

Transposable elements cause the loss of self-incompatibility in citrus

Hu, Jianbing; Liu, Chenchen; Du, Zezhen; Guo, Furong; Song, Dan; Wang, Nan; Wei, Zhuangmin; Jiang, Jingdong; Cao, Zonghong; Shi, Chunmei; Zhang, Siqi; Zhu, Chenqiao; Chen, Peng; Larkin, Robert m.; Lin, Zongcheng; Xu, Qiang; Ye, Junli; Deng, Xiuxin; Bosch, Maurice; Franklin-tong, Veronica e.

DOI:
[10.1111/pbi.14250](https://doi.org/10.1111/pbi.14250)

License:
Creative Commons: Attribution (CC BY)

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Hu, J, Liu, C, Du, Z, Guo, F, Song, D, Wang, N, Wei, Z, Jiang, J, Cao, Z, Shi, C, Zhang, S, Zhu, C, Chen, P, Larkin, RM, Lin, Z, Xu, Q, Ye, J, Deng, X, Bosch, M, Franklin-tong, VE & Chai, L 2023, 'Transposable elements cause the loss of self-incompatibility in citrus', *Plant Biotechnology Journal*. <https://doi.org/10.1111/pbi.14250>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.






When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Transposable elements cause the loss of self-incompatibility in citrus

Jianbing Hu^{1,2} , Chenchen Liu^{1,2}, Zezhen Du^{1,2}, Furong Guo¹, Dan Song¹, Nan Wang¹, Zhuangmin Wei³, Jingdong Jiang¹, Zonghong Cao¹, Chunmei Shi¹, Siqi Zhang¹, Chenqiao Zhu¹, Peng Chen⁴, Robert M. Larkin^{1,2}, Zongcheng Lin^{1,2} , Qiang Xu^{1,2} , Junli Ye¹, Xiuxin Deng^{1,2} , Maurice Bosch⁵ , Veronica E. Franklin-Tong⁶ and Lijun Chai^{1,2,*}

¹National Key Laboratory for Germplasm Innovation and Utilization of Horticultural Crops, College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan, P. R. China

²Hubei Hongshan Laboratory, Wuhan, P. R. China

³Guangxi Subtropical Crops Research Institute, Nanning, P. R. China

⁴Horticultural Institute, Hunan Academy of Agricultural Sciences, Changsha, China

⁵Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Aberystwyth, UK

⁶School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham, UK

Received 10 September 2023;

revised 25 October 2023;

accepted 11 November 2023.

*Correspondence (Tel: +8613006338843;

Fax: 027 87282110; email chailijun@mail.hzau.edu.cn

hzau.edu.cn

Summary

Self-incompatibility (SI) is a widespread prezygotic mechanism for flowering plants to avoid inbreeding depression and promote genetic diversity. Citrus has an *S*-RNase-based SI system, which was frequently lost during evolution. We previously identified a single nucleotide mutation in *S_m-RNase*, which is responsible for the loss of SI in mandarin and its hybrids. However, little is known about other mechanisms responsible for conversion of SI to self-compatibility (SC) and we identify a completely different mechanism widely utilized by citrus. Here, we found a 786-bp miniature inverted-repeat transposable element (MITE) insertion in the promoter region of the *FhiS₂-RNase* in *Fortunella hindsii* Swingle (a model plant for citrus gene function), which does not contain the *S_m-RNase* allele but are still SC. We demonstrate that this MITE plays a pivotal role in the loss of SI in citrus, providing evidence that this MITE insertion prevents expression of the *S-RNase*; moreover, transgenic experiments show that deletion of this 786-bp MITE insertion recovers the expression of *FhiS₂-RNase* and restores SI. This study identifies the first evidence for a role for MITEs at the *S*-locus affecting the SI phenotype. A family-wide survey of the *S*-locus revealed that MITE insertions occur frequently adjacent to *S-RNase* alleles in different citrus genera, but only certain MITEs appear to be responsible for the loss of SI. Our study provides evidence that insertion of MITEs into a promoter region can alter a breeding strategy and suggests that this phenomenon may be broadly responsible for SC in species with the *S-RNase* system.

Keywords: Self-incompatibility, citrus, *S*-RNase, *S*-locus, MITE, evolution.

Introduction

Transposable elements (TEs) are the most variable components of the genome, which can be replicated and integrated into positions and mediate the genetic diversity and evolution in plants by affecting the expression levels of adjacent genes (Bennetzen and Wang, 2014; Lisch, 2013; Serrato-Capuchina and Matute, 2018). Miniature inverted-repeat transposable elements (MITEs) are considered to be deletion derivatives of DNA transposons, which have many copies in the genome and tend to be inserted in or near genes. They play an important role in the regulation of genes through various mechanisms (Chen *et al.*, 2014; Feschotte *et al.*, 2002; Lu *et al.*, 2012). Previous studies have shown that TEs are crucial for plant reproduction (Wang *et al.*, 2017; Xia *et al.*, 2017), development (Xu *et al.*, 2020), disease or stress resistance (Mao *et al.*, 2015; Wu

et al., 2022; Zhang *et al.*, 2016) and domestication (Li *et al.*, 2018; Lisch, 2013; Niu *et al.*, 2019).

Over 40% of flowering plant species are self-incompatible (SI) and thus, can produce fertile seeds after outcrossing but cannot produce zygotes after self-pollination (Igic *et al.*, 2008). SI systems can be divided into sporophytic self-incompatibility (SSI) and gametophytic self-incompatibility (GSI) by genetic mode of action. In SSI, the recognition specificity is determined by the genotype of the pollen-producing plant (the sporophyte). In GSI, the *S*-specificity is controlled by the genotype carried by the haploid gamete (i.e. the pollen) (Fujii *et al.*, 2016; Takayama and Isogai, 2005). The *S-RNase*-based GSI is the most widely distributed system (Zhao *et al.*, 2022). At present, six eudicot families (Solanaceae, Plantaginaceae, Rosaceae, Rubiaceae, Rutaceae and Cactaceae) have been demonstrated to use the *S-RNase*-based SI system (Du *et al.*, 2021; Igic and Kohn, 2001;

Please cite this article as: Hu, J., Liu, C., Du, Z., Guo, F., Song, D., Wang, N., Wei, Z., Jiang, J., Cao, Z., Shi, C., Zhang, S., Zhu, C., Chen, P., Larkin, R.M., Lin, Z., Xu, Q., Ye, J., Deng, X., Bosch, M., Franklin-Tong, V.E. and Chai, L. (2023) Transposable elements cause the loss of self-incompatibility in citrus. *Plant Biotechnol. J.*, <https://doi.org/10.1111/pbi.14250>.

Liang *et al.*, 2020; Ramanauskas and Igc, 2021; Sassa *et al.*, 2007; Sijacic *et al.*, 2004; Ushijima *et al.*, 2003; Wang *et al.*, 2003). This has an *S*-locus composed of highly polymorphic, tightly linked pistil and pollen *S* determinants comprising the *S*-RNase and multiple *S*-locus *F*-box (*SLF*) genes, respectively (Fujii *et al.*, 2016; Takayama and Isogai, 2005; Wang *et al.*, 2003). The *S*-RNase ribonuclease activity is critical for self-pollen rejection (Huang *et al.*, 1994) and RNase activity in style tissue is positively correlated with ability to reject self-pollen (Murfett *et al.*, 1994); the *S*-RNases enter pollen tubes indiscriminately and if not removed, inhibit pollen tube growth. Using a 'non-self' recognition system, groups of *SLFs* recognize and detoxify several *S*-RNases (Fujii *et al.*, 2016).

It is generally acknowledged that SI, by preventing self-fertilization, maintains genetic diversity (Goldberg *et al.*, 2010). The shift from obligatory outcrossing to predominant self-fertilization occurred repeatedly in convergent evolution (Barrett, 2002). Selfing is favoured for its inherent propagative and reproductive advantages in the absence of mates, pollinators or both. When the advantages of selfing outweigh the costs of inbreeding depression (e.g. reduced effective recombination rates, expansion of population subdivision, genetic bottlenecks, reductions in effective population size and selective interference between linkage sites), selfing evolves (Charlesworth and Willis, 2009; Charlesworth and Wright, 2001; Goldberg *et al.*, 2010; Igc *et al.*, 2008; Slotte *et al.*, 2013). The evolution of selfing is usually accompanied by the loss of SI and evidence suggests that the early origin and ancestral status of the *S*-locus and *S*-RNase-based SI in dicotyledonous plants indicates that SI was frequently lost during speciation (Fujii *et al.*, 2016; Zhao *et al.*, 2022).

Citrus is a tropical and subtropical perennial fruit tree from the Aurantioideae subfamily from the Rutaceae family that is widely cultivated worldwide. Many of the main cultivars are SI (Caruso *et al.*, 2012; Chai *et al.*, 2010; Gambetta *et al.*, 2013; Honsho *et al.*, 2009; Kakade *et al.*, 2017; Yamashita and Tanimoto, 1985; Ye *et al.*, 2009; Zhang *et al.*, 2015), which substantially restricts yield and quality improvement efforts for some cultivars. The Aurantioideae consists of three major genera (*Citrus*, *Fortunella* and *Poncirus*) with distinct reproductive mechanisms (e.g. apomixis, outcrossing and self-fertilization) (Shimada *et al.*, 2018; Wang *et al.*, 2017, 2022; Wu *et al.*, 2018, 2021). We previously showed that citrus uses the *S*-RNase-based SI system and identified a dysfunctional mutant allele, *S_m-RNase*, which comprises a single nucleotide mutation in the *S*-RNase. Its fixation and propagation in *Citrus* genus led to the loss of SI (Liang *et al.*, 2020). However, several genera that are SC, such as

Fortunella and *Poncirus* genera do not contain the *S_m-RNase* allele, implicating that there are other mechanisms responsible for the loss of SI in the Aurantioideae.

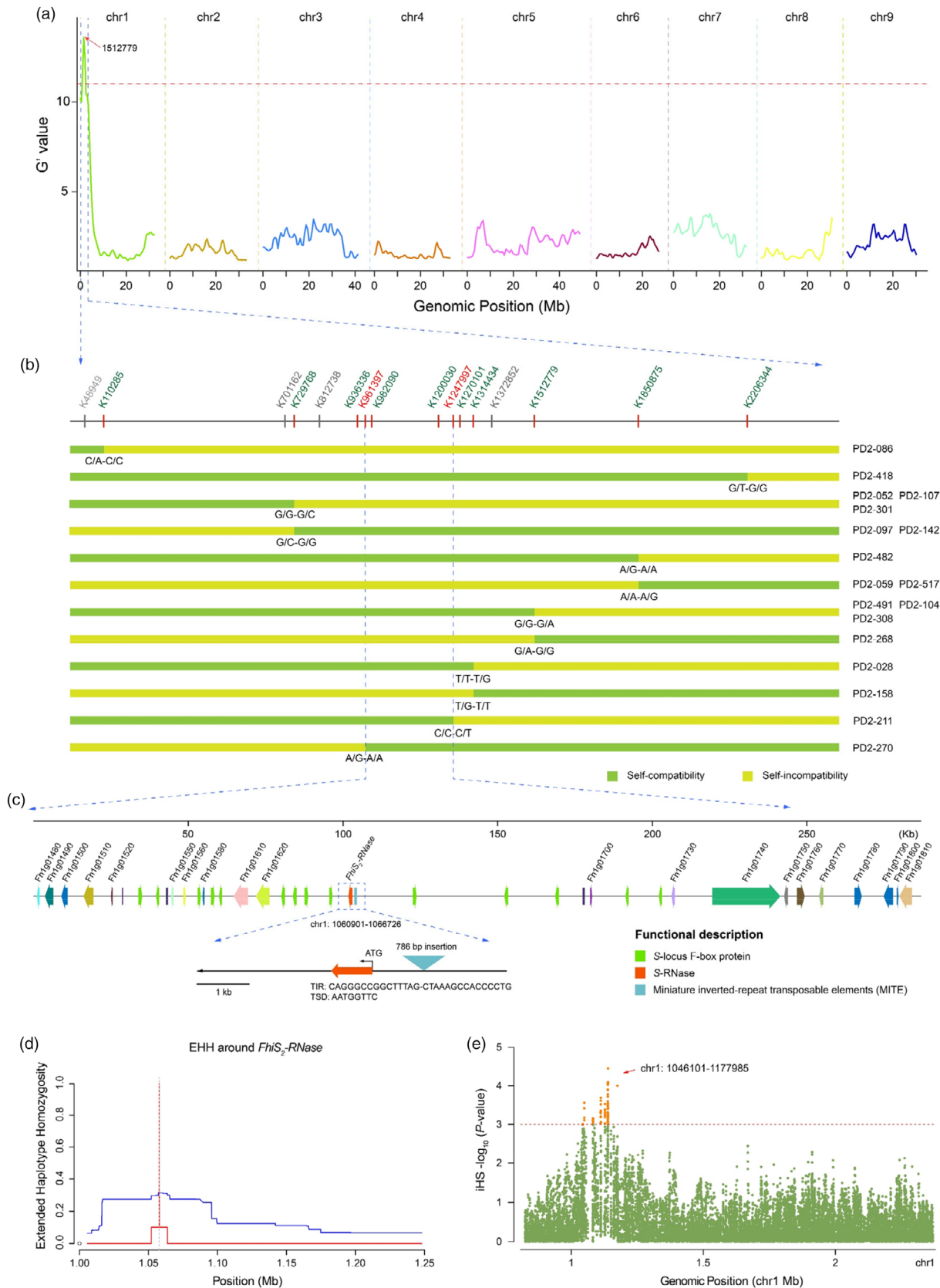
In this study, we document genetic variation at the SC-associated loci in *F. hindsii* (Hongkong kumquat, also known as 'Mini-Citrus') within a hybrid population and identify a 786-bp MITE insertion in the promoter of *FhiS₂-RNase* that results in the loss of SI in *F. hindsii*. Transgenic experiments show that deletion of the 786-bp MITE insertion in the *FhiS₂-RNase* promoter recovered the expression of *FhiS₂-RNase* in the *F. hindsii* style and restored SI. We observed the frequent occurrence of MITE insertions near *S*-RNase alleles in different intergeneric and interspecific of citrus and found that particular MITE insertions affect the expression of host genes, playing an important role in the loss of SI in citrus. In addition, MITE insertions have also been observed near *S*-RNases alleles of other common GSI families, and MITE-mediated SI loss is general in *S*-RNase-based SI system. Our analysis of the role of *S*-RNase-related MITE insertions in the evolution of selfing significantly broadens our understanding of another important mechanism contributing to SC in citrus. As *S*-RNases are widely distributed in several important plant families, our findings in citrus here are potentially relevant to the evolution of other *S*-RNase-based SI systems.

Results

S₂-RNase is associated with the SC phenotype of *F. Hindsii*

To explore the genetic basis underlying the conversion of SI to SC phenotype in *F. hindsii*, we performed a genetic analysis of F₁ hybrids from a cross (PN02 × DB02) between the SI diploid line PN02 (*S₈S₁₉*, *F. hindsii*) and the SC DB02 (*S₂S₂₉*, *F. hindsii*). In 2019 and 2020, analysis of 186 F₁ progeny established a ~ 1:1 segregation ratio of SC to SI (98 SC plants to 88 SI plants; $\chi^2 = 0.54$, $P = 0.46$; Figure S1; Table S1). These data indicate that SC in *F. hindsii* is caused by a single dominant gene or gametophytic factor (Table S2). Next, 32 SC and 32 SI F₁ individuals were selected for a BSA (bulk segregation analysis) utilizing whole-genome sequencing (WGS), which yielded a peak of SNPs linked to SC/SI on the anterior end of chromosome 1 (0.83 ~ 2.38 Mb, G' value = 11, $P = 10^{-3}$, Figure 1a; Figure S2a). KASP (Kompetitive allele-specific PCR) genotyping analysis was used to narrow the initial interval that was obtained from the BSA experiment. Using 12 pairs of markers, eighteen recombinant plants were identified from a total of 186 SC and SI F₁ individuals. Based on two recombinant plants (PD02-211 and PD02-270) and their

Figure 1 Genetic localization of key loci for self-compatibility traits in a *F. hindsii* population derived from a PN02 × DB02 cross. (a) BSA-Seq analysis of the F₁ hybrid population (PN02 × DB02). G' values plotted against the single nucleotide polymorphism (SNP) positions on the *F. hindsii* (S3y-45 v2.0) (Wang *et al.*, 2022) genome. Major trait loci for self-compatibility were identified using QTLseqr (Mansfeld and Grumet, 2018). The significance threshold ($P \leq 10^{-3}$) is indicated by the red horizontal line. The red arrow indicates a peak above the threshold value. (b) Fine mapping of the candidate gene using KASP genotyping. Eighteen recombinant offspring were classified into two groups according to the 16 pairs of KASP markers in the target region. Individuals with chromosome segment substitutions in the target region are shown. The SC haplotype is shown in pale green. The SI haplotype is shown in dark yellow. (c) Candidate genes in the target region containing two KASP markers (K961397 and K1247997). Recombinant analysis suggests the key gene of SC must be located in a 286.8-kb interval that contains 35 genes. Coloured boxes represent different genes associated with the *S₂-locus* of *F. hindsii*. The red box represents an *S*-RNase gene. The green boxes represent *SLF* genes. (d) Extended haplotype homozygosity (EHH) detected in the *FhiS₂-RNase* genomic region. Red and blue line correspond to the long haplotype and alternative variants, respectively. (e) Signatures of selection at SC-associated loci. Manhattan plot reflects the distribution of selection signatures of the initial interval (0.83 ~ 2.38 Mb, G' value = 11, $P = 10^{-3}$, BSA-Seq) with the integrated haplotype homozygosity score (iHS) statistic. The significance threshold ($P \leq 10^{-3}$) is indicated by the red horizontal line.



phenotypes, the key genetic loci controlling SC in *F. hindsii* were pinpointed to an interval between two markers, K961397 and K1247997 (286.6 kb, Figure 1b; Table S3). This

candidate interval overlapped with the *S*-locus that controls SI in citrus and contains 35 annotated genes, including 14 *F*-box genes, one *S₂-RNase* gene (named *FhiS₂-RNase*, *F. hindsii*) and

20 conserved structural genes in the flanking region of the *S*-locus (Figure 1c).

