Published in final edited form as:
Nat Plants. 2020 February ; 6(2): 131–142. doi:10.1038/s41477-020-0597-3.

Evolution of self-compatibility by a mutant S_m-*RNase* in citrus

Mei Liang[1,8], **Zonghong Cao**[1], **Andan Zhu**[2], **Yuanlong Liu**[3], **Mengqin Tao**[1], **Huayan Yang**[1], **Qiang Xu Jr**[1], **Shaohua Wang**[4], **Junjie Liu**[1], **Yongping Li**[1], **Chuanwu Chen**[5], **Zongzhou Xie**[1], **Chongling Deng**[5], **Junli Ye**[1], **Wenwu Guo**[1], **Qiang Xu**[1], **Rui Xia**[3], **Robert M Larkin**[1], **Xiuxin Deng**[1], **Maurice Bosch**[6], **Vernonica E. Franklin-Tong**[7], **Lijun Chai**[1,*]

[1]Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan 430070, P.R. China

[2]Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, P.R. China

[3]State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University, Guangzhou, 510642, P.R. China

[4]Institute of Tropical and Subtropical Cash Crops, Yunnan Academy of Agricultural Sciences, Kunming 650201, P.R. China

[5]Guangxi Key Laboratory of Citrus Biology, Guangxi Academy of Specialty Crops, Guilin 541000, P.R. China

[6]Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Plas Gogerddan, Aberystwyth, UK

[7]School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

[8]Department of Ecology & Evolutionary Biology, University of Connecticut, Storrs, CT 06269, U.S.A.

*Correspondence: Lijun Chai (chailijun@mail.hzau.edu.cn).

Author contributions

L.J.C., M.L. and V.E.F.-T. designed the experiments. M.L., Z.H.C., H.Y.Y., Q.X.(Jr.) and M.Q.T., performed the experiments. J.J.L. completed the collection and sequencing of sour orange. M.L., A.D.Z., Y.L.L., Y.P.L. and R.X. analyzed the bioinformatic data. S.H.W., C.W.C., Z.Z.X. and C.L.D collected the pummelo accessions. J.L.Y., W.W.G., Q.X., R.M.L., X.X.D., M.B., and L.J.C. were involved in the research design and improvement of the manuscript. M.L. and V.E.F.-T. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Data and materials availability

The RNA-seq data shown in Suppl. Table 3 (for pummelo and grapefruit) are available at NCBI Bioproject ID under accession codes PRJNA526584 and PRJNA573625. The sequence data shown in Suppl. Table 7 of the pummelo *S1*-locus and *S2*-locus BAC clones are available at NCBI Bioproject ID under accession codes at PRJNA573817 and PRJNA573818 respectively. The DNA-seq Data shown in Suppl. Table 11 from the different citrus species are available at NCBI Bioproject ID under accession codes PRJNA544805 (*C. maxima*), PRJNA544816 (*C. aurantium*), PRJNA544866 (*C. paradisi*), PRJNA544867 (*C. limon*), PRJNA573624 (*C. reticulata*). The sequence data of the fifteen citrus *S-RNase* genes are available at NCBI Genbank ID under accession codes MN652897, MN652898, MN652899, MN652900, MN652901, MN652902, MN652903, MN652904, MN652905, MN652906, MN652907, MN652908, MN652909, MN652910, MN652911, MN652912, respectively. Any other raw data are available from the corresponding author upon request.

All materials generated during this study are available from the corresponding author upon request.

Europe PMC Funders Author Manuscripts

Abstract

Self-incompatibility (SI) is an important mechanism that prevents self-fertilization and avoids inbreeding in flowering plants. The most widespread SI system utilizes *S*-ribonucleases (*S-RNase*s) and *S*-locus F-boxes (*SLF*s) as *S*-determinants. In citrus, SI is ancestral; *Citrus maxima* (pummelo) is self-incompatible, while *Citrus reticulata* (mandarin) and its hybrids are self-compatible (SC). Here we identified nine highly polymorphic pistil-specific, developmentally expressed *S-RNase*s from pummelo that segregate with *S*-haplotypes in a gametophytic manner and cluster with authentic *S-RNase*s. We provide evidence that these *S-RNase*s function as the female *S*-determinants in citrus. Moreover, we found that each *S-RNase* is linked to ~nine *SLF*s. Analysis of 117 citrus *SLF/SLFL* genes revealed clustering into 12 types and evidence that the *S-RNase*s and intra-haplotypic *SLFs/SLFL*s co-evolved. Our data are consistent with citrus having an *S*-locus comprising a *S-RNase* and several *SLF*s that fit the non-self-recognition model. We identified a predominant single nucleotide mutation, S_m*-RNase*, in SC citrus, which provides a 'natural' loss of function. We present evidence that SI-SC transitions due to the S_m-RNase, initially arose in mandarin, spreading to its hybrids and became fixed. Identification of an evolutionarily distant new genus utilizing the *S*-RNase-based SI system, >100 million years separated from the nearest *S*-RNase family, is a milestone for evolutionary comparative studies.

Self-incompatibility (SI) is a major genetically controlled mechanism used by flowering plants to prevent inbreeding and facilitate outcrossing. SI is usually controlled by a single *S*-locus organised in a haplotype that carries two tightly linked *S*-genes, the pollen and pistil *S*-determinants[1]. The Solanaceae, Rosaceae and Plantaginaceae employ gametophytic SI (GSI)[2–5], with the *S*-genotype being determined by the haploid pollen. The female *S*-determinant in these families is encoded by a class III *S*-ribonuclease (*S*-RNase) expressed in the pistil. This system is therefore referred to as *S*-RNase-based SI[4]; see[5] for a review. The pollen *S*-determinants of *S*-RNase-based SI usually comprise multiple *S*-locus F-box (*SLF*) genes[6]; see[7] for a review. As these families utilizing the S-RNase SI system have a common origin and are the ancestors of ~75% of dicot families, *S*-RNase-based SI is believed to be the ancestral state for the vast majority of dicots[8,9].

Citrus belongs to the Rutaceae family and is a commercially important crop, grown worldwide. Since most citrus species are woody perennial trees with a long juvenile period (taking 5-10 years from seed to flowering)[10], studies involving crosses are very time consuming. Nevertheless, pollination studies have established that many citrus accessions are self-incompatible[11–13]. This is in line with it being a long-lived perennial; reproductive assurance is less of an issue and is outweighed by the cumulative, deleterious effects of inbreeding, so they are generally outcrossers[14]. Moreover, citrus also utilizes sporophytic apomixis, which is an asexual reproduction resulting in seed formation from somatic nucellar cells[15,16]. Data from crosses show that SI in citrus is controlled by a single codominant *S*-locus with multiple *S*-alleles[17,18]. It has been proposed that citrus may employ an *S*-RNase-based SI system, as several *S-RNase* homologues were identified in citrus accessions[19–21]. However, evidence that these genes function as *S*-determinants in citrus has not been demonstrated to date.

In a large SI population, the diversification of *S*-alleles is maintained by negative frequency-dependent selection, because pollen with rare *S*-haplotypes is compatible with more potential pistils than those with common *S*-haplotypes[22,23]. However, when compatible pollen or pollinators are limited, natural selection favours breakdown of SI to self-compatibility (SC), as selfing provides reproductive assurance[24]. Breakdown of SI is common in the S-RNase SI system and can involve gene duplication or mutations in either the *S-RNase* and *SLF* or non *S*-determinants; see[25,26] for reviews.

