mates were also mapped but in the same orientation (hence possibly indicating the presence of a local inversion as compared with the reference). We required that at least 25% of the paired reads indicate the presence of an inversion.

A Neighbor-Joining phylogeny of the amino-acid sequence of the extracellular domain of the *SRK* gene for accessions carrying haplogroup C was obtained in MEGA7 (Kumar et al. 2016). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown above the branches.

Targeted Resequencing across the Recombination Breakpoint

We obtained sequences around the recombination breakpoints and the deletion border by Sanger sequencing. DNA was extracted by Macherey-Nagel plant DNA kit. GoTaq (Promega) was used for PCR. PCR products were purified by using Agencourt AMPure XP (Beckman Coulter) and sequenced by Applied Biosystems 3130xL capillary sequencer. For primers used, see supplementary table S2, Supplementary Material Online. Sequences were inspected by eye, assembled

and aligned by using the CLC Main Workbench. For long-range PCR fragments, internal primers were designed to tile the fragments entirely.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

References

Abbott RJ, Gomes MF. 1989. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* 62:411–418.

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

Barrett SCH. 2002. The evolution of plant sexual diversity. *Nat Rev Genet.* 3:274–284.

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. *Mol Biol Evol.* 23:1741–1750.

Boggs NA, Nasrallah JB, Nasrallah ME. 2009. Independent *S*-locus mutations caused self-fertility in *Arabidopsis thaliana*. *PLoS Genet.* 5:e1000426.

Bomblies K, Yant L, Laitinen RA, Kim S-T, Hollister JD, et al. 2010. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genet.* 6:e1000890.

Busch JW, Joly S, Schoen DJ. 2011. Demographic signatures accompanying the evolution of selfing in *Leavenworthia alabamica*. *Mol Biol Evol.* 28:1717–1729.

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

Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, et al. 2011. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet.* 43:956–963.

Castric V, Bechsgaard JS, Grenier S, Noureddine R, Schierup MH, et al. 2010. Molecular evolution within and between self-incompatibility specificities. *Mol Biol Evol.* 27:11–20.

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

Darwin C. (1876) The effects of cross and self fertilisation in the vegetable kingdom. London: J. Murray. p. 487.

de Nettancourt D. (2001) Incompatibility and incongruity in wild and cultivated plants. 2nd ed. Berlin: Springer. p. 322.

Durand E, Méheust R, Soucaze M, Goubet PM, Gallina S, et al. 2014. Dominance hierarchy arising from the evolution of a complex small RNA regulatory network. *Science* 346:1200–1205.

Dwyer KG, Berger MT, Ahmed R, Hritzo MK, McCulloch AA, et al. 2013. Molecular characterization and evolution of self-incompatibility genes in *Arabidopsis thaliana*: the case of the *Sc* haplotype. *Genetics* 193:985–994.

Fisher RA. 1941. Average excess and average effect of a gene substitution. *Ann Eugen.* 11:53–63.

Gan X, Stegle O, Behr J, Steffen JG, Drewe P, et al. 2011. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477:419–423.

Gervais CE, Castric V, Ressayre A, Billiard S. 2011. Origin and diversification dynamics of self-incompatibility haplotypes. *Genetics* 188(3): 625–636.

Gervais C, Awad DA, Roze D, Castric V, Billiard S. 2014. Genetic architecture of inbreeding depression and the maintenance of gametophytic self-incompatibility. *Evolution* 68:3317–3324.

Goring DR, Indriolo E, Samuel MA. 2014. The *ARC1* E3 ligase promotes a strong and stable self-incompatibility response in *Arabidopsis* species: response to the Nasrallah and Nasrallah commentary. *Plant Cell* 26:3842–3846.

Goubet PM, Bergès H, Bellec A, Prat E, Helmstetter N, et al. 2012. Contrasted patterns of molecular evolution in dominant and recessive self-incompatibility haplotypes in *Arabidopsis*. *PLoS Genet.* 8:e1002495.

Guo YL, Bechsgaard JS, Slotte T, Neuffer B, Lascoux M, Weigel D, Schierup MH, 2009. 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(13):5246–5251.

Guo YL, Zhao X, Lanz C, Weigel D. 2011. Evolution of the *S*-locus region in *Arabidopsis* relatives. *Plant Physiol.* 157:937–946.

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

Igic B, Lande R, Kohn JR. 2008. Loss of self-incompatibility and its evolutionary consequences. *Int J Plant Sci.* 169:93–104.

Indriolo E, Safavian D, Goring DR. 2014. The *ARC1* E3 ligase promotes two different self-pollen avoidance traits in *Arabidopsis*. *Plant Cell* 26:1525–1543.