SC-associated loci are usually under strong selection in the population (Guo *et al.*, 2011). Focusing on this candidate region, to search for genomic regions with signs of recent positive selection, we calculated the extended haplotype homozygosity (EHH) of SC *F. hindsii* accessions (Table S4), which measures decay of haplotypes that carry a specified core allele as a function of distance and obtained the integrated haplotype score (iHS) statistic for each SNP. This identified a strong positive selection signal in the *FhiS₂-RNase* genomic region, which contained 75 significant SNPs ($P = 10^{-3}$) and was associated with 13 of 35 annotated genes mentioned above (Figure 1d,e; Figure S2b; Table S5). The SC phenotype was tightly linked to the *S₂-RNase* allele in the *F₁* progeny of the hybrid population (Table S1). Sequence variations including several synonymous and non-synonymous mutations in the 3' region of the *FhiS₂-RNase* allele from *F. hindsii* were predicted to result in a *FhiS₂-RNase* protein with 12 more amino acid residues relative to the other five alleles that encode *S₂-RNases* (*PtrS₂-RNase*, *CmeS₂-RNase*, *AbuS₂-RNase*, *CgrS₂-RNase* and *CreS₂-RNase*), although the conserved domain (C1-C5) (Liang *et al.*, 2020) was unaltered (Figure 2a; Figure S3).

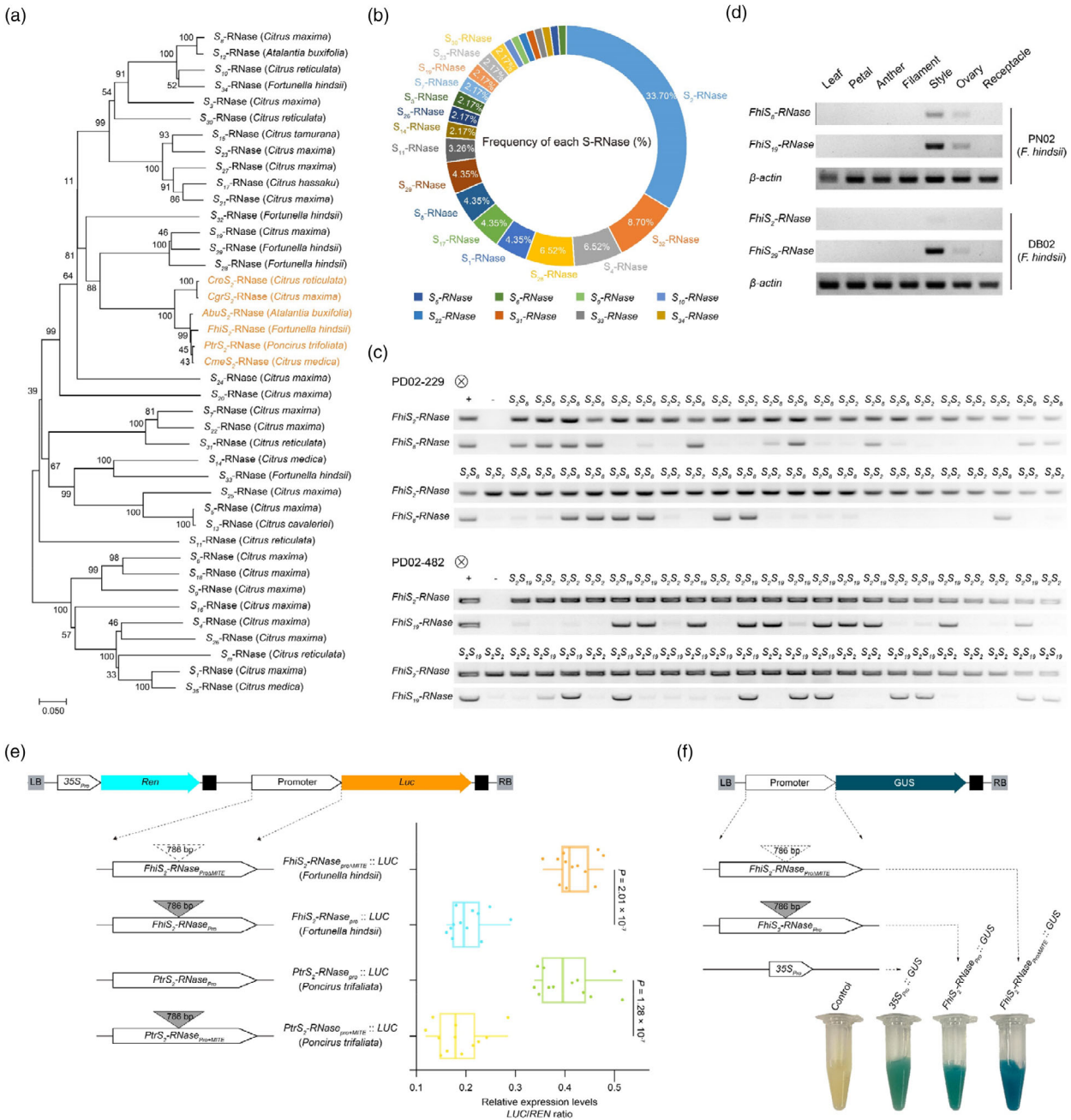
Frequency-dependent selection is a key feature of SI, and it determines the evolutionary dynamics of alleles at the *S*-locus (Castric and Vekemans, 2004; Wright, 1939). Genotypes and SI phenotypes were characterized in the 46 *F. hindsii* accessions, collected from 5 provinces in China (Table S4). The *S₂-RNase* genotype was most abundant (30/46) and was found in up to 33.7% of accessions (Figure 2b). Strikingly, all the accessions containing the *S₂-RNase* genotype were SC (Table S4). To detect the positive selection signatures between SC and SI *F. hindsii*, Cross-Population Extended Haplotype Homozygosity (XP-EHH) tests were performed. Four peaks were detected and 34 putatively advantageous positively selected genes were obtained (Figure S2c; Table S6), of which 17 genes overlapped with the KASP genotyping results, indicating that these genes, including *FhiS₂-RNase*, were strongly selected in the SC *F. hindsii* lines.

To further assess if the *FhiS₂-RNase* was the candidate gene responsible for the change in phenotype from SI to SC in *F. hindsii*, the segregation ratios of the *S* haplotypes in the selfed progeny (*S₁*) were analysed using chi-square tests. The *S* haplotypes of the progeny from the self-pollinated PD02-229 (*S₂S₈*), rather than segregating into a 1:2:1 ratio, segregated into two classes (*S₂S₂:S₂S₈*) at a ratio of 62:63. The *S₈S₈* genotype, which was an expected outcome of this self-cross, was absent from the *S₁* progeny (Table 1, Figure 2c). This is consistent with an SC mutation linked to the *S₂-locus* (1:1, $\chi^2 = 0.01$, $P = 0.93$). Furthermore, 116 individuals from the *S₁* progeny derived from the self-pollinated PD02-482 (*S₂S₁₉*) also segregated as 1:1 with *S₂S₂* (55 plants) and *S₂S₁₉* (61 plants) genotypes (1:1, $\chi^2 = 0.31$, $P = 0.58$), with the expected *S₁₉S₁₉* genotype being absent from the *S₁* progeny (Table 1; Figure 2c). These data show that only pollen carrying the *S₂-RNase* allele, but not the *S₈-RNase* or the *S₁₉-RNase*, are able to transmit to the *S₁* progeny (Figure 2c), further supporting the hypothesis that the *FhiS₂-RNase* is associated with the disruption of SI in *F. hindsii*.

A MITE suppresses the transcription of the *S₂-RNase*

Semi-quantitative real-time RT-PCR (SqRT-PCR) revealed that the expression of the *FhiS₂-RNase* in pistil tissue was significantly lower than the *FhiS₂₉-RNase* expression in SC DB02 (*S₂S₂₉*, *F. hindsii*), as well as the *FhiS₈-RNase* and *FhiS₁₉-RNase* expression in SI PN02 (*S₈S₁₉*, *F. hindsii*) (Figure 2d; Figure S4), suggesting that the down-regulated expression of the *FhiS₂-RNase* might be responsible for the SC phenotype. However, we found that the *FhiS₂-RNase* transcript was intact, with no mutations in it. A 786-bp insertion located 1071 bp upstream of the start codon of the *FhiS₂-RNase*, was identified when comparisons between sequences of *S₂-locus* of DB02 (SC, *S₂S₂₉*, *F. hindsii*) and ZK8 (SC, *S₂S₃₁*, *Poncirus trifoliata*, common wild trifoliata orange, a widely used citrus rootstock) were made (Figure S5). This comparison was made as, due to the significant differences in the non-coding region of *S₂-RNase* in different intergeneric of citrus, only ZK8 (*P. trifoliata*, *S₂S₃₁*) and DB02 (*F. hindsii*, *S₂S₂₉*) share high homology in this region. Using the P-MITE database (Chen *et al.*, 2014) and RepeatMasker, we determined that the

Figure 2 Influence of a MITE on the expression of the *FhiS₂-RNase*. (a) Phylogeny of 6 *S₂-RNase* alleles from different intergeneric and interspecific of citrus. The phylogenetic tree was derived from *PtrS₂-RNase* (*Poncirus trifoliata*), *CmeS₂-RNase* (*Citrus medica*), *FhiS₂-RNase* (*Fortunella hindsii*), *AbuS₂-RNase* (*Atlantia buxifolia*), *CgrS₂-RNase* (*Citrus maxima*), *CreS₂-RNase* (*Citrus reticulata*) (indicated in orange text) and 35 *S-RNase* alleles from different intergeneric and interspecific of citrus. The neighbour-joining phylogenetic tree was constructed using *S-RNase* amino acid sequences and MEGA X (Kumar *et al.*, 2018). (b) Frequency of the *S-RNase* genotype in 46 *F. hindsii* accessions. The *S* haplotypes of these *F. hindsii* accessions were assigned based on PCR-based genotyping of leaf DNA with 36 pairs of *S-RNase*-specific primers. These 46 accessions were collected from 5 provinces in China, and 24 *S-RNase* alleles were identified (see Table S4). Most of the genotypes were present at a low frequency. The *S₂-RNase* is present at the highest frequency of 33.7%. (c) Segregation of *S-RNases* with the *S*-locus in the *F₁* progeny of self-pollinated PD02-229 and PD02-482. Genotyping of the progeny from each pollination combination using PCR with *S₂*-, *S₈*- and *S₁₉-RNase* specific primers showed that PD02-229 and PD02-482 carried the same *S₂-RNase* allele. Pairs of amplified *S-RNase* sequences in three different combinations (*S₂S₈*, *S₂S₁₉* and *S₂S₂*) were amplified from 72 progeny from each self-pollination combination. (d) Semi-quantitative PCR was used to detect the expression of *S-RNase* alleles in different tissues from DB02 and PN02. The *S-RNase* allele was expressed only in the ovary and style tissues of pistils. The expression levels of the *FhiS₂-RNase* allele in the style tissues of DB02 were much lower relative to the other three alleles. The expression levels of these three alleles (*FhiS₈-RNase*, *FhiS₁₉-RNase* and *FhiS₂₉-RNase*) were similar in the style tissue of *F. hindsii*. (e) Activities of promoter fragments from *S₂-RNase* alleles in *F. hindsii* and *P. trifoliata*. The structure of the pGreenII 0800-LUC vector (top) and constructs with different promoter fragments from the *S₂-RNase* allele (left) are shown. The expression of the firefly luciferase (*LUC*) reporter gene was driven by different target promoter fragments. The expression of *Renilla Luciferase* (*Ren*) was used as an internal control. The expression levels of *LUC* relative to *REN* are shown with a boxplot (right). The horizontal lines indicate statistically significant differences ($n = 12$ biological replicates, two-tailed Student's *t*-tests). (f) Activity of *FhiS₂-RNase* promoter fragments in *F. hindsii*. Promoter fragments from the *S₂-RNase* allele were fused to the *GUS* (β -glucuronidase) reporter gene and expressed in the callus of DB02 using an *Agrobacterium*-based transient transformation method. A schematic diagram of the pK7WFS7 vector (above), diagram of the different promoter fragments from the *FhiS₂-RNase* allele (left) and staining of callus (below) are shown. p35S::GUS was used as the control.



786-bp insertion sequence was a MITE transposon that belongs to the *hAT* (*hobo*, *Ac* and *Tam3*) family. To establish if the MITE insertion in the *FhiS*₂-RNase promoter region contributed to changes in the expression of this allele, we generated various constructs by inserting different combinations of promoter fragments from *F. hindsii* into the binary vectors pGreen1 0800-LUC and pKGWFS7. We introduced these constructs into DB02 callus and examined the expression of the luciferase reporter gene (*LUC*). The MITE deletion (construct *FhiS*₂-RNase_{promMITE}::LUC) significantly increased the promoter activity of *FhiS*₂-RNase (Figure 2e, $P = 2.01 \times 10^{-7}$), this was consistent with the results from the GUS histochemical staining (Figure 2f). In addition, when the MITE insertion (construct *PtrS*₂-RNase_{pro+MITE}::LUC) was introduced into the *S*₂-RNase promoter of *P. trifoliata* at same position as the *FhiS*₂-RNase promoter, it significantly

decreased the promoter activity of the *PtrS*₂-RNase ($P = 1.28 \times 10^{-7}$, Figure 2e). These data demonstrate that the MITE insertion can modulate expression levels of *FhiS*₂-RNase in *F. hindsii*, suggesting that this 786-bp MITE could be responsible for the SC phenotype.

Evidence for CHH methylation clustered around the MITE transposon

The results of *in vitro* experiments suggested that the effect of 786-bp insertion on the expression levels of reporter genes (*LUC* and *GUS*) were not as completely suppressed as expected *in vivo* (Figure 2d-f; Figure S4). As cytosine methylation in plants is an epigenetic marker critical for transposon silencing, we investigated if this might be responsible for the attenuated expression of the *FhiS*₂-RNase in the *F. hindsii*. Whole-genome bisulphite

Table 1 Segregation of *S*-RNases in F_1 progeny in a gametophytic manner

Cross	No. of progeny	Possible genotypes ^a	Observed ratio ^b	Expected ratio ^c	χ^2 value	P-value
PD02-229 (S_2S_8) × PD02-229 (S_2S_8)	125	<u>S_2S_2</u> · <u>S_2S_8</u> · <u>S_8S_8</u>	62:63:0	1:1:0	0.01	0.93
PD02-482 (S_2S_{19}) × PD02-482 (S_2S_{19})	116	<u>S_2S_2</u> · <u>S_2S_{19}</u> · <u>$S_{19}S_{19}$</u>	55:61:0	1:1:0	61.51	4.39×10^{-14}
PN01 (S_7S_{28}) × PN03 (S_4S_7)	43	<u>S_4S_7</u> · <u>S_4S_{28}</u> · <u>S_7S_7</u> · <u>S_7S_{28}</u>	23:20:0:0	1:1:0:0	0.21	0.65
PN03 (S_4S_7) × PN01 (S_7S_{28})	53	<u>S_4S_7</u> · <u>S_7S_7</u> · <u>S_4S_{28}</u> · <u>S_7S_{28}</u>	0:0:24:29	0:0:1:1	0.47	0.49
				1:1:1:1	53.94	1.15×10^{-11}

Segregation analysis of *S* haplotypes in S_1 progeny from *F. hindsii* accessions from self-pollination and cross-pollination experiments analysed using PCR (Figure 2C).