Here, we demonstrate that SI citrus species employ the *S*-RNase-based GSI and harbour an *S-RNase* linked to several *SLF*s at each *S*-locus. Notably, we identified a mutant S-RNase, S_m*-RNase*, responsible for SC in citrus; this SI-SC transition occurred first in mandarin and then spread to its hybrids. As citrus is evolutionarily distant from other families using *S-RNase* SI, our data provide new insights into the evolution of this widespread SI system.

Results

Previous studies indicated that some pummelo (*C. maxima*) accessions from Japan are predominantly outcrossers and their self-fertilization barriers are determined by SI[12,13,18]. To test whether this extends to the Chinese accessions of pummelo, manual pollinations comprising self- and cross-pollinations were performed on nine pummelo varieties widely cultivated in China (Supplementary Tables 1, 2). Four accessions (HB, WB, SJ and GX) produced seedless fruit in the absence of pollination, identifying them as parthenocarpic (Supplementary Table 2). All cross pollinations resulted in fruits; the mean number of seeds per fruit was 121 ± 7, while self-pollinations resulted in no seed set (Supplementary Table 2). As they have fully functional pollen and pistils and they set seed when cross-pollinated, this provides good evidence that these Chinese pummelos are self-incompatible.

Identification of pistil-expressed *S-RNase* genes in pummelo

We constructed 64 RNA-seq libraries of style and anther from these Chinese pummelos (Supplementary Table 3). As a previous study had suggested that pummelo had candidate *S-RNase* genes[20], we investigated this further. Nine candidate *S-RNase* genes with complete open reading frames (ORFs) and homology to previously reported *S-RNase*s were identified. We named these genes S_n*-RNase*, with $_n$ denoting the *S*-haplotype (S_1*-RNase* to S_9*-RNase*). Their full-length cDNA clones contain coding regions ranging from 660- to 699-bp (Supplementary Fig. 1) and encode highly polymorphic (38.1% to 76.7% deduced amino acid identity) proteins (Supplementary Fig. 2). Their predicted molecular masses, between 22.96 to 24.47 kDa and alkaline isoelectric points (7.67 to 9.39; Supplementary Table 4) are similar to known *S*-RNases[8]. The highly polymorphic citrus sequences contain key features of known functional *S*-RNases[9] (Fig. 1a, Supplementary Fig. 2). However, comparison of these sequences with known *S*-RNases reveals that although the C2 and C3 domains are relatively well conserved (including the histidine residues implicated in catalysis), other domains are poorly conserved across species (Supplementary Fig. 2). An extra histidine is conserved across all nine citrus *S*-RNases, but not present in the other *S*-RNases. The pummelo *S*-RNases have five hypervariable regions; two correspond to the *HVa* and *HVb* domains in other species, but three are unique to pummelo (Supplementary Figs. 2, 3).

Phylogenetic analysis revealed that pummelo *S*-RNases cluster together with authentic *S*-RNases, but on a separate branch (Supplementary Fig. 4). This provides good evidence that these highly polymorphic pummelo sequences may be *S*-RNases.

We investigated the frequency of these nine *S-RNase* genes within natural pummelo populations comprising 391 individuals from various provinces in China (Supplementary Fig. 5a). These *S*-haplotypes were abundant and found in 76.2% of the accessions and their frequency ranged from 2.3% to 30.2% (Supplementary Fig. 5b). This pattern is consistent with the negative frequency dependent selection utilized by *S*-determinant genes[23].

Analysis of various tissues using qRT-PCR analysis and western blotting (Supplementary Fig. 6a-c) showed that the nine pummelo *S-RNase*s were specifically expressed in the style. Although transcript levels in the style were highest five days before anthesis and decreased thereafter, western blots revealed that the protein was not detectable at this stage, but that it was detected four days before anthesis and levels of the protein progressively increased until the pistils were mature (Supplementary Fig. 6c, Fig. 1b). Thus, these citrus RNases display the tissue- and developmental specificity expected of an *S*-determinant.

The pummelo *S-RNase*s segregate with *S*-haplotype in a GSI manner

The *S*-genotypes of fifteen pummelo accessions were assigned based on pollinations and aniline blue staining (Supplementary Fig. 7, Supplementary Table 5). As many of the pummelo accessions examined contained the nine identified *S-RNase*s, the *S*-RNases identified were then assigned a particular *S*-allele using *S*-allele-specific PCR primers (Supplementary Table 6). This showed that the S_1- to S_9-*RNase*s were uniquely amplified for their assigned *S*-alleles, with each accession having a pair of *S*-RNase bands corresponding to that particular genotype, in each of fifteen pummelo accessions (Fig. 1c).

To confirm our designation and to demonstrate that these *S*-RNases segregated genetically as expected, we used PCR to establish the *S*-genotypes of the progeny of these plants (T1 plants; Fig. 1d, Table 1). For a half compatible cross (e.g. the SJ × WB cross, $S_5S_6 \times S_2S_5$), the *S*-RNases assigned to the parental *S*-alleles and the 118 progeny *S*-RNase genotypes (assigned by PCR), segregated into the two expected classes and no other genotypes were observed (i.e. absence of S_5S_5 and S_5S_6 genotypes). All of the 118 T1 plants examined had either the S_2S_5 (56 plants) or S_2S_6 (62 plants) genotypes in the expected 1:1 ratio (χ^2 = 0.31, P = 0.58, Table 1). They lacked S_5S_5 or S_5S_6 genotypes, demonstrating that only S_2 pollen was compatible with S_5S_6 pistils, as expected for a GSI system. Reciprocal crosses (WB × SJ) yielded 59 T1 progenies with either the S_2S_6 or S_5S_6 genotypes in a 1:1 ratio (χ^2 = 1.37, P = 0.24). This half-compatibility was also observed in other tests where the parents shared a common *S*-allele (Table 1). For a fully compatible cross, (SJ × WB cross, $S_5S_6 \times S_1S_2$), four *S*-genotypes were identified, segregating 1:1:1:1 as expected (χ^2 = 2.32, P = 0.51, Table 1). These data provide genetically-based evidence that the outcomes of these pollinations segregate as expected for a GSI system. Moreover, they show that the pummelo *S-RNase*s assigned to the *S*-genotypes segregate as expected for *S*-alleles at the *S*-locus. Antibodies raised against the recombinant S_1- and S_2-RNases also confirmed that the product of the cloned *S*-RNases was associated with the *S*-alleles assigned by pollination (Fig. 1e).