Indriolo E, Tharmapalan P, Wright SI, Goring DR. 2012. The *ARC1* E3 ligase gene is frequently deleted in self-compatible Brassicaceae species and has a conserved role in *Arabidopsis lyrata* self-pollen rejection. *Plant Cell* 24:4607–4620.

Iwano M, Ito K, Fujii S, Kakita M, Asano-Shimosato H, Igarashi M, Kaothien-Nakayama P, Entani T, Kanatani A, Takehisa M, et al. 2015. Calcium signalling mediates self-incompatibility response in the Brassicaceae. *Nat Plants* 11:15128.

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M. 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet.* 39(9): 1151–1155.

Kitashiba H, Liu P, Nishio T, Nasrallah JB, Nasrallah ME. 2011. Functional test of *Brassica* self-incompatibility modifiers in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A.* 108:18173–18178.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 33 (7): 1870–1874.

Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, 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.

Lee CR, Svardal H, Farlow A, Exposito-Alonso M, Ding W, Novikova P, Alonso-Blanco C, Weigel D, Nordborg M 2017. On the post-glacial spread of human commensal *Arabidopsis thaliana*. *Nat Communications* 8:14458.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760.

Liu P, Sherman-Broyles S, Nasrallah ME, Nasrallah JB. 2007. A cryptic modifier causing transient self-incompatibility in *Arabidopsis thaliana*. *Curr Biol.* 17:734–740.

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

Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, et al. 2013. Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat Genet.* 45:884–890.

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

Mable B, Robertson A, 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.

Mable BK, Hagmann J, Kim ST, Adam A, Kilbride E, Weigel D, Stift M. 2017. What causes mating system shifts in plants? *Arabidopsis lyrata* as a case study. *Heredity* 118:52–63.

Murase K, Shiba H, Iwano M, Che F-S, Watanabe M, et al. 2004. A membrane-anchored protein kinase involved in *Brassica* self-incompatibility signaling. *Science* 303:1516–1519.

Nasrallah JB, Nasrallah ME. 2014. Robust self-incompatibility in the absence of a functional *ARC1* gene in *Arabidopsis thaliana*. *Plant Cell* 26:3838–3841.

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. *Proc Natl Acad Sci U S A.* 101:16070–16074.

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, et al. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* 3:e196.

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, et al. 2010. The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet.* 6:e1000843.

Samuel MA, Chong YT, Haasen KE, Aldea-Brydges MG, Stone SL, et al. 2009. Cellular pathways regulating responses to compatible and self-incompatible pollen in *Brassica* and *Arabidopsis* stigmas intersect at Exo70A1, a putative component of the exocyst complex. *Plant Cell* 21:2655–2671.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nature Genetics.* 46:919–925.

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

Shimizu KK, Shimizu-Inatsugi R, Tsuchimatsu T, Purugganan MD. 2008. Independent origins of self-compatibility in *Arabidopsis thaliana*. *Mol Ecol.* 17:704–714.

Shimizu KK, Tsuchimatsu T. 2015. Evolution of selfing: recurrent patterns in molecular adaptation. *Annu Rev Ecol Evol Syst.* 46:593–622.

Smyth DR, Bowman JL, Meyerowitz EM. 1990. Early flower development in *Arabidopsis*. *Plant Cell* 2:755–767.

Stebbins GL. (1974) Flowering plants: evolution above the species level. Cambridge: Belknap Press. p. 399.

Stone S, Anderson E, Mullen R, Goring D. 2003. ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible *Brassica* pollen. *Plant Cell* 15:885–898.

Takayama S, Isogai A. 2005. Self-incompatibility in plants. *Annu Rev Plant Biol.* 56:467–489.

Tang C, Toomajian C, Sherman-Broyles S, Plagnol V, Guo YL, et al. 2007. The evolution of selfing in *Arabidopsis thaliana*. *Science* 317:1070–1072.

The 1001 Genomes Consortium 2016. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166:481–491.

Tsuchimatsu T, Suwabe K, Shimizu-Inatsugi R, Isokawa S, Pavlidis P, et al. 2010. Evolution of self-compatibility in *Arabidopsis* by a mutation in the male specificity gene. *Nature* 464:1342–1346.

Tsuchimatsu T, Kaiser P, Yew C-L, Bachelier JB, Shimizu KK. 2012. Recent loss of self-incompatibility by degradation of the male component in allotetraploid *Arabidopsis kamchatica*. *PLoS Genet.* 8(7): e1002838. doi:10.1371/journal.pgen.1002838

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

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