^aThe observed genotypes are underlined.

^bThe *S*-genotype ratios observed in all progeny.

^cThe upper segregation ratio is the ratio expected from a single genotype mutation model from the GSI system. The lower segregation ratio is the ratio expected from simple Mendelian inheritance.

sequencing (WGBS) profiling of DNA methylation in the anther and style tissues from SC DB02 (S_2S_{29}), revealed that the CHH methylation levels in the style tissue of the *FhiS₂-RNase* region were significantly higher than in the anther tissue, with methylation rates of 16.12% and 4.13%, respectively, while the depth of coverage at cytosines and the average CpG, CHG and CHH methylation levels across the whole genome were similar in each tissue (Tables S7 and S8). Moreover, CHH methylation was mainly clustered around the MITE transposon that was inserted in the *FhiS₂-RNase* promoter region (Figure 3a). In contrast, CpG, CHG and CHH methylation levels in the other *S* allele (*FhiS₂₉-RNase*) region of the SC DB02 were not significantly different in anther and style tissues (Figure S6a). These data suggest that CHH methylation may contribute to the attenuated expression of the *FhiS₂-RNase* in the style of *F. hindsii*.

Studies have shown that TE-related methylation was usually associated with small RNAs (Borges and Martienssen, 2015; Zhang et al., 2018). Using small RNA sequencing, we found that a set of partially overlapping 24-nt siRNAs (small interfering RNAs) were generated from the flanking regions of the MITE in the upstream promoter of *FhiS₂-RNase* and that the partially overlapping siRNA peaks at both ends of the MITE were significantly higher in styles relative to levels in anther tissue (Figure 3b; Tables S9 and S10). In contrast, no siRNA peaks were detected in the *FhiS₂₉-RNase* region (Figure S6b). This is consistent with the CHH methylation of the *FhiS₂-RNase* promoter region in the style tissue of SC DB02 and with the idea that the 24-nt siRNAs generated from the MITE regulates *FhiS₂-RNase* expression by mediating the methylation of the *FhiS₂-RNase* promoter region.

Recovered expression of *S₂-RNase* restores the SI phenotype of *F. hindsii*

As our data suggested that lower expression levels of *FhiS₂-RNase* in style tissue might be responsible for the loss of SI in *F. hindsii*, we wished to ascertain what might be responsible for this. We, therefore, generated transgenic SC DB02 (S_2S_{29}) *F. hindsii* plants with increased expression of *FhiS₂-RNase* using 35S promoter ($35S_{pro}::FhiS_2-RNase-DB02$), the native promoter of *FhiS₂₉-RNase* ($FhiS_{29}-RNase_{pro}::FhiS_2-RNase-DB02$) and its own promoter of *FhiS₂-RNase* without 786-bp MITE insertion ($FhiS_2-RNase_{pro,AMITE}::FhiS_2-RNase-DB02$) (Figure S7a–c). The relative expression levels of the *FhiS₂-RNase* were significantly elevated in transgenic lines

relative to the empty vector control plants. Twelve stably transformed and flowering lines (seven $35S_{pro}::FhiS_2-RNase-DB02$ and five $FhiS_{29}-RNase_{pro}::FhiS_2-RNase-DB02$ lines) were selected for a follow-up study (Figure 4a,f; Figure S7a).

The expression levels of *FhiS₂-RNase* in these SC *F. hindsii* transgenic lines ranged from 0.8-fold to 143.3-fold increase (Figure 4a,f; Figure S7d). In the line $FhiS_{29}-RNase_{pro}::FhiS_2-RNase-DB02$ #14, where the expression of *FhiS₂-RNase* was increased by more than 14.9-fold (which corresponded to ~76% of the expression level of the other endogenous *S* allele, *FhiS₂₉-RNase*; Figure 4f), the SI phenotype was restored (Figure 4b–e,i; Figure S7e). The relative expression levels of the other five lines ($35S_{pro}::FhiS_2-RNase-DB02$ #4, $FhiS_{29}-RNase_{pro}::FhiS_2-RNase-DB02$ #1, #4, #12, #17) were elevated by 5.5-fold or less (which corresponded to ~25% of the expression level of the other endogenous *S* allele, *FhiS₂₉-RNase*; Figure 4a,f), and aniline blue staining of pollinations showed that these lines and the empty vector control plants all maintained the SC phenotype (Figure 4g,h,j; Figure S7e). These data suggest that when *FhiS₂-RNase* is expressed beyond a certain threshold level (~75% of *S*-RNase expression in SI plants) in SC *F. hindsii*, the SI phenotype can be restored. In contrast, when the expression levels of the *FhiS₂-RNase* allele in the style tissue of the transgenic lines were similar or only increased by ~5-fold to the empty vector control plants, the transgenic lines remained SC (Figure 4g,h,j; Figure S7e). These data demonstrate that the loss of SI in *F. hindsii* is due to the low expression level of the *FhiS₂-RNase*. When the *FhiS₂-RNase* is expressed above a threshold level, it acts as a functional female *S*-determinant that is responsible for the specific inhibition of self-pollen tubes. These data provide evidence that *FhiS₂-RNase* performs the same function as *S₂-RNases* in the other intergeneric of citrus. Moreover, they demonstrate that highly expressed *FhiS₂-RNase* in styler tissues can restore the SI phenotype. As overexpression of the *S₂-RNase* was sufficient to restore the SI phenotype in self-fertile *F. hindsii*, this suggests that all the components required for SI are retained except for the *S₂-RNase* allele.

Transgenic lines demonstrate that removal of the 786-bp MITE insertion restores the SI phenotype of *F. hindsii*

Although making transgenic lines in woody tree species is highly time-consuming, to confirm whether the MITE insertion in the promoter of *FhiS₂-RNase* was responsible for its low expression

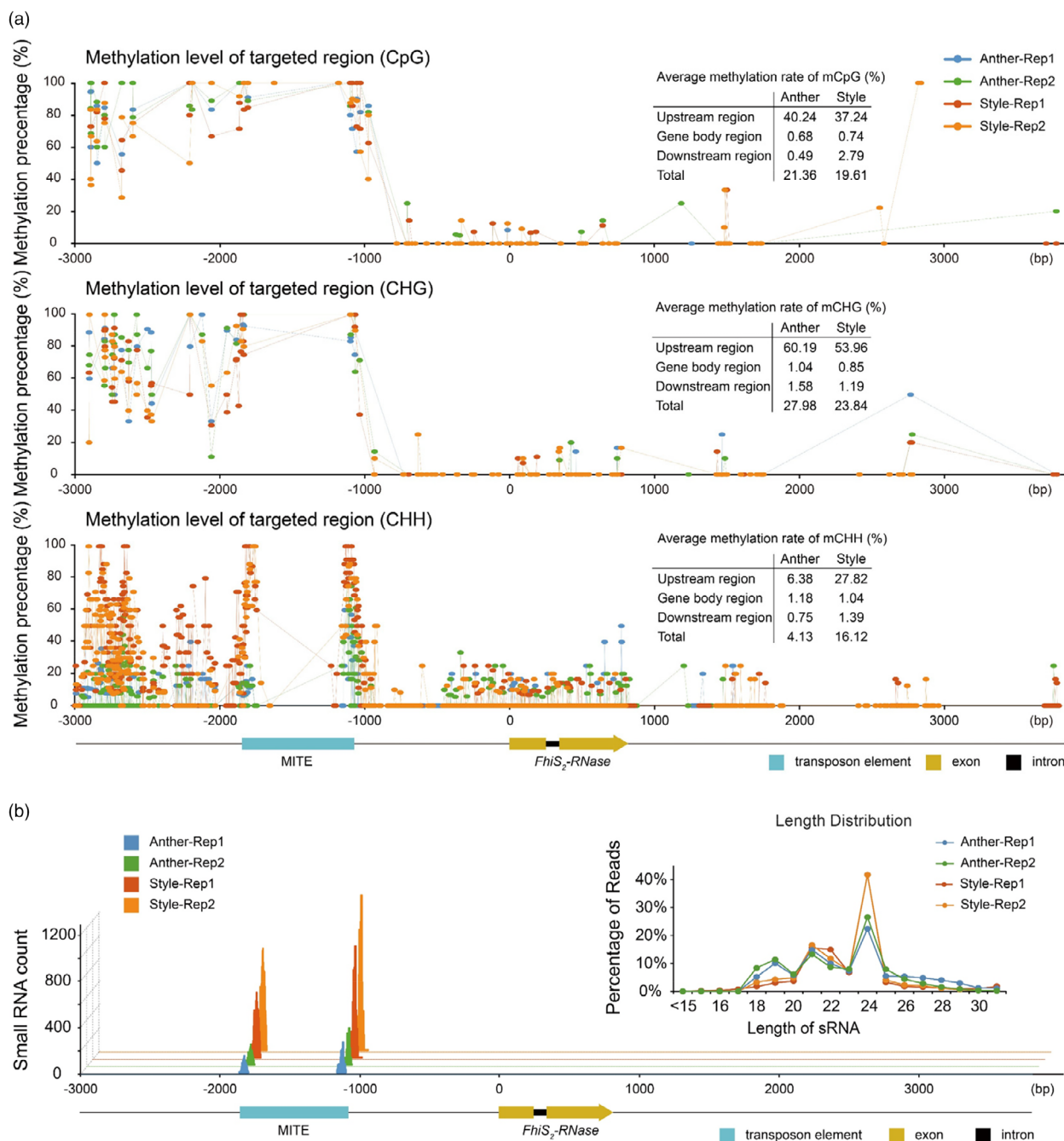


Figure 3 Influence of methylation may be involved in the expression of the *FhiS₂-RNase*. (a) Methylation levels of the *FhiS₂-RNase* allele in the leaf, anther and style of *F. hindсии*. Methylation levels in the CpG (top row), CHG (middle row) and CHH (bottom row) contexts were quantified in 3-kb 5'-flanking regions, exons, introns and 3-kb 3'-flanking regions. The percentage of total mCs is the number of mCs/total number of Cs. (b) Numbers of small RNAs from anthers and styles that map to the 3-kb 5'-flanking region of the *FhiS₂-RNase* allele in DB02. Most small RNAs map to the MITE region. The linear graph in the upper right corner shows the length distribution of small RNAs in anther and style tissue of DB02.

level, we constructed complementary transgenic lines with 786-bp MITE insertion removed (*FhiS₂-RNase_{proΔMITE}::FhiS₂-RNase-DB02*). Among the obtained 3 flowering transgenic lines, the relative expression of 2 lines was increased about 20-fold (Figure 4k; Figure S7a), and the results of aniline blue staining showed that their styles could reject the growth of self-pollen tubes, indicating that the SI phenotype had been restored compared with the control (Figure 4m–o). While aniline blue

staining of pollinations showed that the *FhiS₂-RNase_{proΔMITE}::FhiS₂-RNase-DB02* #3 line still maintained the SC phenotype (Figure 4p), and its relative expression level was just elevated by ~3-fold (Figure 4k). Furthermore, we also found that the expression of *S₂-RNase* (*PtrS₂-RNase*) in ZK8 (*P. trifoliata*, *S₂S₃₁*, common wild trifoliata orange) style was significantly higher than that of the DB02 (*F. hindсии*, *FhiS₂-RNase*), and its promoter lacked 786-bp MITE insertion compared to *FhiS₂-RNase* (Figure 4l;

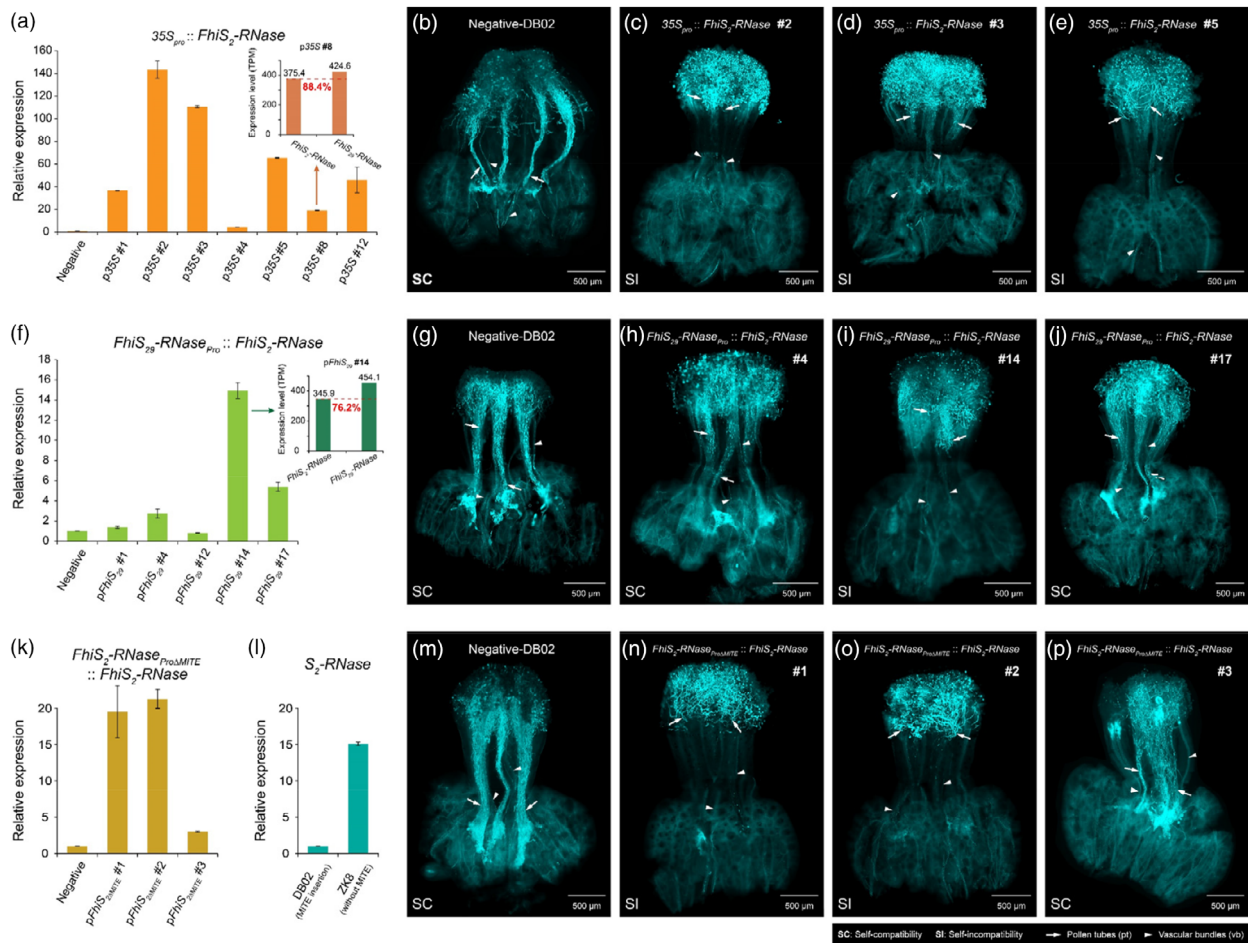


Figure 4 Recovered expression of *FhiS₂-RNase* restores the SI phenotype of *F. hindsii*. (a) Comparison of the relative expression levels of the *FhiS₂-RNase* allele in the styles of a negative transgenic line and 7 overexpression lines ($n = 3$ biological replicates). The small bar chart (upper right) shows the transcript levels of *FhiS₂-RNase* and *FhiS₂₉-RNase* alleles in transgenic line (*35S_{pro}::FhiS₂-RNase* #8). (b–e) Representative aniline blue staining of self-pollinated styles from *35S_{pro}::FhiS₂-RNase-DB02* transgenic plants ($n = 12$). Negative indicates plants from the transformation that lost the transgene because of segregation. (f) Relative expression levels of the *FhiS₂-RNase* allele in the styles of negative controls and 5 transgenic lines (*FhiS₂₉-RNase_{pro}::FhiS₂-RNase*). The small bar chart (upper left) shows the transcript levels of *FhiS₂-RNase* and *FhiS₂₉-RNase* alleles in transgenic line (*FhiS₂₉-RNase_{pro}::FhiS₂-RNase* #14). TPM values were calculated with the reads mapped to the DB02 assembled reference genome. The percentage below the dotted red line indicates the ratio of the *FhiS₂-RNase* transcription level in the transgenic line to the endogenous other *S* allele (*FhiS₂₉-RNase*) expression level. (g–j) Representative aniline blue staining images of pollen tubes in pistils from *FhiS₂₉-RNase_{pro}::FhiS₂-RNase-DB02* transgenic plants ($n = 17$). (k) Relative expression levels of the *FhiS₂-RNase* allele in the styles of negative controls and 3 transgenic lines (*FhiS₂-RNase_{pro:AMITE}::FhiS₂-RNase*). Data are presented as means \pm SE. ($n = 3$ biological replicates). (l) Comparison of the relative expression levels of the *S₂-RNase* allele in the styles of DB02 (*F. hindsii*, *S₂S₂₉*; the promoter of *FhiS₂-RNase* contains a 786-bp MITE insertion) and ZK8 (*P. trifoliata*, *S₂S₃₁*; Compared with *FhiS₂-RNase*, the promoter of *PtrS₂-RNase* does not contain the MITE insertion, Figure S5). Data are presented as means \pm SE. ($n = 3$ biological replicates). (m–p) Representative aniline blue staining images of pollen tubes in pistils from *FhiS₂-RNase_{pro:AMITE}::FhiS₂-RNase-DB02* transgenic plants ($n = 3$). Five images were acquired for each pollination combination. Representative fluorescence images from aniline blue stained pistils are shown that were acquired at 1 day before anthesis and at 2 days after pollination. Scale bars = 500 μ m. Pollen tubes (pt) are indicated with arrows. Vascular bundles (vb) are indicated with arrowheads.