The *S-RNase*s are responsible for *S*-specific pollen inhibition in pummelo

We expressed recombinant citrus S_1- and S_2-RNases as GST fusion proteins (Fig. 2a) and confirmed they exhibited RNase activity (Fig. 2b, c). To establish whether these *S-RNase*s function as *S*-determinants in citrus, we examined if these fusion proteins could function to specifically inhibit incompatible pollen tube growth *in vitro* (Fig. 2d, Supplementary Figs. 8, 9). This used a bioassay similar to that used for *Papaver* SI[27,28]. While this does not fully mimic the *in vivo* pollen-pistil interaction, it does provide a measure of *S*-specific pollen inhibitory activity exhibited by the female *S*-determinant. Because pollen from a heterozygous plant comprises two *S*-haplotypes, a single recombinant *S*-RNase should induce a half-incompatible reaction (i.e., inhibition of pollen tube growth for 50% of the pollen tubes). The recombinant S_1-RNase-GST inhibited pollen tubes from plants with genotype S_1S_3 (half-compatible, blue bar) by ~54% ($P < 0.001$, **) compared to its untreated (UT) control, while the compatible pollen genotype S_5S_6 (grey bars), was only inhibited by 9% compared to its untreated control ($P = 0.067$, N.S., Fig. 2d(i)). Likewise, the S_2-RNase-GST inhibited pollen tubes from plants with genotype S_2S_8 by ~51% ($P < 0.001$, **)compared to its untreated control, while the compatible pollen genotype S_5S_6 (grey bars), was only inhibited by 1% compared to its untreated control ($P = 0.763$, N.S., Fig. 2d(ii)). Combined recombinant S_1- and S_2-RNase fusion proteins inhibited pollen from plants with genotype S_1S_2 (an incompatible combination) by 62% ($P < 0.001$, **,) compared to untreated pollen tubes. The same proteins had reduced inhibitory activity against compatible pollen from plants with genotype S_5S_6, with a 7% ($P = 0.113$, N.S.) reduction in length compared to its untreated control. The combined S_1- and S_2-RNases had an intermediate effect on half-compatible pollen from plants with genotypes S_1S_3 and S_2S_8 with a mean reduction of 44 and 45% of pollen tube length compared to their respective untreated controls ($P < 0.001$, ** for both). Together these data provide evidence that the *S*-RNases have *S*-specific pollen inhibitory activity. These data demonstrate that, although pummelo pollen does not grow to the same extent as *in vivo* (probably because of absence of key pistil components absent *in vitro*), and despite some non-specific inhibitory activity by the recombinant proteins, pollen of different haplotypes was affected specifically and differentially by the recombinant S_1- and S_2-RNase fusion proteins. Although they may not reflect the *in vivo* situation exactly, and further studies are required to validate how representative of an *in vivo* response they are, these data demonstrate that the pummelo *S-RNase* genes identified here can induce *S*-specific inhibition of incompatible pollen, providing confirmatory data to support the genetic evidence that they function as the female *S*-determinant.

Identification of *SLF* genes linked to the *S-RNase* gene

The *S*-locus in other *S*-RNase SI systems has the male *S*-determinant, F-box proteins[5] linked to the female *S*-determinant, the *S*-RNase. To identify the pollen *S*-determinant, a BAC library covering the S_1- and S_2-loci was constructed from a pummelo accession with a S_1S_2 genotype (Supplementary Table 7). Approximately 240-kb of the S_1-locus and approximately 198-kb of the S_2-locus were assembled. Harr plot analysis of the S_1- and S_2-allele sequences showed that both ends of the *S*-loci were largely syntenic, while the remaining region was highly divergent (Supplementary Fig. 10). Each *S*-locus had twelve *F-box* genes associated with it, as well as other genes, including transposons (Supplementary

Tables 8, 9). The *F-box* genes on the S_1-locus have 33.6% to 74.2% deduced amino acid identity which is comparable to that of the *F-box* genes at the S_2-locus (33.5% to 73.9%). Nine *F-box* genes exhibited a relatively high sequence divergence (78.1% to 93.7% deduced amino acid identity) between the two *S*-loci; three *F-box* genes were highly conserved (99.5% to 99.7% deduced amino acid identity), and may be *SLF-like* (*SLFL*) genes.

RNA-seq analysis revealed that all the *SLF*s were specifically expressed in anthers (Supplementary Fig. 11); qRT-PCR verified this, identifying expression of the *SLF*s in anthers, pollen and pollen tubes (Supplementary Fig. 12). Linkage analysis confirmed that plants from segregating families with the S_1-*RNase* expressed S_1-*SLF1* to S_1-*SLF9* and that those with the S_2-*RNase* carried S_2-*SLF1* to S_2-*SLF9* (Supplementary Fig. 13a, b). This is consistent with *SLF*s being the pollen *S*-determinants in pummelo.

Identification of the *S*-locus in *Citrus*

Based on the two conserved sequences at both ends of the *S*-loci, we identified seven additional *S*-loci from the reported seven citrus genomes. *S*-loci were found to span 198- to 370-kb; each of these contained one *S-RNase* and 11 to 17 *SLF*s/*SLFL*s (Fig. 3a; Supplementary Fig. 14). Analysis of 117 *SLF*s/*SLFL*s revealed clustering into 12 types; we designated the *F-box* of each locus as S_n-*SLFx*/ S_n-*SLFLx* ($_n$ indicating *S*-haplotype and *x* the type; Supplementary Fig. 15). The pollen and pistil *S*-determinants should exhibit evidence of co-evolution. Examining the synonymous (*Ks*) and non-synonymous (*Ka*) substitution rates, revealed that those of the *S*-RNases (*Ks* = 0.814, *Ka* = 0.503) and each SLF/SLFL type (*Ks* = 0.977 - 1.047, *Ka* = 0.422 - 0.461) were quite similar, and much higher than the inter-allelic K_a and K_s values of each SLF/SLFL type (*Ks* = 0.015 - 0.476, *Ka* = 0.009 - 0.156, Fig. 3b). These data suggest that the *S-RNase* and intra-haplotypic *SLF*s co-evolved and are likely to be similarly ancient.

Like Petunia[6], the citrus SLFs/SLFLs show extensive polymorphism between types (44.24% to 46.52% identity), while the sequence identities between allelic variants of each type were more highly conserved (74.78% to 97.49% identity; Fig. 3b). The clustering of the SLF sequences, together with intra-haplotypic *vs.* inter-allelic differences is consistent with the non-self-recognition model of *S*-RNase/SLF evolution, which proposes that divergent/deleted *SLFs* predict the specific target *S*-RNase, with one missing, mutated or diverged SLF in each haplotype[6,29]. Within each “type” of SLF, amino acid sequence polymorphism varied; we observed some alleles with high sequence conservation and others with moderate conservation (Fig. 3c, Supplementary Table 10). The non-self recognition model predicts that the *S*-RNase is the target of the non-self SLFs[6]; thus, in the citrus type 1 *SLF* group, S_1-*SLF1* is the most diverged, so the S_1-RNase is predicted to be the target of the more conserved SLF1s (S_2-, S_6-, S_{13}-, S_{12}-, S_{14}-, S_{10}- and S_{11}-SLF1; Fig. 3c, Supplementary Table 10). In Petunia, *SLF* copy number varied from 0 to 2[6,29]; we also found missing SLFs. Within the type 9 *SLF*s the S_{11}-, S_{13}- and S_{14}-SLF9 alleles were absent; moreover, two copies of SLF within one type were often found (Fig. 3c, Supplementary Fig. 15).

Our data provide evidence that the *S-RNase* genes and intra-haplotypic *SLF/SLFL* genes are likely to have co-evolved and that the divergence of the inter-allelic *SLF/SLFL* genes from each type occurred more recently. Together, our findings are consistent with the non-self-

recognition model of *S*-RNase/SLF evolution. As this is well established to be utilized by species with *S*-RNase/SLFs confirmed to function as *S*-determinants in SI, this contributes to the evidence that SI in pummelo is likely to be controlled by *S-RNase* and *SLF* genes acting as *S*-determinants.

Identification of a mutant S_m-*RNase* responsible for SC in *Citrus*

Among the 15 identified *S*-RNases, we unexpectedly found that the coding sequence of a *S*-RNase from *C. sinensis* was shorter than the others (Supplementary Fig. 16). Cloning of this gene (named S_m-*RNase*) revealed a single nucleotide deletion at position 443, resulting in a frameshift mutation and premature stop codon at position 498 (Fig. 4a, Supplementary Fig. 16). The truncated predicted S_m-RNase protein contains the catalytic histidines, but lacks the C4 and C5 conserved domains, the HV4, HV5 hypervariable domains and also four conserved cysteine residues (Fig. 4a). Because the unmutated progenitor of the S_m-*RNase* was not identified in the accessions, we engineered a "recovered" version (named S_m^R-RNase) by insertion of a single adenine in the deleted position. The S_1-*RNase* has the nearest sequence identity to the S_m-*RNase* and it has adenine at this position; this is predicted to result in a normal transcript length (Fig. 4a, Supplementary Fig. 17).