Figures S4a,c, and S5). These results indicated that the 786-bp MITE insertion inhibited the expression of *FhiS₂-RNase* and was responsible for the loss of SI in *F. hindsii*.

Diverse SC-linked structural variations do not directly affect loss of SI in *F. hindsii*

It has been reported that SVs are involved in the loss of SI and play an important role in the evolution of *S*-locus (Li et al., 2019; Okada et al., 2008); in *Arabidopsis* SC haplotypes are mainly generated by recombination and SVs (Goubet et al., 2012;

Tsuchimatsu et al., 2017). Due to the high polymorphic nature of the *S*-locus, the identification of SC-linked SVs is challenging. To further assess the structural differences and to identify SC-specific features of *S* haplotype in *F. hindsii*, we sequenced and assembled the genomes of SC DB02 and SI PN02. The genomes were based on single molecule real-time (SMRT) DNA sequencing and *de novo* assembled with Canu (Koren et al., 2017), which produces two partially phased diploid genomes (a DB02gv1 contig and a PN02gv1 contig; Table S11). Four complete *S*-locus sequences (*FhiS₂*-locus, *FhiS₈*-locus, *FhiS₁₉*-locus and *FhiS₂₉*-locus) were

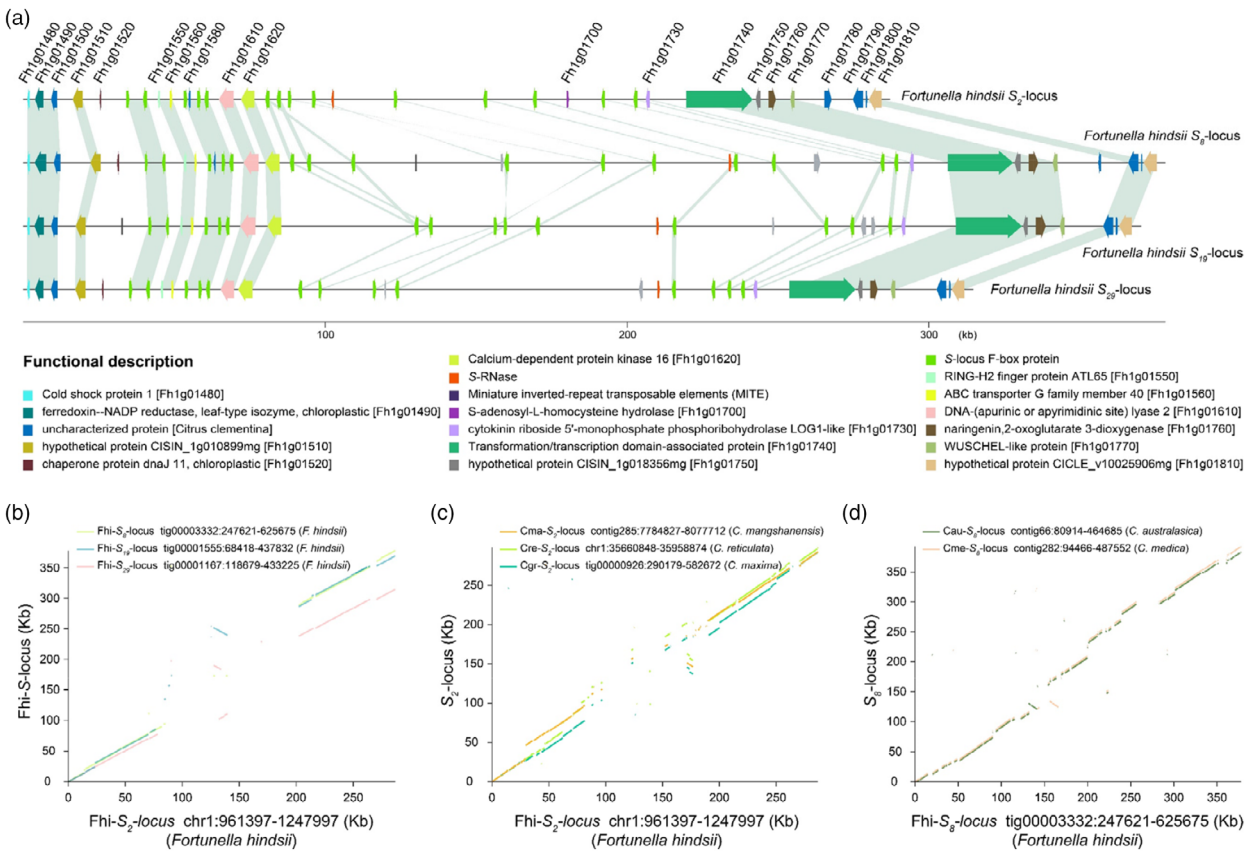


Figure 5 Structural variant analysis of *S*-locus in DB02 and PN02. (a) Schematic diagram for the gene annotation and collinearity analysis of 4 *S*-locus in the genomes of DB02 and PN02. The blue and grey lines represent the syntenic sequences among the genes at the four *S*-locus of *F. hindsii*. The coloured boxes represent different genes associated with the *Fhi-S₂*-locus, *Fhi-S₈*-locus, *Fhi-S₁₉*-locus and *Fhi-S₂₉*-locus. The red boxes represent *S*-RNase alleles. The green boxes represent the *SLF* alleles. (b) Whole-sequence alignments of the *S*-locus region in DB02 and PN02. The 200-bp sequences of two non-recombinant KASP markers (K961397 and K1247997) were used as the boundaries of the *S*-locus. Sequences of the *Fhi-S₂*-locus, *Fhi-S₈*-locus, *Fhi-S₁₉*-locus and *Fhi-S₂₉*-locus were obtained from the DB02 and PN02 genomes (DB02gv1 and PN02gv1). (c) Whole-sequence alignment analysis of the *S₂*-locus region in different intergeneric and interspecific of citrus. The sequences of the *Cma-S₂*-locus, *Cre-S₂*-locus and *Cgr-S₂*-locus were derived from the genomes of ‘mangshanyejū’ (*Citrus mangshanensis*), ‘Ponkan’ (*Citrus reticulata* Blanco cv. Ponkan) and pummelo (*Citrus maxima*), respectively. The *Fhi-S₂*-locus sequence was used as reference. (d) Whole-sequence synteny analysis of the *S₈*-locus region for different interspecific of citrus. *Cau-S₈*-locus and *Cme-S₈*-locus were derived from the genomes of *C. australasica* and *C. medica*, respectively. The *Fhi-S₈*-locus sequences were used as references. Detailed information on *S*-locus in different intergeneric and interspecific of citrus is provided in Table S14.

obtained using the 200-bp sequences of two non-recombinant KASP markers (K961397 and K1247997) as the boundary (Table S3).

These four *S*-locus were annotated based on available transcriptome data (Figure 5a) and the SI haplotypes were aligned to the *F. hindsii* (S3y-45 v2.0) (Wang et al., 2022) *Fhi-S₂*-locus to assess structural differences and to identify SC-specific features. There were differences in length among the SI and SC haplotypes in *F. hindsii* (Figure 5b). The *S*-locus ranged in size from ~286.8 to ~378.1 kb, but the *Fhi-S₂*-locus were significantly shorter relative to other SI haplotypes (Figure 5a). These length differences appear to be related to SC-linked structural variations. The SI haplotypes share 15 small insertions and 6 small deletions relative to the *Fhi-S₂*-locus, encompassing a total length of 30.2 kb. These SVs were located mainly in the flanking regions of the *S*-locus. In addition, there were 7 translocations and 8 inversions in the *Fhi-S₂*-locus relative to the three other SI haplotypes, including 2 large inversions located between two *SLFs* alleles (*SLF4* and *SLF5*; Figure 5b). Although the ~125-kb middle region of different *S*-locus in *F. hindsii* was highly polymorphic,

the sequences flanking both ends of the *S*-locus were conserved and contained a highly collinear group of genes. Most of the polymorphic regions were composed of repeat sequences and TEs, including MITEs, long terminal-repeat (LTR) retrotransposons, gypsy elements and copia elements (Figure 5a,b; Table S12). It is noteworthy that, compared to the SC *Fhi-S₂*-locus of *F. hindsii*, the other three SI *S₂* haplotypes (*Cma-S₂*-locus, *Cre-S₂*-locus and *Cgr-S₂*-locus) share a common large inversion (*Fhi-S₂*-locus: 171501–176731) located between *FhS₂*-*SLF2* and *Fh1g01700*. In addition, compared to the *Fhi-S₂*-locus, the three *S₂* haplotypes have 26 small deletions and 12 small insertions with a total length of 39.9 kb (Figure 5c).

Phylogenetic analysis revealed that all the *SLF* alleles located at *S*-locus in citrus could be divided into 14 types (Figure S8; Table S13). Translocations, inversions and tandem repeats in *SLF* alleles surrounding the *S*-RNase were identified; these may be caused by SVs in the *S*-locus (Table S12). All 14 types of *SLF* alleles were found in the *Fhi-S₂*-locus of *F. hindsii* and there was no significant difference in the expression of these alleles compared to *SLF* alleles in other *S*-locus (Figure S9a,b). Thus, although there

were many SVs in self-compatible Fhi- S_2 -locus relative to other SI haplotypes, they did not appear to affect gene content or expression and they do not seem to be responsible for the SC phenotype in *F. hindsii*. The MITE insertion in the promoter of the S_2 -RNase allele was unique to *F. hindsii* and took place after the divergence of genera in citrus. Our data suggest that the fixation and propagation of the MITE insertion in the Fhi- S_2 -locus was responsible for the loss of SI in the *F. hindsii* population.

MITE insertions are extensive in citrus and are associated with a SC phenotype

There are multiple possible routes for the loss of SI. However, to date, only one route has been found in citrus, which involved the inactivation of the S -RNase (S_m -RNase), which led to the loss of SI from mandarin and its hybrids (*Citrus* genus) (Liang et al., 2020; Zhao et al., 2022). Here we have identified a new, alternative route for loss of SI in *Fortunella* genus, which was caused by the insertion of a MITE in the promoter region. We examined whether this phenomenon is widespread. An alignment of S -locus shows that there are many SVs in the same S haplotypes of different intergeneric and interspecific citrus, which is consistent with the S haplotypes evolving in parallel during the species differentiation of citrus (Figure 5c,d). Moreover, these SVs contain many repeated sequences and TEs, including the MITE insertion found in the Fhi S_2 -RNase gene promoter of *F. hindsii*, which we identified as being responsible for the loss of SI in *Fortunella*.

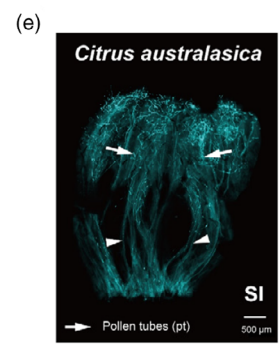
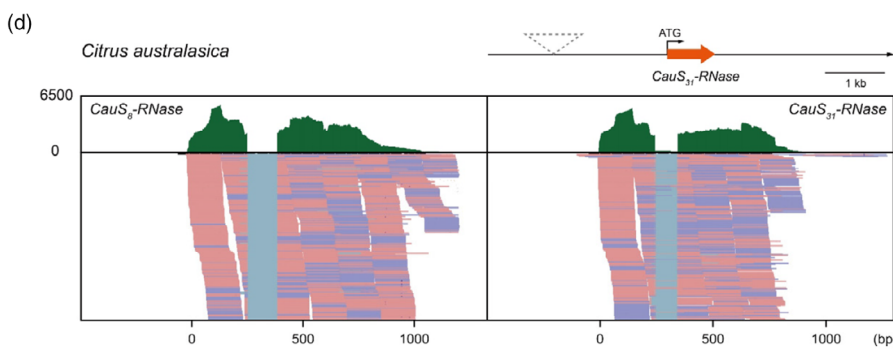
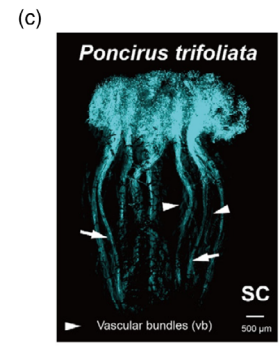
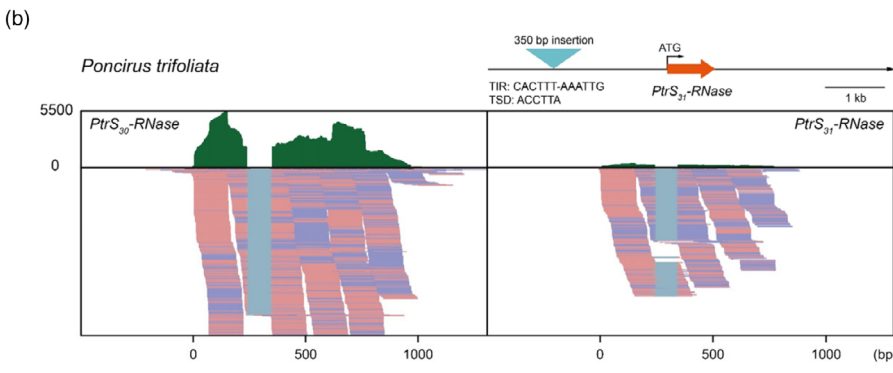
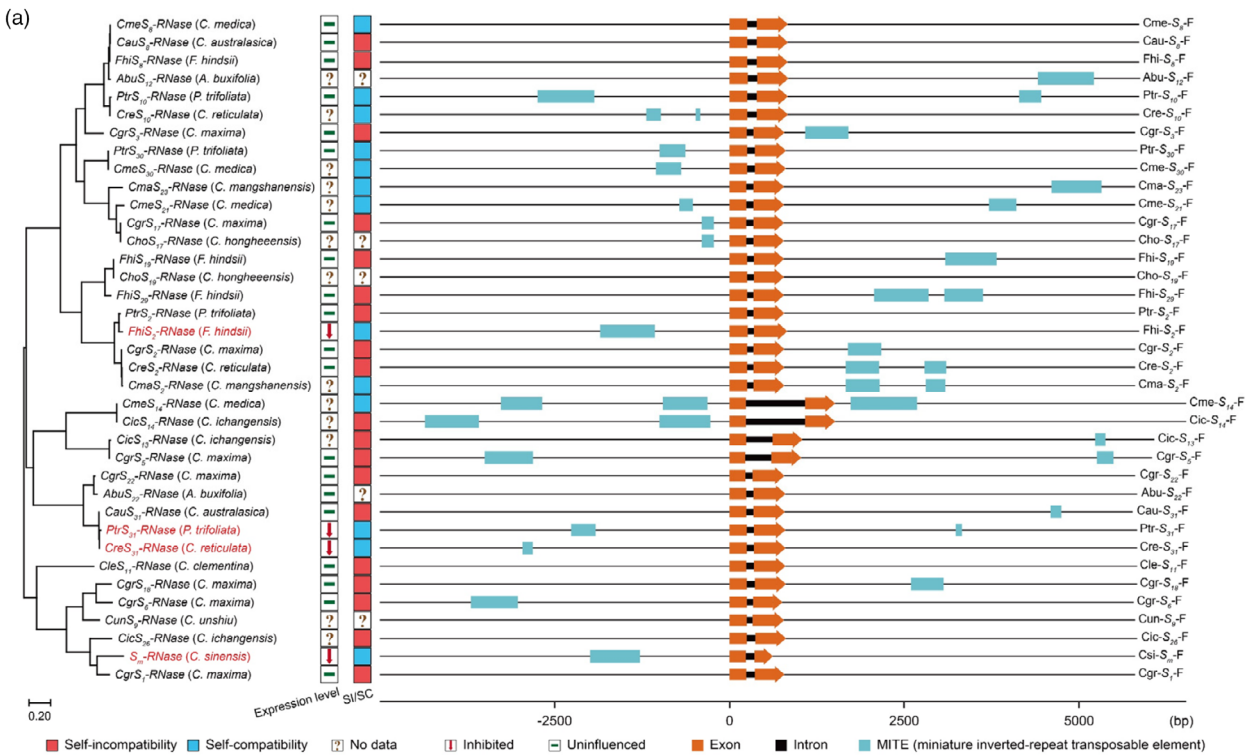
To investigate the MITE insertions of adjacent S -RNase alleles in different citrus populations, we obtained 37 ~ 10.7-kb genomic fragments that cover the entire S -RNase allele, including 5 kb from the 5'-flanking region and 5 kb from the 3'-flanking region, using published citrus genomes data and SMRT sequencing (The sequence fragments can be found in Tables S11, S14 and S15). The MITE annotation was performed using the Repeatmasker Web Server (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>), MITE-Hunter (Han and Wessler, 2010) and the P-MITE database (Chen et al., 2014). A total of 37 annotation results were obtained. These MITEs were mainly derived from three subfamilies and their lengths ranged from 91 to 953 bp (Table S15). The MITE annotation indicates

that the MITE insertions near S -RNase alleles occurred frequently in different intergeneric and interspecific of citrus (Figure 6a). We, therefore, examined if they might have a similar effect on SI as that found in *F. hindsii*, which led to the loss of SI in the *Fortunella* population.

Nucleotide alignments revealed the presence of a 350-bp sequence containing a MITE in the promoter region of the $PtrS_{31}$ -RNase in SC *P. trifoliata* (Table S15; Figures S10a,b and S11). The expression levels of the $PtrS_{31}$ -RNase in the styles of *P. trifoliata* were significantly lower relative to other alleles and aniline blue staining showed that *P. trifoliata* was SC (Figure 6b,c). We also found that, in contrast, *Citrus australasia* and *Citrus ichangensis* which are SI, lacked the 350-bp sequence containing a MITE and the expression levels of $AbuS_{31}$ -RNase in the styles of *C. australasia* were similar to another allele ($AbuS_8$ -RNase) (Figure 6d,e). This is consistent with our interpretation that the presence of a 350-bp MITE-containing sequence in the promoter region is responsible for a SC phenotype. Furthermore, in *Citrus reticulata*, the 'Mashuiju' and 'Mingliutianju' mandarins are SC, despite not having the S_m -RNase frameshift mutation previously shown to be responsible for the loss of SI in *Citrus* genus (Liang et al., 2020). We identified another MITE insertion (115-bp sequence) in the upstream region of the $CreS_{31}$ -RNase, which accompany significantly lower expression levels in the style (Figures S11 and S12A). We also identified a MITE insertion in the promoter of the S_m -RNase allele, which was accompanied by significantly reduced expression of the S_m -RNase in the style relative to the other S allele (Figure 6a; Figure S12b; Table S15). We also found a 961-bp MITE insertion in the 3'-flanking regions of $CicS_{14}$ -RNase allele in SC *Citrus medica* (Figures S13 and S14). Together, this suggests that MITE insertions occur frequently adjacent to S -RNase alleles and only certain MITEs appear to be responsible for its attenuated expression and a SC phenotype.

The automatic advantage of selfing favour the spread of SC haplotypes in populations, and we previously showed that the S_m -RNase is prevalent (90/153) in 153 citrus accessions (Goldberg et al., 2010; Liang et al., 2020). We, therefore, examined the frequency of different S -RNase genotypes. We found that the Fhi S_2 -RNase with the MITE insertion responsible for conferring SC

Figure 6 MITE insertions are extensive in citrus and are associated with a SC phenotype. (a) Analysis of MITE insertions in the 5-kb flanking regions of S -RNase alleles. Detailed information is provided for 37 genomic fragments from different S haplotypes of citrus that were analysed. MITE annotation in the 5-kb flanking regions of S -RNase alleles were made using RepeatMasker (4.1.2), MITE_Hunter (Han and Wessler, 2010) and the P-MITE database (Chen et al., 2014) (see Table S15 for detailed information). Orange blocks represent S -RNase alleles, cyan represents MITE transposons (right). A neighbour-joining phylogenetic tree was constructed using the amino acid sequences of S -RNases and MEGA X (Kumar et al., 2018). Red text indicates SC haplotypes whose expression of S -RNase is inhibited due to the MITE insertion (left). Different colour boxes represent SC or SI haplotypes. A downward red arrow indicates significant inhibition of S -RNase, a short green horizontal line indicates normal S -RNase allele expression, and a question mark in the box indicates a lack of data (middle). (b) Sequence read clusters from the $PtrS_{30}$ -RNase and $PtrS_{31}$ -RNase (*P. trifoliata*) alleles. The sequence read clusters are from the RNA-Seq data generated from the styles of *P. trifoliata* and are shown in the Integrative Genomics Viewer (Robinson et al., 2011). There were significantly more reads mapped to $PtrS_{30}$ -RNase than $PtrS_{31}$ -RNase in the styles of *P. trifoliata*. (c) Representative aniline blue staining images from pollen tubes in the pistils of *P. trifoliata*. Five images were acquired 5 days after pollination for aniline blue stained pistils at stage -1 DBA (1 day before anthesis) for each combination of pollination. Representative fluorescence images are shown. (d) Sequence read clusters from the $CauS_8$ -RNase and $CauS_{31}$ -RNase (*C. australasia*) alleles. The sequence read clusters are from the RNA-Seq data generated from the styles of *C. australasia* and are shown in the Integrative Genomics Viewer (Robinson et al., 2011). The reads mapped to the $CauS_8$ -RNase and $CauS_{31}$ -RNase alleles were equivalent in the styles of *C. australasia*. The green bars depict the number of reads mapped to the S -RNase sequences that included 500-bp of 5'-flanking sequence, exons, introns and 500-bp of 3'-flanking sequence). A partial alignment of the RNA mapping data is shown (below). Pink and blue represent the reads from the different strands. (e) Representative aniline blue staining of self-pollinated styles from an *C. australasia* accession collected at Huazhong Agriculture University. Styles from five pollinations from each accession were examined. The growth of the pollen tubes was inhibited near the top of the styles, which is a SI phenotype. Bunches of pollen tubes extending through the pistil is a SC phenotype. Scale bars = 500 μ m. Pollen tubes (pt) are indicated with arrows. Vascular bundles (vb) are indicated with arrowheads.



in the *Fortunella* population occurs at a high frequency (Figure 2b; Table S4), which is consistent with the notion that SC is advantageous. We examined 113 accessions of *P. trifoliata* and found that the number of polymorphisms in the *S-RNase* alleles was generally low (~12 *S-RNase* genotypes, with the exception of the *PtrS31-RNase*, whose promoter contained a MITE insertion accompanied by low expression of this *S-RNase*, which had a high frequency (89.38%, Figure S15a). In the 127 *C. medica* accessions examined, the frequency of SC *CmeS14-RNase* and

CmeS21-RNase alleles were also relatively high (Figure S15b). As we identified MITE insertions in the 3'-flanking and promoter regions of *CmeS14-RNase* and *CmeS21-RNase*, respectively, this high frequency could be attributed to the SC phenotype caused by the MITE insertion. However, it should be noted that not all MITEs inserted near *S-RNase* alleles influence expression, as illustrated by the *CgrS2-RNase*, *CgrS3-RNase*, *CgrS5-RNase*, *CgrS6-RNase*, *CreS2-RNase* (Hu *et al.*, 2021; Liang *et al.*, 2020), *FhiS19-RNase*, *FhiS29-RNase* (Figure S4), *PtrS10-RNase* (Figure S16),

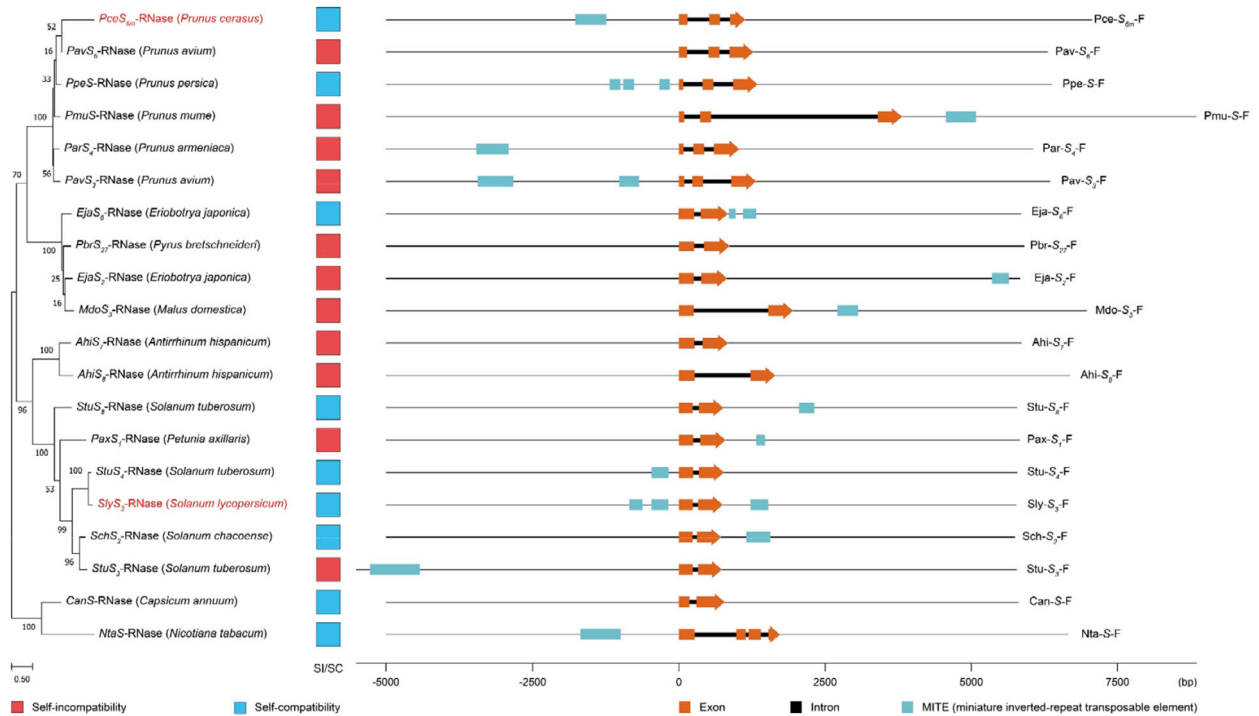


Figure 7 MITE insertions identified in the 5-kb flanking regions of *S-RNase* alleles form three common eudicot families (Solanaceae, Plantaginaceae, Rosaceae). The sequences of twenty genomic fragments from different *S* haplotypes of fifteen GSI species from three common eudicot families (Solanaceae, Plantaginaceae, Rosaceae) that utilize the *S-RNase* SI system were analysed. A neighbour-joining phylogenetic tree was constructed using the amino acid sequences of *S-RNases* and MEGA X (Kumar et al., 2018) (left). Annotation of MITEs occurring in the 5-kb flanking regions of *S-RNase* alleles was made using RepeatMasker (v 4.1.2, <https://www.repeatmasker.org/>), MITE_Hunter (Han and Wessler, 2010) and the P-MITE database (Chen et al., 2014) (see Table S16 for details). Orange blocks indicate *S-RNase* alleles, cyan indicate MITE transposons (right). Red boxes and blue boxes (left) indicate SI and SC haplotypes, respectively. Red text indicates SC haplotypes whose expression of *S-RNase* is known to be inhibited (Kondo et al., 2002b; Yamane et al., 2003).

PtrS₃₀-RNase (Figure 6b) and *CauS₃₁-RNase* alleles (Figure 6d). Our findings indicate that a large number of TEs, including MITEs, were inserted and enriched around the *S*-locus, were inserted randomly near *S-RNase* alleles and evolved in parallel throughout intergeneric and interspecific of citrus (Table S15).

MITE-mediated SI loss is a more general phenomenon in *S-RNase*-based SI systems

To investigate the MITE insertions of adjacent *S-RNase* alleles in other GSI families, we examined the sequences from fifteen GSI species that utilize the *S-RNase* SI system. The data revealed the presence of MITE insertions adjacent to the *S-RNase* alleles were also observed in GSI families. We found that in species known to be SC, MITE insertions were generally found to be located in the promoter region of the *S-RNase* genes, while in the SI species either MITEs were absent or inserted further away from the *S-RNase* (Figure 7; Table S16). This is consistent with our findings in citrus, suggesting that MITE-mediated SI loss may be a more general phenomenon in *S-RNase*-based SI systems.

In summary, here we demonstrate for the first time that MITEs provide an important mechanism that is responsible for loss of SI. Our study, shows that MITE insertions in the promoter regions of *FhiS₂-RNase* and *PtrS₃₁-RNase* reduce *S-RNase* expression and transgenic lines demonstrate that removal of the 786-bp MITE insertion restores the SI phenotype of *F. hindsii*. We further show that the propagation of this 786-bp MITE in the *FhiS₂-RNase* promoter region is responsible for the SC phenotype in a

population of *Fortunella*. Our study suggests that this MITE insertion led to the loss of SI in *Fortunella* and *Poncirus* genera, respectively, and that MITEs are likely to have played an important role in SI to SC transitions (Figure S17), as these transposons were inserted frequently at the *S*-locus in several SI species that utilize the *S-RNase* system.

Discussion

Transposable elements (TE) are a rich source of genetic variation in plant genomes, which can be fixed and propagated by selection and can mediate phenotypic diversity (Feschotte et al., 2002). It is well established that the insertion of TEs can cause extensive changes in gene composition and function and this can play an important part in plant adaptation and evolution (Chuong et al., 2017; Lisch, 2013). The development of genomics and phenomics has made it possible to systematically identify and evaluate the role that TEs have potentially played. For example, the insertion of a terminal-repeat retrotransposon in miniature (TRIM) element upstream of *Ms2* allele resulted in a male sterility phenotype in wheat (Xia et al., 2017); drought tolerance in maize was closely associated with the insertion of a MITE in the promoter of *ZmNAC111* (Mao et al., 2015); studies on rice have shown that the susceptibility of rice to bacterial blight was related to the TE insertion in the intron of *WRKY45-1* (Zhang et al., 2016); in potato, the insertion of a 533-bp MITE in the promoter of *Sli* (*S*-locus inhibitor) enables its expression in pollen,

resulting in the breakdown of SI (Eggers *et al.*, 2021; Ma *et al.*, 2021). Studies have shown that TEs have contributed to the regulation and evolution of important traits in plants, such as the red coloration of fruit peels (Zheng *et al.*, 2019), the formation of bloody sarcocarp (Butelli *et al.*, 2012; Lisch, 2013) and apomixis (Shimada *et al.*, 2018; Wang *et al.*, 2017) in citrus.