We hypothesized that the truncated S_m-RNase could be responsible for the loss of functional SI in the SC accessions. We first examined the level of *mRNA* expression of the *Sm-RNase*, as several reports of SC in other species being due to low *S-RNase* expression exist[30,31]. Analysis of a range of tissues revealed that the expression of S_m-*RNase* was minimal compared with that of the S_2-*RNase* (Fig. 4b). Absolute qRT-PCR confirmed that the expression of the S_m-*RNase* transcript was greatly reduced (Fig 4c, Supplemental Fig. 18); RNA-seq confirmed this (Fig. 4d). These data suggest that the SC phenotype could potentially be due to the reduced S_m-*RNase* transcript level. We next expressed the recombinant S_m- and S_m^R-RNase-GST fusion proteins (Fig. 4e); both exhibited RNase activity (Fig. 4f, g), so SC cannot be due to lack of this activity.

To further test how the S_m-*RNase* mutation might confer SC, we tested the activity of recombinant S_m-RNase GST-fusion protein and its "recovered" S_m^R-RNase GST-fusion version on pollen from a plant with genotype S_2S_m (a half-compatible combination as no homozygous plants exist) in the SI *in vitro* bioassay (Supplemental Fig 19, Fig. 4h). The recombinant S_m-RNase GST-fusion protein did not significantly inhibit pollen tube growth compared to the untreated control ($P = 0.156$, N.S., *ANOVA*). This lack of pollen inhibitory activity for the S_m-RNase suggests that this mutation could potentially be responsible for the SC phenotype. As the S_m-*RNase* has lost hypervariable domains HV4 and HV5 it is possible that specificity may reside here. Supportive of this idea, predicted structural analysis suggests these regions reside at the surface of the protein (Supplemental Fig. 20). In contrast, treatment of pollen from a S_2S_m genotype plant with the "recovered" S_m^R-RNase-GST resulted in inhibition of growth, with pollen tube lengths significantly reduced compared to the S_m-RNase ($P < 0.001$, **, *ANOVA*) and not significantly different from the (half-compatible) pollen inhibitory activity displayed by the S_2-RNase-GST ($P = 0.787$, N.S., *ANOVA*). Moreover, as the "recovered" S_m^R-RNase exhibited a gain of pollen inhibitory activity, this is consistent with the explanation that the truncation of this gene may

be responsible for loss of activity and the SC phenotype. Thus, although the *S_m-RNase* is a functional RNase, it does not display *S*-specific pollen inhibitory activity. However, as expression of the *S_m-RNase* is almost zero in the SC accessions, we cannot conclude that this lack of pollen inhibitory activity is responsible for the SC phenotype.

Evolution of SI and SC in *Citrus*

We examined the frequency of the *S*-haplotypes of 153 citrus accessions by mapping the paired reads to the 15 *S-RNase*s identified (Supplementary Table 11). These 15 *S-RNase*s were present in 132 of these accessions. Each *S-RNase* occurred at a low frequency, in keeping with it being maintained by negative frequency-dependent selection (Supplementary Fig. 21). Analysis of the relationships of the 15 *S-RNase*s to investigate how they spread through the citrus species revealed that the phylogeny of the *S-RNase*s did not fit the phylogeny of citrus species as described by Wang et al[32] (Supplementary Fig. 22), suggesting that the divergence of these *S*-RNases occurred before citrus diverged.

Ninety accessions contained the *S_m-RNase* (Supplementary Fig. 21). All of those with the *S_m-RNase* were SC and the *S_m-RNase* was absent in all the SI accessions (Supplementary Fig. 23). *Ichang papeda* (an ancient near-citrus), is SI (Supplementary Fig. 24) and diverged earlier than the SC accessions, mandarin and its hybrids[33,34]. This suggests that SI is (as expected) the ancestral trait. Because the *S_m-RNase* was found in wild and cultivated mandarin and its hybrids (Supplementary Fig. 23), it suggests that the SI-SC transition arose initially in mandarin and then spread to its hybrids through mating or introgression (Fig. 5). Data suggest that the *S_m-RNase* is fixed in the hybrid citrus populations; exactly how this SI-SC transition became fixed is unknown, but selfing and apomixis, which enables breeders to fix valuable traits and heterozygosity[16,32] may have played a role.

Discussion

Studies of the verified *S*-RNase-based SI systems have to date been confined to the Rosaceae (Rosids) and the Solanaceae and Plantaginaceae (Asterids)[2–5,9] (Fig. 6). Here we identify several polymorphic pistil-expressed *S-RNase*s from pummelo and show that they segregate with *S*-haplotypes. We provide strong evidence that citrus utilizes the *S*-RNase-based SI system and that S-RNases function as pistil *S*-determinants, inhibiting pollen in an *S*-specific manner. Phylogenetically, *S-RNase*s are found in several divergent families and whether this SI system evolved several times remains controversial[3,8], as few families with *S-RNase*s shown to function in SI have been identified in the last 25 years, although putative *S*-RNases have been identified in the Rubiaceae[35,36]. Our identification of a functional *S*-RNase SI system in citrus, which diverged ~110 *m.y.a.* from the Solanaceae (Fig. 6), supports the idea that *S-RNase*s have a single origin, prior to the divergence of these families, as a common ancestor is more likely than >three separate independent gains of *S-RNase*.

In contrast to other SI systems, which have female and male *S*-determinants displaying co-evolutionary relationships, the *S-RNase*s and *SLF*s in the Solanaceae and Maloideae do not show this. Instead they utilize a collaborative “non-self recognition” system[6,8,29]. In this scenario, multiple *SLF*s are required for pollen *S*-specificity; as a functional *S*-haplotype

cannot encode a *SLF* that recognizes its own *S*-RNase, either a diverged or deleted allele of that *SLF* type is utilized. Thus, within an *S*-haplotype, the product of each type of SLF interacts with a group of non-self *S*-RNases that are collectively recognized and detoxified[6,29]. Our identification of multiple pollen-expressed *F-box*es (*SLF*s) tightly linked to each *S-RNase* suggests that the *S*-locus in citrus fits this model. Analysis of their highly polymorphic sequences revealed that the *SLF* types display evidence of co-evolution with *S-RNase*s. Moreover, the clustering of the citrus *SLF*s is consistent with the "non-self" recognition model[6,8,29] with a missing or diverged *SLF* for each haplotype. This substantiates the idea that these genes are likely to be involved in SI.

For many species, the evolutionary history of the SI-SC transition(s) is unclear[37]. Here, we begin to decipher the evolutionary history of the SI-SC transition in citrus. It is interesting to note that the SC trait in citrus is strongly associated with apomixis. Reproductive system change is a striking feature of crop domestication[14] and apomixis, which enables breeders to fix valuable traits and heterozygosity, is a powerful tool for breeders[16,32]. Although further studies are required, it is possible that apomixis, in conjunction with selection of the SC mutant, may have played an important role in citrus domestication. The SI trait is ancestral *in Citrus*[33,34]; while pummelo retained SI, mandarin and its hybrids became SC. Notably, here we identify a frameshift mutation in the female *S*-determinant, S_m*-RNase*, which yields a truncated *S-RNase*, and provide evidence that it is responsible for the loss of SI. The prevalence of this mutant in citrus populations suggest that SC has a single origin: S_m*-RNase* arose in mandarin and subsequently became prevalent and nearly fixed in its hybrids. Although this mutant *S*-RNase has extremely low expression *in planta*, which is sufficient to explain the SC phenotype, the S_m*-RNase* has RNase activity. This contrasts with how SC was achieved in many other *S*-RNase families, where loss of SI is often accompanied by the complete deletion of the *S-RNase* from the *S*-locus[38,39], with the exception of a report by Golz *et al*[40]. In citrus, although low expression could explain the SC phenotype, the functionally active S_m-RNase does not inhibit pollen. Thus, as the S_m-RNase is missing two hypervariable domains, which are predicted to be at the surface of the protein, this hints that *S*-specificity may be located in this region.