Studies on TE distribution in the genomes suggest that chromosomal regions with low recombination frequencies are associated with accumulation of TEs (Kent *et al.*, 2017). The *S*-locus, which determines SI, has long been thought to be a genomic region with suppressed recombination (Boyes *et al.*, 1997; Boyes and Nasrallah, 1993; Vieira *et al.*, 2003). Therefore, a feature of the *S*-locus is the occurrence of repetitive sequences and transposable elements in this region. The *S*-locus of *Arabidopsis lyrata*, *Brassica oleracea*, *Ipomoea trifida*, *Antirrhinum*, *Petunia inflata* and *Prunus dulcis* all exhibit a higher TE density than the genomic background (Fujimoto *et al.*, 2006; Goubet *et al.*, 2012; Lai *et al.*, 2002; Tomita *et al.*, 2004; Ushijima *et al.*, 2003; Wang *et al.*, 2004). In *Prunus* there are several reports of TE insertion in the coding region of the pollen *S*-determinant *SFB* alleles resulting in loss of function. For example, in *Prunus cerasus* there is an insertion of a 615-bp Ds-like element into *SFB*₁' (Hauck *et al.*, 2006); in *P. persica* there is a 155-bp direct repeat inserted in *SFB*₁ (Tao *et al.*, 2007); in *P. avium* and *P. mume* there is a 6.8-kp LTR-like TE insertion in *SFB*₁' (Ushijima *et al.*, 2004); a *FaSt* MITE insertion disrupted the open reading frame in *SFB*_C allele in *Prunus* (Halász *et al.*, 2014). In contrast to the impact caused by TE insertions in the coding region of *S*-locus resulting in the loss of SI through deletion or inactivation of *S* determinants (Broz and Bedinger, 2021; Claessen *et al.*, 2019; Nasrallah, 2017), the role of TE insertion in the non-coding region of *S*-locus is far more ambiguous and has rarely been reported before (Kondo *et al.*, 2002a; Tsukamoto *et al.*, 2010). Here, we established that MITE transposons are enriched in the non-coding region of the *S*-locus. Our data provide the first evidence that a MITE insertion in the promoter region of the female *S* allele can disrupt SI by affecting the expression levels of adjacent *S*-RNase alleles, thereby causing a SC phenotype.

The breakdown of SI and shift from SI to SC is one of the most common and frequent evolutionary transitions in flowering plants (Goldberg *et al.*, 2010). Natural selection tends to switch from outcrossing (SI) to self-fertilization (SC) because of the inherent reproductive assurance of SC. Indeed, it has been shown that SI was frequently lost during species differentiation, which may involve duplication or mutation of the *S*-RNase, *SLF* or non-*S*-determinant genes (Broz and Bedinger, 2021; Claessen *et al.*, 2019; Wu *et al.*, 2013). However, our understanding of the mechanisms responsible for SI breakdown is still relatively incomplete. Although it is known that SI status can be influenced by genetic modifiers that are unlinked to the *S*-locus, mutations in the genes comprising the male and female *S* determinants are usually responsible for the loss of SI (Baldwin and Schoen, 2017; Li *et al.*, 2016). In this study, we established that a 786-bp MITE insertion in the promoter of the *S*₂-RNase allele is responsible for inhibited *S*-RNase expression, resulting in the loss of SI in *Fortunella* genera. This seems to be a common occurrence in citrus, since the expression of other *S*-RNase alleles in other intergeneric or interspecific populations are affected in a similar manner by MITE insertions in their promoters. Our data implicate that these transposons may be responsible for the transition from SI to SC and provide an example of parallel evolution in different populations.

Due to the *S*-locus being highly polymorphic, the major *cis*-acting elements and *trans*-acting factors that influence the expression of the *S*-locus are poorly understood. In the *S*-RNase-based SI system, current research has focused in-depth on the function of the *S*-RNase/*SLF* genes and the mechanisms of SI (Broz and Bedinger, 2021; Wu *et al.*, 2013); there have been few studies on the relative expression levels of the *S*-RNase/*SLF* genes since the original studies establishing the *S*-RNase SI system (Hu *et al.*, 2021; McClure *et al.*, 1989; Murfett *et al.*, 1994; Tsukamoto *et al.*, 2003). Previous transgenic studies in tobacco have shown an increase in *S*_{A2}-RNase transcripts was associated with increased style RNase activity (Murfett *et al.*, 1994) and in antisense *S*_{A2}-RNase-transformed plants, a decrease in *S*_{A2}-RNase transcripts was accompanied by decreased style RNase activity (Murfett *et al.*, 1995). RNase activity in style tissue was positively correlated with ability to reject self-pollen (Murfett *et al.*, 1994) and a high level of *S*-RNase expression in style tissues is necessary for functional *S*-RNase-based SI (Lee *et al.*, 1994; Murfett *et al.*, 1994). In this study, we found that the expression level of *S*-RNase in style tissues was correlated with its ability to reject self-pollen. The insertion of the MITE in the promoter region resulted in a low expression level of *S*-RNase in styles and a SC phenotype. Our transgenic data showed that the transcript level of *S*-RNase reached about 75% of the endogenous *S* allele and that this was sufficient for rejection of self-pollen.

We also found that MITEs were significantly enriched around the *S*-RNase alleles in Aurantioideae and were more likely to be inserted in the promoter region of the SC haplotypes. Moreover, this finding has also been confirmed in other GSI families that utilize the *S*-RNase SI system. In three common eudicot families (Solanaceae, Plantaginaceae, Rosaceae), there are marked differences in the occurrence of MITEs adjacent to the *S*-RNase alleles in various SI and SC species; in species known to be SC, MITE insertions were generally found to be located in the promoter region of the *S*-RNase genes, while in the SI species either MITEs were absent or inserted further away from the *S*-RNase. In addition, previous studies have shown that the low expression level of *S*-RNase alleles lead to the loss of SI in sour cherry (*Prunus cerasus*) and cultivated tomato (*Solanum lycopersicum*) (Kondo *et al.*, 2002b; Yamane *et al.*, 2003). Surprisingly, we also found MITE insertions in the promoter region of these two *S*-RNase genes (*PceS_{6m}-RNase* and *SlyS₃-RNase*), which may affect the expression of host genes and mediate the loss of SI. In conclusion, studies on the four common GSI families confirm that MITE-mediated SI loss is a more general phenomenon in *S*-RNase-based SI systems.

It is important to note that not all MITE insertions around *S*-RNase alleles could mediate the loss of SI. We suggest that the insertion positions and types of MITEs may affect expression of *S*-determinant *S*-RNase genes at the *S*-locus, as extensive TE insertions at various locations in *FLC* in *Capsella rubella* has been shown to affect *FLC* expression and led to natural variation in flowering time (Niu *et al.*, 2019). It is possible that the insertion of TEs might introduce new regulatory elements that may alter the expression of host genes (Lisch, 2013; Shen *et al.*, 2017; Xia *et al.*, 2017; Zhang *et al.*, 2020). This clearly deserves further investigation. In summary, our identification of a role for MITEs in citrus and other GSI families significantly improves how the transition from SI to SC can be achieved.

It is known that the fixation and propagation of the SC haplotypes, which have a reproductive assurance in a population increases the frequency of these genotypes and we show that this

happens in SC accessions (Halasz *et al.*, 2007; Tsuchimatsu *et al.*, 2017). Here, we found that multiple nonfunctional SC haplotypes (such as Fhi-*S*₂, Ptr-*S*₃₁, *S*_m and Cre-*S*₃₁ haplotypes) caused by a MITE insertion were maintained in the citrus species-wide loss of SI. All these SC haplotypes observed were associated with insertion of TEs; this is quite different from the loss of SI caused by multiple independent recombination of *S* haplotypes in *Arabidopsis* (Tsuchimatsu *et al.*, 2017). Analysis of the insertion events of different TEs and the statistics of *S* haplotype frequencies demonstrate that these SC haplotypes were not derived from their ancestors, but arose independently after intergeneric differentiation and were then fixed in the population. Similar to *Capsella rubella* (Guo *et al.*, 2011), the SC loci were also under strong positive selection in the population. The widespread apomixis of citrus may also contribute to the fixation and maintenance of multiple independent nonfunctional SC haplotypes in the population (Liang *et al.*, 2020; Wang *et al.*, 2017), which appears to be unique in adaptive evolution. However, the evolutionary relationship between apomixis and SC is still unclear.

In summary, we identified the first evidence for a pivotal role for MITE insertions in the promoter region at the *S*-locus reducing the expression of *S*-RNase alleles, with the consequent breakdown of SI. Our findings are of significance for our understanding of alterations in breeding strategies by SI to SC transitions and the evolution of the *S*-RNase-based SI system in angiosperms. Our comprehensive identification of TEs associated with *S*-locus-genes provide useful data for evolutionary models proposing the maintenance of high polymorphism of *S* loci and the generation mechanism of SC in different populations (Guo *et al.*, 2011; Tsuchimatsu *et al.*, 2017). The loss of SI plays a crucial role in the process of adaptive evolution and since an increasing number of functional TEs are being identified, further effects of TEs on species divergence and domestication are likely to be identified in the future.

Methods

Plant materials and sample extraction

The hybrid population of *F. hindsii* containing approximately 500 individual plants derived from a PNO2 × DBO2 cross was generated in 2013 and was planted in the germplasm garden at the National Citrus Breeding Center at Huazhong Agricultural University. PNO2 (*S*_{8S₁₉) is self-incompatible and DBO2 (*S*_{2S₂₉) is self-compatible. Two individual plants (PDO2-229 and PDO2-482) from this population were used for segregation ratio analysis by genotyping the *S*-locus in the *S*₁ progeny derived from self-pollinated individuals. The 46 accessions of *F. hindsii* that were used in this study were collected from 5 provinces in China. Details on these accessions are provided (Table S4). Floral organs from PNO2 and DBO2 were collected and divided into six tissues (petals, anthers, filaments, styles, ovaries and receptacle) that were frozen immediately in liquid nitrogen and stored at −80 °C for further analysis. Fresh leaf tissues were used for genomic sequencing.}}

Evaluation of self-compatibility

Nearly opened flowers from each individual in the hybrid population were collected. Pollen was obtained by manually separating and drying the anthers at 28 °C for 12 h. At the flowering stage, approximately 10 flowers (1 day before anthesis) were self-pollinated on each individual plant using a

fine paint brush and covered with a paper pollination bag. The pollinated styles were collected after 48 h. Aniline blue staining was performed as described by Liang *et al.*, (2017). In addition, for the visualization of pollen tube growth in hardy orange, mandarin and Australian limes, pollinated styles were collected after 96–120 h. Selected styles were imaged using Inverted Microscopes for Industry Leica DMI8 M. The images were made using the associated software package (Public Experimental Platform of Key Laboratory of Horticultural Plant Biology, HZAU, MOE).

More than 5 carpels were taken from each individual for observation. The carpels were scored as SC if more than 50% of the pollen tubes reached the ovary and are indicated with a '√'. If the pollen tubes stopped growing in the upper third of the stigma, the carpel was scored as incompatible and indicated with an '×'. Other phenotypes were scored as uncertain and indicated with a '-'. If 3 or more styles were compatible, the individual was scored as self-compatible. If all styles were scored as incompatible, the individual was scored as self-incompatible. In other cases, the individual was scored as undetermined.

Bulked segregant analysis sequencing (BSA-Seq)

PNO2 was crossed with DBO2 to generate F₁ individuals that segregated the SC and SI phenotypes at a 1:1 ratio. DNA was extracted from pools of tissue collected from 32 self-incompatible F₁ individuals and 32 self-compatible F₁ individuals. Two 300-bp libraries were constructed to generate 150-bp paired-end short reads (90×) using the BGISEQ-500 platform (BGI-Shenzhen, China). Clean reads were obtained by quality evaluation using fastp (v. 0.23.1) (Chen *et al.*, 2018). Subsequently, the filtered short reads were aligned against the reference genome (*S*3y-45 v2.0) (Wang *et al.*, 2022). The variant calling of SNPs was excavated using the GATK (v. 4.1.2.0) (McKenna *et al.*, 2010) software. The heterozygous and inconsistent SNPs between the two DNA pools were selected by calculating *G'* values using the *G* prime algorithm in the QTLseqr package (Li and Xu, 2021; Mansfeld and Grumet, 2018).

Kompetitive allele-specific PCR (KASP) analysis

KASP genotyping reactions (KASP-TF V4.0 2 × Master Mix 384, Low ROX) for allelic discrimination was performed on a Roche LightCycler® 480 II (Roche, Switzerland) using KASP assays designed to be specific for SNPs from the target region. KASP assays were conducted according to the protocol supplied by the manufacturer (LGC, Laboratory of the Government Chemist, UK) with FAM (Excitation: 465 nm; Emission: 510 nm) and HEX (Excitation: 533 nm; Emission: 580 nm). The genotyping results after the standard KASP thermal cycle were analysed and visualized using the lightCycler® software (Endpoint Genotyping program) to confirm correct segregation and genotype. KASP markers used in this article are listed in Table S17.

Identification of selection signatures

For the selection scan, we used the imputed SNP data set of 46 accessions (Table S8) that was controlled using VCFtools (v. 0.1.16) (Danecek *et al.*, 2011). The integrated haplotype score (iHS) test was performed to investigate selection signatures of SC-associated loci in the SC *F. hindsii* population. The iHS statistic compares the Extended haplotype homozygosity (EHH) between two alleles by controlling for the allele frequency of each SNP. The R library rehh (Gautier and Vitalis, 2012) was used to calculate EHH and iHS statistic. Cross-Population Extended Haplotype

Homozygosity (XP-EHH) statistics based on the extended haplotype was calculated for SC and SI *F. hindsii* populations (Table S8) using selscan (v. 2.0.0) (Szpiech and Hernandez, 2014). The SC *F. hindsii* was defined as observed population, and the SI *F. hindsii* was the reference population. Significant genomic regions were identified by *P*-value <0.001.

Quantitative RT-PCR analysis

Total RNA was extracted with RNAiso Plus (TaKaRa, Japan) according to the manufacturer's instructions. cDNA was synthesized with HiScript® II Q RT SuperMix for qPCR (+gDNA wiper) (Vazyme Biotech, China) according to the supplier's protocol. RT-PCR was conducted as described by Hu *et al.* (2021). All samples were run in triplicate, and there were two separate runs for each plate. For qRT-PCR, expression of β -actin was used as an internal control for normalization. The relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method. The pertinent primers are listed in Table S17.

GUS assays

The cloned promoter fragments (*FhiS₂-RNase_{PRO}*, 2689 bp; *FhiS₂-RNase_{PROΔMITE}*, 1903 bp) were constructed in pKGWFS7 using the Gateway method (Invitrogen, USA). The recombinant plasmids were transformed into DB02 callus that had been subcultured for 21 days in the dark using the *Agrobacterium*-mediated method. For GUS staining, callus was submerged in X-gluc buffer (X-Gluc, 50 mg; 50 mM/L, pH 7.0, PBS, 80 mL; 20% methanol, 20 mL; 0.5 M/L EDTA, 2 mL; 1% Triton-100, 1 mL) overnight at 37 °C, de-stained with 70% ethanol and photographed.

Dual-luciferase assays

The promoter fragments (*FhiS₂-RNase_{PROΔMITE}*, 1903 bp; *PtrS₂-RNase_{PRO+MITE}*, 2688 bp; *FhiS₂-RNase_{PRO}*, 2689 bp; *PtrS₂-RNase_{PRO}*, 1902 bp) were inserted into the pGreenII 0800-LUC vector to drive the expression of the firefly luciferase (*LUC*) gene using the ClonExpress II One Step Cloning Kit (Vazyme, China). The recombinant plasmids were transformed into DB02 callus and cocultured at 21 °C for 3 days in constant darkness on MT-Acetosyringone (AS) medium (4.46 g/L Murashige and Tucker Medium [PhytoTechnology Laboratories, LLC], 40 g/L sucrose, 8 g/L agar, 20 mg/L AS, pH 5.8). The activity of LUC and REN was measured using the Dual-Luciferase® Reporter Assay System (Promega, USA). Promoter activity was quantified as the ratio of LUC/REN with 12 replicates. Relevant primers are listed in Table S17.

Agrobacterium-mediated genetic transformation

The binary plasmid vectors pK7WG2D.1 containing p35S::*FhiS₂-RNase* and DsRED2 containing p*FhiS_{2g}*::*FhiS₂-RNase* and p*FhiS_{2ΔMITE}*::*FhiS₂-RNase* were provided by Prof. Jin (Sun *et al.*, 2018) and were transferred into *Agrobacterium tumefaciens* strain EHA105. Surface sterilized mature seeds were inoculated on MT medium and cultured at 25 °C in the dark. After 6–8 weeks, the seedlings were transferred to a photoperiod containing 16 h of light and 8 h of dark for 14 days. After chlorophyll visibly accumulated, the seedlings were cut into 1–1.5 cm long stem segments that were used for transformation experiments. *Agrobacterium*-mediated genetic transformation of *F. hindsii* was performed as described by Zhu *et al.* (2019). The transgenic lines were identified using PCR with specific primers (Table S17).

Genomic and transcriptomic sequencing

The sequencing of PNO2 and DB02 was performed using the PacBio Sequel II platform (Pacific Biosciences, USA) at Beijing Biomarker Technologies Co. Ltd. (Beijing, China). A SMRTbell library with a size of 30 kb was constructed for genomic sequencing following the standard protocol provided by Pacific Biosciences. Two 300-bp libraries were constructed to survey the genome information using the BGISEQ-500 platform (BGI-Shenzhen, China). Total RNA was extracted from the anther and style tissues of PNO2 and DB02. Paired-end libraries were constructed and sequenced on BGISEQ-500 platform (BGI-Shenzhen, China). Three biological replicates were analysed.

RNA-seq analysis

The anther and style tissues of PNO2 and DB02 were collected 1 day before anthesis. Clean reads were mapped to the reference genome (S3y-45 v2.0) using Hisat2 (Kim *et al.*, 2019) with default parameters. The TPM (transcripts per kilobase per million mapped reads) of each gene and the DEGs (fold change ≥ 2 , FDR <0.01) between PNO2 and DB02 were calculated (Love *et al.*, 2014). The reproducibility of transcriptome samples was assessed using a Spearman rank correlation.

Whole-genome bisulphite sequencing (WGBS) and analysis

BS-seq libraries from leaves, anthers and style tissues of DB02 were constructed and sequenced using 150-bp reads from the BGISEQ-500 platform (BGI-Shenzhen, China). The clean reads were aligned to the DB02gv1 contig genome using Bismark (v. 0.22.2) (Krueger and Andrews, 2011), allowing up to three mismatches. Only cytosines that were covered by at least four reads were considered and used for the calculation. The methylation levels of individual cytosines were calculated as the ratio of mC to the total cytosines [mC/(mC + un-mC)].

Small RNA transcriptome analysis

Small RNAs were extracted from anther and style tissues 1 day before anthesis using RNAiso Plus (TaKaRa) according to the manufacturer's instructions. Then the 18–30 nt small RNAs were excised and ligated to adenylated 3' adapters annealed to unique molecular identifiers (UMI), followed by the ligation of 5' adapters. Library preparation and single-end sequencing were conducted using the DNBSEQ platform (BGI-Shenzhen, China). After filtering, the clean reads were mapped to the *F. hindsii* genome (S3y-45 v2.0) or within the *S*-locus in DB02 using Bowtie (Langmead *et al.*, 2009) (v. 1.2.3) with the following parameters: a -m 20 -v 0 -p 1. The remaining parameters were set at default values. Mapped reads were extracted using the SAMtools (Li *et al.*, 2009) (v. 1.14) suite with default parameters. The alignment results were visualized using the Integrated Genome Viewer (Robinson *et al.*, 2011) (v. 2.11.3). The RMP (reads per million) of siRNAs in the target region between anther and style were calculated.

Genome assembly and S haplotype reconstruction

The CLR reads were assembled using Canu (Koren *et al.*, 2017) (v2.1.1) with the following parameters: --pacbio (minReadLength = 2000, minOverlapLength = 500, corOutCoverage = 150, corMinCoverage = 2 'batOptions = -dg 3 -db 3 -dr 1 -ca 500 -cp 50', corrected error rate = 0.035). The reads were polished iteratively

using two rounds of Pilon (Walker *et al.*, 2014) (v. 1.24) with ~12 Gb of WGS BGISEQ data to generate the draft genome (PN02gv1 contig, DB02gv1 contig and AZMgv1 contig). The sequences of *S* haplotypes were identified from genomes using the *S*-*RNase* alleles. The Fhi-*S*₈-locus, Fhi-*S*₁₉-locus, Fhi-*S*₂-locus, Fhi-*S*₂₉-locus, Cau-*S*₈-locus and Cau-*S*₃₇-locus were obtained using the 200-bp sequences from two non-recombinant KASP markers (K961397 and K1247997). Gene predictions and annotations for each *S*-locus were made using FGENESH (Solovyev *et al.*, 2006) and nonredundant protein sequence (nr) (Pruitt *et al.*, 2007) databases. The sequences of the other *S*-locus were obtained from the published genome (see Table S14 for more details). The syntenic regions of the *S*-locus were identified using the all-vs-all BLASTP (e-value = 1e-5) method with a threshold value of 0.95.

Whole-sequence alignments and structural variation analysis

Pairwise alignments of all *S*-locus were performed using the NUCmer program in MUMmer (Marcais *et al.*, 2018) (v.4.0.0). The Fhi-*S*₂-locus in the S3y-45 v2.0 genome (*F. hindsii*) were used as a reference. The filtering parameters were as follows: 'Delta-filter-i 90-L 200-1'. The filtered data were plotted using Mummerplot. Show-diff and show-SNPs were used for structural variation (SVs, > 50 bp). show-SNPs in MUMmer was used to detect SNPs between the reference and query sequences. The parameters were as follows: '-c-H-I-T-r-L'. The online website Assemblytics (Nattestad and Schatz, 2016) (<http://assemblytics.com/>) was used to analyse and visualize structural variation.

Isolation and identification of MITEs

The *S* haplotypes in the genomes were identified by BLAST alignment with the *S*-*RNase* allele sequences (see Table S14 for more details), and genomic fragments covering the entire *S*-*RNase* gene, the 5-kb 5'-flanking region and the 5-kb 3'-flanking region were obtained based on the position of *S* alleles. The genomic fragments sequences data can be found in the GenBank data libraries under accession numbers (Table S15). The MITE annotation was performed using MITE-Hunter (Han and Wessler, 2010), Repeatmasker Web Server (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>) and the P-MITE database (Chen *et al.*, 2014).

Acknowledgements

This research was financially supported by the National Key Research and Development Programme of China (grant no. 2021YFD1200200), the National Natural Science Foundation of China (grant nos. 32122075, 32072523, 32270357, 32302489), Key-Area Research and Development Program of Guangdong Province (grant no. 2022B0202070002), Hubei Provincial Natural Science Foundation of China (grant no. 2023AFB094), the China Agriculture Research System (grant no. CARS-26). The work was also supported by the Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/T00486X/1 to MB and VEF-T.

Author contributions

JH and LC conceived the project. JH conducted KASP experiments and performed BSA-Seq, transcriptomic, genomic, siRNAs and WGBS sequencing and data acquisition. ZD conducted the genome assembly and MITE annotation. CL, FG and DS

conducted PCR, GUS and LUC assays. NW, SZ, CZ, PC and QX provided the plant materials. ZW, JJ, ZC and CS collected the plant samples and conducted pollination assays. XD and LC supervised the project and revised the manuscript. JH wrote the manuscript with the help of JY, RL, ZL, MB and VEF-T. All authors contributed to the article and approved the manuscript before submission.

Conflict of interest

The authors declare no conflicts of interest.

Data availability statement

All sequence data have been deposited at the NCBI under project number PRJNA860018, PRJNA827463, PRJNA827498. The genome assembly have been deposited in the CNGB Sequence Archive (CNSA) of the China National GenBank DataBase (CNGBdb) with accession numbers CNP0003278. The RNA sequencing data from anther and style tissues are available on the NCBI short read archive under accession PRJNA827480. The BSA-Seq data has been deposited in NCBI under project number PRJNA827489. Other data are available in the source data file or will be made available upon request. Source data are provided with this paper. *S*-*RNase* and *SLF* sequences data from this article can be found in the GenBank data libraries under accession numbers (see Table S13, S15, S16 and S18).

References

- Baldwin, S.J. and Schoen, D.J. (2017) Genetic variation for pseudo-self-compatibility in self-incompatible populations of *Leavenworthia alabamica* (Brassicaceae). *New Phytol.* **213**, 430–439.
- Barrett, S.C. (2002) The evolution of plant sexual diversity. *Nat. Rev. Genet.* **3**, 274–284.
- Bennetzen, J.L. and Wang, H. (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* **65**, 505–530.
- Borges, F. and Martienssen, R.A. (2015) The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* **16**, 727–741.
- Boyes, D.C. and Nasrallah, J.B. (1993) Physical Linkage of the *Slg* and *Srk* Genes at the Self-Incompatibility Locus of Brassica-Oleracea. *Mol. Gen. Genet.* **236**, 369–373.
- Boyes, D.C., Nasrallah, M.E., Vrebalov, J. and Nasrallah, J.B. (1997) The self-incompatibility (*S*) haplotypes of Brassica contain highly divergent and rearranged sequences of ancient origin. *Plant Cell* **9**, 237–247.
- Broz, A.K. and Bedinger, P.A. (2021) Pollen-Pistil Interactions as Reproductive Barriers. *Annu. Rev. Plant Biol.* **72**, 615–639.
- Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G. *et al.* (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell*, **24**, 1242–1255.
- Caruso, M., Merelo, P., Distefano, G., La Malfa, S., Lo Piero, A.R., Tadeo, F.R., Talon, M. *et al.* (2012) Comparative transcriptome analysis of stylar canal cells identifies novel candidate genes implicated in the self-incompatibility response of *Citrus clementina*. *BMC Plant Biol.* **12**, 20.
- Castric, V. and Vekemans, X. (2004) Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. *Mol. Ecol.* **13**, 2873–2889.
- Chai, L., Ge, X., Biswas, M.K., Xu, Q. and Deng, X. (2010) Self-sterility in the mutant 'Zigui shatian' pummelo (*Citrus grandis* Osbeck) is due to abnormal post-zygotic embryo development and not self-incompatibility. *Plant Cell, Tissue Organ Cult.* **104**, 1–11.

- Charlesworth, D. and Willis, J.H. (2009) The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–796.
- Charlesworth, D. and Wright, S.I. (2001) Breeding systems and genome evolution. *Curr. Opin. Genet. Dev.* **11**, 685–690.
- Chen, J., Hu, Q., Zhang, Y., Lu, C. and Kuang, H. (2014) P-MITE: a database for plant miniature inverted-repeat transposable elements. *Nucleic Acids Res.* **42**, D1176–D1181.
- Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**, i884–i890.
- Chuong, E.B., Elde, N.C. and Feschotte, C. (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86.
- Claessen, H., Keulemans, W., Van de Poel, B. and De Storme, N. (2019) Finding a compatible partner: Self-incompatibility in European pear (*Pyrus communis*); molecular control, genetic determination, and impact on fertilization and fruit set. *Front. Plant Sci.* **10**, 407.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E. et al. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
- Du, J., Ge, C., Li, T., Wang, S., Gao, Z., Sassa, H. and Qiao, Y. (2021) Molecular characteristics of S-RNase alleles as the determinant of self-incompatibility in the style of *Fragaria viridis*. *Hortic Res.* **8**, 185.
- Eggers, E.J., van der Burgt, A., van Heusden, S.A.W., de Vries, M.E., Visser, R.G.F., Bachem, C.W.B. and Lindhout, P. (2021) Neofunctionalisation of the Sli gene leads to self-compatibility and facilitates precision breeding in potato. *Nat. Commun.* **12**, 4141.
- Feschotte, C., Jiang, N. and Wessler, S.R. (2002) Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* **3**, 329–341.
- Fujii, S., Kubo, K. and Takayama, S. (2016) Non-self- and self-recognition models in plant self-incompatibility. *Nat Plants* **2**, 16130.
- Fujimoto, R., Okazaki, K., Fukai, E., Kusaba, M. and Nishio, T. (2006) Comparison of the genome structure of the self-incompatibility (S) locus in interspecific pairs of S haplotypes. *Genetics* **173**, 1157–1167.
- Gambetta, G., Gravina, A., Fasiolo, C., Fornero, C., Galiger, S., Inzaurrealde, C. and Rey, F. (2013) Self-incompatibility, parthenocarpy and reduction of seed presence in ‘Afourer’ mandarin. *Sci. Hortic.* **164**, 183–188.
- Gautier, M. and Vitalis, R. (2012) rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*, **28**, 1176–1177.
- Goldberg, E.E., Kohn, J.R., Lande, R., Robertson, K.A., Smith, S.A. and Igic, B. (2010) Species selection maintains self-incompatibility. *Science* **330**, 493–495.
- Goubet, P.M., Berges, H., Bellec, A., Prat, E., Helmstetter, N., Mangenot, S., Gallina, S. et al. (2012) Contrasted patterns of molecular evolution in dominant and recessive self-incompatibility haplotypes in Arabidopsis. *PLoS Genet.* **8**, e1002495.
- Guo, Y.L., Zhao, X., Lanz, C. and Weigel, D. (2011) Evolution of the S-locus region in Arabidopsis relatives. *Plant Physiol.* **157**, 937–946.
- Halasz, J., Pedryc, A. and Hegedus, A. (2007) Origin and dissemination of the pollen-part mutated SC haplotype which confers self-compatibility in apricot (*Prunus armeniaca*). *New Phytol.* **176**, 792–803.
- Halász, J., Kodad, O. and Hegedűs, A. (2014) Identification of a recently active Prunus-specific non-autonomous Mutator element with considerable genome shaping force. *Plant J.* **79**, 220–231.
- Han, Y. and Wessler, S.R. (2010) MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, e199.
- Hauck, N.R., Ikeda, K., Tao, R. and Iezzoni, A.F. (2006) The mutated S1-haplotype in sour cherry has an altered S-haplotype-specific F-box protein gene. *J. Hered.* **97**, 514–520.
- Honsho, C., Kotsubo, M., Fukuda, Y., Hamabata, Y., Kurogi, Y., Nishiwaki, A. and Tetsumura, T. (2009) Reproductive Characteristics for Self-compatibility and Seedlessness in ‘Nishiuchi Konatsu’, a Bud Mutation of Hyuganatsu (*Citrus tamurana* hort. ex Tanaka). *Hortscience*, **44**, 1547–1551.
- Hu, J., Xu, Q., Liu, C., Liu, B., Deng, C., Chen, C., Wei, Z. et al. (2021) Downregulated expression of S2-RNase attenuates self-incompatibility in “Guiyou No. 1” pummelo. *Hortic Res.* **8**, 199.
- Huang, S., Lee, H.S., Karunanandaa, B. and Kao, T.H. (1994) Ribonuclease activity of *Petunia inflata* S proteins is essential for rejection of self-pollen. *Plant Cell*, **6**, 1021–1028.
- Igic, B. and Kohn, J.R. (2001) Evolutionary relationships among self-incompatibility RNases. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13167–13171.
- Igic, B., Lande, R. and Kohn, J.R. (2008) Loss of Self-Incompatibility and Its Evolutionary Consequences. *International Journal of Plant Sciences* **169**, 93–104.
- Kakade, V., Dubey, A.K., Sharma, R.M. and Malik, S.K. (2017) Gametophytic self-incompatibility causes seedlessness in ‘Kagzi Kalan’ lemon (*Citrus limon*). *J. Hortic Sci. Biotech.* **92**, 303–312.
- Kent, T.V., Uzunovic, J. and Wright, S.I. (2017) Coevolution between transposable elements and recombination. *Philos. Trans. R. Soc. London B Biol. Sci.* **372**, 20160458.
- Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915.
- Kondo, K., Yamamoto, M., Itahashi, R., Sato, T., Egashira, H., Hattori, T. and Koyama, Y. (2002a) Insights into the evolution of self-compatibility in Lycopersicon from a study of stylar factors. *Plant J.* **30**, 143–153.
- Kondo, K., Yamamoto, M., Matton, D.P., Sato, T., Hirai, M., Norioka, S., Hattori, T. et al. (2002b) Cultivated tomato has defects in both S-RNase and HT genes required for stylar function of self-incompatibility. *Plant J.* **29**, 627–636.
- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736.
- Krueger, F. and Andrews, S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572.
- Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549.
- Lai, Z., Ma, W.S., Han, B., Liang, L.Z., Zhang, Y.S., Hong, G.F. and Xue, Y.B. (2002) An F-box gene linked to the self-incompatibility (S) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. *Plant Mol. Biol.* **50**, 29–42.
- Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.
- Lee, H.S., Huang, S. and Kao, T. (1994) S proteins control rejection of incompatible pollen in *Petunia inflata*. *Nature*, **367**, 560–563.
- Li, Z. and Xu, Y. (2021) Bulk segregation analysis in the NGS era: a review of its teenage years. *Plant J.* **109**, 1355–1374.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Li, W., Yang, Q., Gu, Z., Wu, C., Meng, D., Yu, J., Chen, Q. et al. (2016) Molecular and genetic characterization of a self-compatible apple cultivar, ‘CAU-1’. *Plant Sci.* **252**, 162–175.
- Li, Z.W., Hou, X.H., Chen, J.F., Xu, Y.C., Wu, Q., Gonzalez, J. and Guo, Y.L. (2018) Transposable Elements Contribute to the Adaptation of Arabidopsis thaliana. *Genome Biol. Evol.* **10**, 2140–2150.
- Li, M., Zhang, D., Gao, Q., Luo, Y., Zhang, H., Ma, B., Chen, C. et al. (2019) Genome structure and evolution of *Antirrhinum majus* L. *Nat Plants*, **5**, 174–183.
- Liang, M., Yang, W., Su, S., Fu, L., Yi, H., Chen, C., Deng, X. et al. (2017) Genome-wide identification and functional analysis of S-RNase involved in the self-incompatibility of citrus. *Mol. Genet. Genomics*, **292**, 325–341.
- Liang, M., Cao, Z., Zhu, A., Liu, Y., Tao, M., Yang, H., Xu, Q., Jr. et al. (2020) Evolution of self-compatibility by a mutant Sm-RNase in citrus. *Nat Plants*, **6**, 131–142.
- Lisch, D. (2013) How important are transposons for plant evolution? *Nat. Rev. Genet.* **14**, 49–61.
- Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
- Lu, C., Chen, J., Zhang, Y., Hu, Q., Su, W. and Kuang, H. (2012) Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Mol. Biol. Evol.* **29**, 1005–1017.
- Ma, L., Zhang, C., Zhang, B., Tang, F., Li, F., Liao, Q., Tang, D. et al. (2021) A nonS-locus F-box gene breaks self-incompatibility in diploid potatoes. *Nat. Commun.* **12**, 4142.