In summary, we provide evidence that SI in citrus utilizes an *S*-RNase-based SI system. Our identification of a new genus utilizing this SI system is a milestone for evolutionary comparative studies[8]. As citrus is >100 *m.y.* separated from the nearest *S*-RNase family, our data will help clarify the distribution of *S*-RNase-based SI systems and their evolution. We provide evidence that SI is ancestral and show that a truncated S_m*-RNase* is responsible for the loss of SI. This has allowed us to decipher the evolutionary history of the SI to SC transition in >150 citrus accessions. Selfing, combined with apomixis, in conjunction with selection of SC by breeders makes this an interesting example of evolution of plant reproductive strategies.

Methods

Plant materials

To analyze the *S*-allele that controls SI in citrus, a natural population with 391 pummelo accessions were collected in the wild (Supplementary Fig. 3a). Among them, fifteen

pummelo cultivars were used to perform the pollination assay and the aniline blue staining (Supplementary Table 1). Leaves and various floral tissues, including petals, anthers, filaments, styles, ovaries, and pedicels, were collected. We collected pistils from flowers at different developmental stages before anthesis. These tissues were immediately frozen in liquid nitrogen and stored at -80 °C. Fresh anthers were collected, dehisced, dried, and stored in a bottle containing desiccant at -20 °C.

Phenotypic characterization of pollination

Cross-, self- and non-pollinations were performed 1 day before anthesis. Five days after pollination, pistils were excised and fixed in a mixture of alcohol and acetic acid (4:1). The growth of the pollen tubes within the style was observed using the aniline blue fluorescence staining method[20] (see Supplementary Fig. 7). The fruit set ratio and the seed number were determined for mature pummelo fruits (Supplementary Table 2).

mRNA sequencing

Total RNA was extracted from citrus anthers and styles based on Liu and Liu's method[41]. The RNA was used for high-throughput RNA-seq library construction and sequenced using the Illumina Hiseq 2500 platform (Supplementary Table 3). Approximate 13 Gb reads per sample (read length = 150-bp) were generated. Clean data were *de novo* assembled separately for each citrus accession using Trinity version 2.8.4[42]. Reads from each library were then mapped back to the assembled transcripts using the align_and_estimate_abundance.pl script in the Trinity package in combination with bowtie 2[43], and the value of fragments per kilobase of transcripts per million mapped reads (FPKM) of each gene was estimated using the RSEM method.

S-RNase identification

To identify candidate *S-RNase*s involved SI, six nucleotide sequences encoding *S-RNases* from species with *S*-RNase-base SI were downloaded from NCBI (HE805271.1 and AJ315593.1 from Antirrhinum; D63887.1 and AB568389.1 from Solanaceae; FJ543097.1 and AF327223.1 from Rosaceae) and aligned by codons using ClustalW in MEGA7[44]. Using this alignment, an *S-RNase* HMM (Hidden Markov Model) profile was built with the Hmmbuild subprogram in HMMER[45]. The Trinity transcripts were queried with this profile using nhmmer.

SLF identification

A bacterial artificial chromosome (BAC) library from S_1S_2 pummelo was constructed using the pIndigoBAC536-S vector with ~110-kb insertion in size. BAC clones that we screened using multiple long PCR primers for S_1- and S_2-*RNase* were sequenced using the Illumina Hiseq 2500 platform (Supplementary Table 4). S_1- and S_2-loci were separately assembled using SOAP denovo[46]. The *Citrus* genomes for *C. sinensis*, *C. maxima*, *C. medica*, *C. ichangensis*, *A. buxifolia* and *C. reticulata* were downloaded from http://211.69.140.136/orange/index.php; the genome of *C. Clementina* was downloaded from https://phytozome.jgi.doe.gov/pz/portal.html. Based on the conserved sequences at each end of S_1- and S_2-loci (Fig. 3a), seven additional *S*-loci were identified from these citrus genomes.

Gene predictions and annotations of all *S*-loci were made using FGENESH and SWISS-PROT databases. Genes containing an F-box domain and FBA (F-box associated) motif were designated *SLF*s. Syntenic regions among all *S*-loci were identified using the blastn program with a threshold value of 0.95 identity, and the regions above 500-bp were plotted using Circos.

Sequence analysis of the candidate pistil and pollen *S*-determinants

Primers for the amplification of S_1- to S_9-*RNase* were designed based on the unigenes from the RNA-seq; primers of the *SLF* sequences were designed based on the genomic sequences assembled from the BAC (Supplementary Table 6). The complementary DNA fragments were amplified using polymerase chain reaction (PCR) with reverse transcription using the standard PCR protocols. All PCR products were cloned into the pEASY-Blunt Cloning Vector (TransGen Biotech) and were sequenced using Sanger sequencing technology. Deduced amino acid sequences were aligned using ClustalW in MEGA7[44] and sequence similarity was illustrated by shading with GeneDoc 2.602. Normed variability index (NVI) value for *S-RNase* genes was calculated with a sliding window of seven residues as described by Kheyr-pour et al[47].

S_m-RNase identification and sequence cloning

The mutated S_m-RNase was identified within the S_m-locus from sweet orange genome. Primers for the amplification of S_m-*RNase* were designed based on the genomic sequence. The complementary DNA fragment of S_m-*RNase* was amplified as above. To examine the function of the unmutated S_m-*RNase*, an adenine nucleotide was introduced at position 443 in the S_m-*RNase* to engineer a “recovered” version of the S_m-*RNase*: S_m^R-*RNase* (Supplementary Fig. 10), using overlap PCR technology (for primers see Supplementary Table 6). Secondary structure predictions for the S_1-, S_m-, S_m^R-RNases were made using the I-TASSER server (https://zhanglab.ccmb.med.umich.edu/)[48] and the PyMol molecular visualization package[49].

Quantitative analysis

Methodology for quantitative RT-PCR (qRT PCR) and western blot were carried out to check gene expression and translation in different tissues as described previously[20]. Heat maps for the expression were drawn using TBtools[50]. RNA-seq reads were aligned to *S*-locus using TopHat2 and the alignment result was visualized using the Integrative Genomics Viewer (IGV)[51,52]. The uniquely mapped reads without any mismatch were used to calculated the FPKM of the genes on *S*-locus with Cufflinks[53].

Absolute quantification method was employed to analyze S_2- and S_m-*RNase* expression levels. The plasmids inserted the full length of S_2- and S_m-*RNase* were used to make 10-fold serial dilutions DNA template from 15 ng μL^{-1} down to 1.5 fg/μL. The PCR system and thermocycler conditions were same with that of qRT PCR. The Ct values (Y-axis) and the log gene copy number (X-axis) were used to generate a standard curve plot and the PCR efficiency were calculated as described by Workenhe et al[54]. Plasmid DNA standard curve equations (Supplementary Fig. 18) were used to calculate the absolute copy number of S_2- and S_m-*RNase* within 50 ng pistil mRNA from the S_2S_m plant.