- Mansfeld, B.N. and Grumet, R. (2018) QTLseqr: An R package for bulk segregant analysis with next-generation sequencing. *Plant Genome*, **11**, 1–5.
- Mao, H., Wang, H., Liu, S., Li, Z., Yang, X., Yan, J., Li, J. et al. (2015) A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nat. Commun.* **6**, 8326.
- Marcais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L. and Zimin, A. (2018) MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944.
- McClure, B.A., Haring, V., Ebert, P.R., Anderson, M.A., Simpson, R.J., Sakiyama, F. and Clarke, A.E. (1989) Style self-incompatibility gene products of *Nicotiana glauca* are ribonucleases. *Nature* **342**, 955–957.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K. et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.
- Murfett, J., Atherton, T.L., Mou, B., Gasser, C.S. and McClure, B.A. (1994) S-RNase expressed in transgenic *Nicotiana glauca* causes S-allele-specific pollen rejection. *Nature* **367**, 563–566.
- Murfett, J., Bourque, J.E. and McClure, B.A. (1995) Antisense suppression of S-RNase expression in *Nicotiana glauca* using RNA-Polymerase-II-transcribed and RNA-polymerase-III-transcribed gene constructs. *Plant Mol. Biol.* **29**, 201–212.
- Nasrallah, J.B. (2017) Plant mating systems: self-incompatibility and evolutionary transitions to self-fertility in the mustard family. *Curr. Opin. Genet. Dev.* **47**, 54–60.
- Nattestad, M. and Schatz, M.C. (2016) Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics* **32**, 3021–3023.
- Niu, X.M., Xu, Y.C., Li, Z.W., Bian, Y.T., Hou, X.H., Chen, J.F., Zou, Y.P. et al. (2019) Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 6908–6913.
- Okada, K., Tonaka, N., Moriya, Y., Norioka, N., Sawamura, Y., Matsumoto, T., Nakanishi, T. et al. (2008) Deletion of a 236 kb region around S4-RNase in a stylar-part mutant S4sm-haplotype of Japanese pear. *Plant Mol. Biol.* **66**, 389–400.
- Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–D65.
- Ramanauskas, K. and Igic, B. (2021) RNase-based self-incompatibility in cacti. *New Phytol.* **231**, 2039–2049.
- Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26.
- Sassa, H., Kakui, H., Miyamoto, M., Suzuki, Y., Hanada, T., Ushijima, K., Kusaba, M. et al. (2007) S locus F-box brothers: multiple and pollen-specific F-box genes with S haplotype-specific polymorphisms in apple and Japanese pear. *Genetics* **175**, 1869–1881.
- Serrato-Capuchina, A. and Matute, D.R. (2018) The role of transposable elements in speciation. *Genes (Basel)* **9**, 254.
- Shen, J., Liu, J., Xie, K., Xing, F., Xiong, F., Xiao, J., Li, X. et al. (2017) Translational repression by a miniature inverted-repeat transposable element in the 3' untranslated region. *Nat. Commun.* **8**, 14651.
- Shimada, T., Endo, T., Fujii, H., Nakano, M., Sugiyama, A., Daido, G., Ohta, S. et al. (2018) MITE insertion-dependent expression of CitRKD1 with a RWP-RK domain regulates somatic embryogenesis in citrus nucellar tissues. *BMC Plant Biol.* **18**, 166.
- Sijacic, P., Wang, X., Skirpan, A.L., Wang, Y., Dowd, P.E., McCubbin, A.G., Huang, S. et al. (2004) Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature* **429**, 302–305.
- Slotte, T., Hazzouri, K.M., Agren, J.A., Koenig, D., Maumus, F., Guo, Y.L., Steige, K. et al. (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* **45**, 831–835.
- Solovyev, V., Kosarev, P., Seledsov, I. and Vorobyev, D. (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* **7 Suppl 1**, S10.1–S10.12.
- Sun, L., Alariqi, M., Zhu, Y., Li, J., Li, Z., Wang, Q., Li, Y. et al. (2018) Red fluorescent protein (DsRed2), an ideal reporter for cotton genetic transformation and molecular breeding. *Crop J.* **6**, 366–376.
- Szpiech, Z.A. and Hernandez, R.D. (2014) selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824–2827.
- Takayama, S. and Isogai, A. (2005) Self-incompatibility in plants. *Annu. Rev. Plant Biol.* **56**, 467–489.
- Tao, R., Watari, A., Hanada, T., Habu, T., Yaegaki, H., Yamaguchi, M. and Yamane, H. (2007) Self-compatible peach (*Prunus persica*) has mutant versions of the S haplotypes found in self-incompatible *Prunus* species. *Plant Mol. Biol.* **63**, 109–123.
- Tomita, R.N., Suzuki, G., Yoshida, K., Yano, Y., Tsuchiya, T., Kakeda, K., Mukai, Y. et al. (2004) Molecular characterization of a 313-kb genomic region containing the self-incompatibility locus of *Ipomoea trifida*, a diploid relative of sweet potato. *Breed. Sci.* **54**, 165–175.
- Tsuchimatsu, T., Goubet, P.M., Gallina, S., Holl, A.C., Fobis-Loisy, I., Berges, H., Marande, W. et al. (2017) Patterns of polymorphism at the self-incompatibility locus in 1083 arabidopsis thaliana genomes. *Mol. Biol. Evol.* **34**, 1878–1889.
- Tsakamoto, T., Ando, T., Takahashi, K., Omori, T., Watanabe, H., Kokubun, H., Marchesi, E. et al. (2003) Breakdown of self-incompatibility in a natural population of *Petunia axillaris* caused by loss of pollen function. *Plant Physiol.* **131**, 1903–1912.
- Tsakamoto, T., Hauck, N.R., Tao, R., Jiang, N. and lezzoni, A.F. (2010) Molecular and genetic analyses of four nonfunctional S haplotype variants derived from a common ancestral S haplotype identified in sour cherry (*Prunus cerasus* L.). *Genetics* **184**, 411–427.
- Ushijima, K., Sassa, H., Dandekar, A.M., Gradziel, T.M., Tao, R. and Hirano, H. (2003) Structural and transcriptional analysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. *Plant Cell*, **15**, 771–781.
- Ushijima, K., Yamane, H., Watari, A., Kakehi, E., Ikeda, K., Hauck, N.R., lezzoni, A.F. et al. (2004) The S haplotype-specific F-box protein gene, SFB, is defective in self-compatible haplotypes of *Prunus avium* and *P. mume*. *Plant J.* **39**, 573–586.
- Vieira, C.P., Charlesworth, D. and Vieira, J. (2003) Evidence for rare recombination at the gametophytic self-incompatibility locus. *Heredity (Edinb)* **91**, 262–267.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A. et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, **9**, e112963.
- Wang, Y., Wang, X., Skirpan, A.L. and Kao, T.H. (2003) S-RNase-mediated self-incompatibility. *J. Exp. Bot.* **54**, 115–122.
- Wang, Y., Tsakamoto, T., Yi, K.W., Wang, X., Huang, S., McCubbin, A.G. and Kao, T.H. (2004) Chromosome walking in the *Petunia inflata* self-incompatibility (S-) locus and gene identification in an 881-kb contig containing S2-RNase. *Plant Mol. Biol.* **54**, 727–742.
- Wang, X., Xu, Y., Zhang, S., Cao, L., Huang, Y., Cheng, J., Wu, G. et al. (2017) Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nat. Genet.* **49**, 765–772.
- Wang, N., Song, X., Ye, J., Zhang, S., Cao, Z., Zhu, C., Hu, J. et al. (2022) Structural variation and parallel evolution of apomixis in citrus during domestication and diversification. *Natl. Sci. Rev.* **9**, nwac114.
- Wright, S. (1939) The distribution of self-sterility alleles in populations. *Genetics* **24**, 538–552.
- Wu, J., Gu, C., Khan, M.A., Wu, J., Gao, Y., Wang, C., Korban, S.S. et al. (2013) Molecular determinants and mechanisms of gametophytic self-incompatibility in fruit trees of rosaceae. *Crit. Rev. Plant Sci.* **32**, 53–68.
- Wu, G.A., Terol, J., Ibanez, V., Lopez-Garcia, A., Perez-Roman, E., Borreda, C., Domingo, C. et al. (2018) Genomics of the origin and evolution of Citrus. *Nature*, **554**, 311–316.
- Wu, G.A., Sugimoto, C., Kinjo, H., Azama, C., Mitsube, F., Talon, M., Gmitter, F.G., Jr. et al. (2021) Diversification of mandarin citrus by hybrid speciation and apomixis. *Nat. Commun.* **12**, 4377.
- Wu, N., Yao, Y., Xiang, D., Du, H., Geng, Z., Yang, W., Li, X. et al. (2022) A MITE variation-associated heat-inducible isoform of a heat-shock factor confers heat tolerance through regulation of JASMONATE ZIM-DOMAIN genes in rice. *New Phytol.* **234**, 1315–1331.
- Xia, C., Zhang, L., Zou, C., Gu, Y., Duan, J., Zhao, G., Wu, J. et al. (2017) A TRIM insertion in the promoter of Ms2 causes male sterility in wheat. *Nat. Commun.* **8**, 15407.

- Xu, L., Yuan, K., Yuan, M., Meng, X., Chen, M., Wu, J., Li, J. *et al.* (2020) Regulation of rice tillering by RNA-directed DNA methylation at miniature inverted-repeat transposable elements. *Mol. Plant* **13**, 851–863.
- Yamane, H., Ikeda, K., Hauck, N.R., Iezzoni, A.F. and Tao, R. (2003) Self-incompatibility (S) locus region of the mutated S6-haplotype of sour cherry (*Prunus cerasus*) contains a functional pollen S allele and a non-functional pistil S allele. *J. Exp. Bot.* **54**, 2431–2437.
- Yamashita, K. and Tanimoto, S. (1985) Studies on Self-incompatibility of Hassaku (*Citrus hassakuhort.* ex Tanaka). *J Jpn Soc Hort Sci* **54**, 178–183.
- Ye, W., Qin, Y., Ye, Z., Silva, J.A.T.d., Zhang, L., Wu, X., Lin, S. *et al.* (2009) Seedless mechanism of a new mandarin cultivar 'Wuzhishatangju' (*Citrus reticulata* Blanco). *Plant Sci.* **177**, 19–27.
- Zhang, S., Ding, F., He, X., Luo, C., Huang, G. and Hu, Y. (2015) Characterization of the 'Xiangshui' lemon transcriptome by de novo assembly to discover genes associated with self-incompatibility. *Mol. Genet. Genomics*, **290**, 365–375.
- Zhang, H., Tao, Z., Hong, H., Chen, Z., Wu, C., Li, X., Xiao, J. *et al.* (2016) Transposon-derived small RNA is responsible for modified function of WRKY45 locus. *Nat. Plants*, **2**, 16016.
- Zhang, H., Lang, Z. and Zhu, J.K. (2018) Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **19**, 489–506.
- Zhang, K., Su, J., Xu, M., Zhou, Z., Zhu, X., Ma, X., Hou, J. *et al.* (2020) A common wild rice-derived BOC1 allele reduces callus browning in indica rice transformation. *Nat. Commun.* **11**, 443.
- Zhao, H., Zhang, Y., Zhang, H., Song, Y., Zhao, F., Zhang, Y., Zhu, S. *et al.* (2022) Origin, loss, and regain of self-incompatibility in angiosperms. *Plant Cell* **34**, 579–596.
- Zheng, X., Zhu, K., Sun, Q., Zhang, W., Wang, X., Cao, H., Tan, M. *et al.* (2019) Natural variation in CCD4 promoter underpins species-specific evolution of red coloration in citrus peel. *Mol. Plant* **12**, 1294–1307.
- Zhu, C., Zheng, X., Huang, Y., Ye, J., Chen, P., Zhang, C., Zhao, F. *et al.* (2019) Genome sequencing and CRISPR/Cas9 gene editing of an early flowering Mini-Citrus (*Fortunella hindsii*). *Plant Biotechnol. J.* **17**, 2199–2210.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

- Figure S1** Fluorescence images of pollen tubes in pistils of a *F. hindsii* cross-pollination population (PN02 × DB02).
- Figure S2.** A signature of selection at the S loci.
- Figure S3** Nucleotide and amino acid sequence alignment of *S*₂-RNase in different intergeneric or interspecific of citrus.
- Figure S4** Analysis of *S*-RNase alleles expression in style tissues of DB02, PN02, and ZK8.
- Figure S5** Nucleotide sequence alignment analysis of the *S*₂-RNase promoter between *F. hindsii* and *P. trifoliata*.
- Figure S6** Methylation status of *FhiS*₂₉-RNase in the anther and style of *F. hindsii*.
- Figure S7** Transgenic assays with the *FhiS*₂-RNase gene from *F. hindsii*.
- Figure S8** Phylogenetic tree analysis of SLFs in citrus.
- Figure S9** TPM values for the *SLF* and *S*-RNase alleles from the *F. hindsii* S-locus.
- Figure S10** Nucleotide and amino acid sequence alignments of *S*₃₇-RNase from different intergeneric and interspecific of citrus.

Figure S11 Nucleotide sequences alignment of *S*₃₇-RNase promoter regions from SC *P. trifoliata*, SI *C. australasia*, SI *C. ichangensis*, and SC *C. reticulata*.

Figure S12 Expression of *S*-RNase in the style tissues of SC mandarins.

Figure S13 Nucleotide and amino acid sequence alignment of *S*₁₄-RNase in different intergeneric and interspecific of citrus.

Figure S14 Nucleotide sequence alignment of the downstream region of the *S*₁₄-RNase allele from SI *C. ichangensis* and SC *C. medica*.

Figure S15 Analysis of the S-haplotype of *P. trifoliata* and *C. medica* accessions.

Figure S16 Expression of *S*-RNase in the style tissues in different intergeneric and interspecific of citrus.

Table S1 Identification of self-incompatibility/compatibility in a *F. hindsii* cross-pollination population.

Table S2 Statistical analysis of SI/SC phenotypes for the progeny from a PN02 × DB02 hybrid population.

Table S3 KASP genotyping analysis of a *F. hindsii* cross-pollination population.

Table S4 Genotype and SC phenotype identification for 46 *F. hindsii* accessions.

Table S5 List of the significant integrated haplotype score measures (iHS) for the single-nucleotide polymorphism (SNP) markers and their closest genes.

Table S6 SC-related genomic region with mean XP-EHH value >1 and associated genes.

Table S7 Mapping summary for WGBS reads.

Table S8 Summary information of DNA methylation in leaf, anther, and style tissues from DB02.

Table S9 Summary of small RNA sequencing data for the anther and style tissues from DB02.

Table S10 List of siRNAs associated with flanking regions of *FhiS*₂-RNase allele in anther and style tissues (RPM ≥ 2).

Table S11 Summary statistics for the assembly of 3 genome sequences.

Table S12 Summary of transposable elements in S-locus of *F. hindsii*.

Table S13 Isolation and identification of S-locus *F*-box alleles in citrus.

Table S14 Detailed information relating to whole-sequence alignments of S-locus.

Table S15 Detailed information for MITE annotation in the 5-kb flanking regions of *S*-RNase alleles in citrus.

Table S16 Detailed information for MITE annotation near the S-RNase genes in three common GSI families.

Table S17 List of primers used in this study.

Table S18 Isolation and identification of S-ribonuclease alleles in citrus.

Table S19 Detailed information relating to checking the S alleles for 24 mandarin accessions.

Figure S17 Multiple routes for the loss of SI in the three major genera (*Citrus*, *Fortunella*, *Poncirus*) of the Aurantioideae.