Phylogenetic analysis

The deduced amino acid sequences were aligned by MAFFT[55], and manually adjusted by removing spurious alignments and long gaps. RAxML[56] was used to construct maximum likelihood (ML) tree under the substitution model PROTGAMMAWAG with 1000 bootstrap replications. To estimate synonymous (*Ks*) and non-synonymous (*Ka*) substitution rates in DnaSP[57], the aligned protein sequences were converted to nucleotide alignments. Species divergence was obtained from the mean estimate time in TimeTree[58].

Expression of *S*-RNase recombinant proteins

The open reading frames from S_1-, S_2-, S_m- and S_m^R-RNase without signal peptide region were expressed in *Escherichia coli* strain BL21 (DE3) (TransGen) as glutathione *S*-transferase (GST) fusion proteins using pGEX-6P-1 (GE Healthcare). *E. coli* strains were induced by 0.2 mM isopropyl-1-thio-β-galatoside (IPTG) for 16 h at 18 °C and the glutathione Sepharose 4B beads (GE Healthcare) protein was used to protein purification according to the manufacturer's protocol. These GST fusion proteins were analysed on SDS-PAGE and western blots. Anti-S_1- and anti-S_2-RNase antibodies were raised against the S_1- and S_2-RNase GST fusion proteins respectively in rabbit and used at a 1:2,000 dilution (anti-S_1-) and 1:1,000 dilution (anti-S_2-), with a goat anti-rabbit IgG-HRP secondary antibody (GenScript, A00098, at 1:5,000 dilution). The RNase activity and pollen inhibitory activity of the GST fusion proteins were tested (see below).

RNase activity in-gel and in-solution assay

In-gel RNase activity assays were performed according to Yeh and Green[59] with slight modification. 20 μg recombinant *S*-RNase protein in standard sample buffer was electrophoresed on 12.5% SDS-PAGE without yeast RNA. After electrophoresis, the SDS-PAGE gel was washed, incubated, stained, and destained[59]. The gel was incubated in 0.1 M Tris-HCl containing 2.4 mg ml^{-1} *Torulopsis utilis* RNA (torula yeast RNA; Sigma) for 1 h at 37 °C. The Tris-HCl buffers used for the in-gel RNase assay were pH 8.0.

We also performed an RNase activity assay of the recombinant *S*-RNase using citrus RNA from pistils as a target. 10 μg *S*-RNase and 2 μg total RNA isolated from citrus pistil were incubated at 37°C for 1 h in a 20 μL reaction mixture[60]; rRNA was then separated on a 1% agarose gel and stained with ethidium bromide and examined for degradation.

Assessment of *S*-RNase pollen inhibitory activity in *in vitro* pollen bioassays

As no homozygous citrus accessions were available, pollen from the plants with genotypes S_1S_3, S_1S_2, S_2S_8 and S_5S_6 were used to test the *S*-specific inhibition of the recombinant S_1- and S_2-RNase-GSTs in an *in vitro* pollen bioassay. Thus, an incompatible combination was achieved by combining the S_1- and S_2-RNases and testing against pollen from a plant of genotype S_1S_2. Half-compatible combinations were achieved with recombinant S_1-RNase *vs* pollen from plants genotype S_1S_3 or S_1S_2, and recombinant S_2-RNase *vs* pollen from plants genotype S_1S_2 or S_2S_8. Fully compatible tests used pollen from plants genotype S_5S_6. For the functional examination of the pollen inhibitory activity of the recombinant S_m- and the "recovered" S_m^R-RNase-GST, pollen from plants with genotype S_2S_m provided a half-

Europe PMC Funders Author Manuscripts

compatible test; recombinant S_2-RNase-GST provided a positive control for maximal inhibitory activity (half-compatible). GST was used as a the "untreated" control for all tests.

A germination medium (GM) described by Liang et al[20] was used to grow pollen tubes *in vitro*. Before each bioassay, the recombinant GST-fusions were dialyzed against GM without PEG-4000 using a Millipore Amicon Ultra-10 kDa centrifugal filter device. A dilution series of the recombinant *S*-RNase-GST fusion proteins against pollen was performed to assess the optimal concentration to use for the bioassays, to obtain maximal inhibitory activity with minimal non-specific activity (see Supplemental Fig. 8). For the bioassay tests, pollen was grown on 200 μL aliquots of GM for 2 h, before the addition of 10 μg mL^{-1} recombinant GST-fusion protein and then incubated for a further 5 h. Each pollen bioassay was independently performed at least three times and the length of pollen tubes (≥ 50 tubes per assay, so $n \geq 150$ in total) were measured with Image-Pro Plus v6.0 (Media Cybernetics, Bethesda, MD, USA). Because we show actual pollen tube lengths, we had to display the data in pairwise comparisons, so each test had its appropriate control. Data was displayed using box and whisker plots to show the full range of pollen tube lengths and analysed using ANOVA.

S-allele mapping and diversification analysis

To characterize the *S*-alleles of the citrus accessions, paired-end reads of whole genome sequences from 153 accessions (Supplementary Table 5) in citrus were mapped to 15 different *S-RNase* genes that we identified using Bowtie 2 with the following parameters: "-D 5 -R 1 -N 0 -L 22 -i S,0,2.50 --fr --no-mixed --no-discordant". These bowtie parameters only retained the uniquely mapped reads with zero mismatches per seed. Bedtools[61] was used for statistical analysis of the nonzero coverage (≥1 reads) of each alignment. For the non-complete alignments with nonzero coverage > 0.9 < 1.0, we cloned the full-length sequence used for these alignments and analyzed these sequences using Sanger sequencing. The *S*-haplotypes of 153 citrus accessions are summarized in Supplementary Table 11.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to Prof. Juyou Wu from Nanjing Agricultural University for providing the sample of *Pyrus bretschneideri*. This research was financially supported by the National Key Research and Development Program of China (Grant Number 2018YFD1000107), the National Natural Science Foundation of China (nos. 31772259, 31630065, 31521092), the Fundamental Research Funds for the Central Universities (2662019PY044) and the China Agriculture Research System (CARS-27). The Biotechnology and Biological Sciences Research Council (BBSRC) funds research in the labs of MB & VEF-T (BB/P005489/1). We would like to thank Prof. Tianzhong Li (China Agricultural University) and Prof. Chris Franklin (School of Biosciences, University of Birmingham, UK) for their valuable suggestions.

References

1. de Nettancourt, D. Incompatibility and Incongruity in Wild and Cultivated Plants. Vol. 3. Springer-Verlag; 2001.

2. Xue Y, Carpenter R, Dickinson HG, Coen ES. Origin of allelic diversity in antirrhinum *S* locus RNases. Plant Cell. 1996; 8:805–814. [PubMed: 8672882]
3. Sassa H, et al. Self-incompatibility (*S*) alleles of the Rosaceae encode members of a distinct class of the T_2/*S* ribonuclease superfamily. Mol Gen Genet. 1996; 250:547–557. [PubMed: 8676858]
4. McClure BA, et al. Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. Nature. 1989; 342:955–957. [PubMed: 2594090]
5. Takayama S, Isogai A. Self-incompatibility in plants. Annu Rev Plant Biol. 2005; 56:467–489. [PubMed: 15862104]
6. Kubo K, et al. Collaborative non-self recognition system in S-RNase–based self-incompatibility. Science. 2010; 330:796–799. [PubMed: 21051632]
7. Fujii S, Kubo K-i, Takayama S. Non-self- and self-recognition models in plant self-incompatibility. Nat Plants. 2016; 2
8. Ramanauskas K, Igić B. The evolutionary history of plant T2/S-type ribonucleases. PeerJ. 2017; 5:e3790. [PubMed: 28924504]
9. Igić B, Kohn JR. Evolutionary relationships among self-incompatibility RNases. Proc Natl Acad Sci USA. 2001; 98:13167–13171. [PubMed: 11698683]
10. Krajewski AJ, Rabe E. Citrus flowering: A critical evaluation. J Hortic Sci. 1995; 70:357–374.
11. Ngo BX, Wakana A, Kim JH, Mori T, Sakai K. Estimation of self-incompatibility *S* genotypes of *Citrus* cultivars and plants based on controlled pollination with restricted number of pollen grains. J Fac Agr Kyushu Univ. 2010; 55:67–72.
12. Ngo BX, Wakana A, Park SM, Nada Y, Fukudome I. Pollen tube behaviors in self-incompatible and self-compatible *Citrus* cultivars. J Fac Agr Kyushu Univ. 2001; 45:443–457.
13. Yamamoto M, Kubo T, Tominaga S. Self- and cross-incompatibility of various citrus accessions. J Japan Soc Hort Sci. 2006; 75:372–378.
14. Miller AJ, Gross BL. From forest to field: perennial fruit crop domestication. Am J Bot. 2011; 98:1389–1414. [PubMed: 21865506]
15. Zhang S, et al. Reproduction in woody perennial Citrus: an update on nucellar embryony and self-incompatibility. Plant Reprod. 2018; 31:43–57. [PubMed: 29457194]
16. Hand ML, Koltunow AMG. The genetic gontrol of apomixis: asexual seed formation. Genetics. 2014; 197:441–450. [PubMed: 24939990]
17. Soost RK. The incompatibility gene system in citrus. Proc First Intl Citrus Sym. 1968; 1:189–190.
18. Kim JH, et al. Determination of self-incompatible *Citrus* cultivars with S_1 and/or S_2 alleles by pollination with homozygous S_1 seedlings (S_1S_1 or S_2S_2) of 'Banpeiyu' pummelo. J Japan Soc Hort Sci. 2011; 80:404–413.
19. Miao H, Ye Z, Silva JA, Qin YH, Hu G. Identifying differentially expressed genes in pollen from self-incompatible "Wuzishatangju" and self-compatible "Shatangju" mandarins. Int J Mol Sci. 2013; 14:8538–8555. [PubMed: 23595002]
20. Liang M, et al. Genome-wide identification and functional analysis of S-RNase involved in the self-incompatibility of citrus. Mol Genet Genomics. 2017; 292:325–341. [PubMed: 27933381]
21. Zhang S, et al. Characterization of the 'Xiangshui' lemon transcriptome by de novo assembly to discover genes associated with self-incompatibility. Mol Genet Genomics. 2015; 290:365–375. [PubMed: 25252890]
22. Castric V, Vekemans X. Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. Mol Ecol. 2004; 13:2873–2889. [PubMed: 15367105]
23. Wright S. The distribution of self-sterility alleles in populations. Genetics. 1939; 24:538–552. [PubMed: 17246937]
24. Goodwillie C, Kalisz S, Eckert CG. The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. Annu Rev Ecol Evol Syst. 2005; 36:47–79.
25. Stone JL. Molecular mechanisms underlying the breakdown of gametophytic self-incompatibility. Q Rev Biol. 2002; 77:17–32. [PubMed: 11963459]
26. McClure B, Cruz-García F, Romero C. Compatibility and incompatibility in S-RNase-based systems. Ann Bot-London. 2011; 108:647–658.

27. Franklin-Tong VE, Lawrence MJ, Franklin FCH. An *in vitro* bioassay for the stigmatic product of the self-incompatibility gene in *Papaver rhoeas* L. New Phytol. 1988; 110:109–118.
28. Franklin-Tong, VE. Self-incompatibility in Papaver rhoeas: Progress in understanding mechanisms involved in regulating self-incompatibility in PapaverSelf-Incompatibility in Flowering Plants: Evolution, Diversity, and Mechanisms. Franklin-Tong, VE, editor. Springer-Verlag; 2008. 237–258.
29. Kubo K, et al. Gene duplication and genetic exchange drive the evolution of S-RNase-based self-incompatibility in *Petunia*. Nat Plants. 2015; 1
30. Royo J, et al. Loss of a histidine residue at the active site of S-locus ribonuclease is associated with self-compatibility in Lycopersicon peruvianum. Proceedings of the National Academy of Sciences. 1994; 91:6511–6514.
31. Sassa H, Hirano H, Nishio T, Koba T. Style-specific self-compatible mutation caused by deletion of the S-RNase gene in Japanese pear (*Pyrus serotina*). Plant J. 1997; 12:223–227.
32. Wang X, et al. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat Genet. 2017; 49:765–772. [PubMed: 28394353]
33. Yang XM, et al. Molecular phylogeography and population evolution analysis of *Citrus ichangensis* (Rutaceae). Tree Genet Genomes. 2017; 13
34. Wu GA, et al. Genomics of the origin and evolution of *Citrus*. Nature. 2018; 554:311–316. [PubMed: 29414943]
35. Nowak MD, Davis AP, Anthony F, Yoder AD. Expression and trans-specific polymorphism of self-incompatibility RNases in *coffea* (Rubiaceae). Plos One. 2011; 6:e21019. [PubMed: 21731641]
36. Asquini E, et al. S-RNase-like sequences in styles of *Coffea* (Rubiaceae). Evidence for S-RNase based gametophytic self-Incompatibility? Trop Plant Biol. 2011; 4:237–249.
37. Igić B, Lande R, Kohn JR. Loss of self-incompatibility and its evolutionary consequences. Int J Plant Sci. 2008; 169:93–104.
38. Li M, et al. Genome structure and evolution of *Antirrhinum majus* L. Nat Plants. 2019; 5:174–183. [PubMed: 30692677]
39. Okada K, et al. Deletion of a 236 kb region around S_4-*RNase* in a stylar-part mutant S_{4sm} haplotype of Japanese pear. Plant Mol Biol. 2008; 66:389–400. [PubMed: 18175198]
40. Golz JF, Clarke AE, Newbigin E, Anderson M. A relic S-RNase is expressed in the styles of self-compatible Nicotiana sylvestris. Plant J. 1998; 16:591–599. [PubMed: 10036777]
41. Liu, Y, Liu, Q. Effcient Isolation of RNA from Fruit Peel and Pulp of Ripening Navel Orange (*Citrus sinensis* Osbeck). Huazhong Agricultural University; 2006.
42. Grabherr MG, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2011; 29:644–654. [PubMed: 21572440]
43. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. [PubMed: 22388286]
44. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016; 33:1870–1874. [PubMed: 27004904]
45. Eddy SR. A new generation of homology search tools based on probabilistic inference. Genome Inform. 2009; 23:205–211. [PubMed: 20180275]
46. Luo R, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience. 2012; 1:18. [PubMed: 23587118]
47. Kheyr-Pour A, et al. Sexual plant reproduction sequence diversity of pistil S-proteins associated with gametophytic self-incompatibility in *Nicotiana alata*. Sex Plant Reprod. 1990; 3:88–97.
48. Yang J, et al. The I-TASSER Suite: protein structure and function prediction. Nat Methods. 2015; 12:7–8. [PubMed: 25549265]
49. Chen C, Xia R, Chen H, He Y. TBtools, a Toolkit for biologists integrating various HTS-data handling tools with a user-friendly interface. 2018
50. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14:178–192. [PubMed: 22517427]

51. Robinson JT, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29:24–26. [PubMed: 21221095]
52. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28:511–515. [PubMed: 20436464]
53. Workenhe ST, Kibenge MJT, Iwamoto T, Kibenge FSB. Absolute quantitation of infectious salmon anaemia virus using different real-time reverse transcription PCR chemistries. J Virol Methods. 2008; 154:128–134. [PubMed: 18789975]
54. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30:772–780. [PubMed: 23329690]
55. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30:1312–1313. [PubMed: 24451623]
56. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25:1451–1452. [PubMed: 19346325]
57. Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. Tree of life reveals clock-like speciation and diversification. Mol Biol Evol. 2015; 32:835–845. [PubMed: 25739733]
58. Yen Y, Green PJ. Identification and Properties of the Major Ribonucleases of *Arabidopsis Thaliana*. Plant Physiol. 1991; 97:1487–1493. [PubMed: 16668575]
59. Besbes F, Franz-Oberdorf K, Schwab W. Phosphorylation-dependent ribonuclease activity of Fra a 1 proteins. J plant physiol. 2019; 233:1–11. [PubMed: 30572279]
60. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
61. Group, A. P. An update of the angiosperm phylogeny group classifcation for the orders and families of flowering plants: APG III. Bot J Linn Soc. 2009; 161:105–121.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

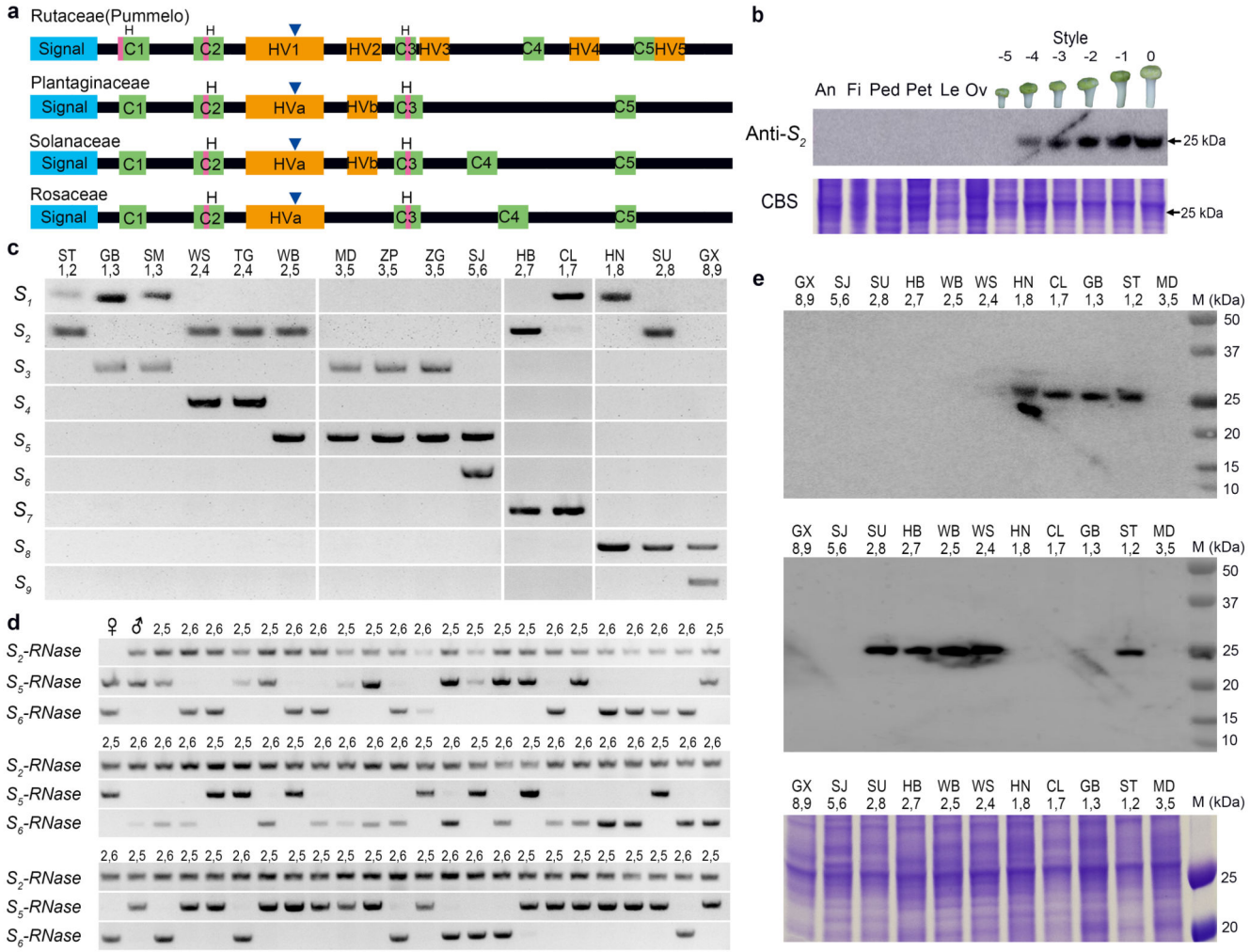

Figure 1. The citrus *S*-RNases exhibit key features of *S*-RNases.

a, Cartoon showing key features of the pummelo *S*-RNase sequences compared to other *S*-RNases, including five conserved domains (C1-C5, green boxes) and hypervariable domains (HV1-HV5, orange boxes). All of the *S*-RNases have a signal peptide (blue box), two or three conserved histidines (pink) and a single intron (triangle). **b**, Western blot showing tissue-specific and developmental expression of the S_2-RNase protein in pistils. Antibody raised against recombinant S_2-RNase cross-reacts with a ~25 kDa protein in extracts from mature pistils (0, open flower); no protein is detected at -5 days before anthesis, low expression is detected at -4 days and this increases over time as the pistil matures. The S_2-RNase antisera did not crossreact with a protein in other tissues, anther (An), filament (Fi); pedicel (Ped), petal (pet), leaf (Le), ovary (Ov). Coomassie blue staining shows equal loading (lower panel). **c**, The *S*-genotypes of fifteen pummelo accessions (indicated above each lane) were assigned using aniline blue staining of pollinated pistils (see Supplementary Table 3). PCR of leaf DNA, using *S*-RNase specific primers (indicated left: S_1- *to* S_9), shows two *S*-allele-specific transcripts for S_1-*RNase* to S_9-*RNase* (S_1-S_9) amplified from each pistil, corresponding to those assigned by pollination. **d**, *S-RNase*s segregate with the *S*-locus in F_1 progeny. A pistil (♀) from a pummelo plant (accession SJ) assigned genotype

Europe PMC Funders Author Manuscripts

S_5S_6 (lane 1) was pollinated with pollen (♂) from a plant (accession WB) assigned genotype *S_2S_5* (lane 2) using pollinations. Here, genotyping of seedling progeny from this cross using PCR with *S_2-, S_5- and S_6-RNase* primers shows that the parental pistils carry *S_5-* and *S_6-RNase* sequences and pistils used for the pollen donor carry *S_2-* and *S_5-RNase* sequences; the seventy progeny shown here display pairs of amplified *S-RNase* sequences corresponding to either *S_2S_5* (2,5) or *S_2S_6* (2,6). **e**, Western blots of pummelo pistil extracts (accessions and *S*-genotypes indicated above lanes), using antibody raised against the recombinant *S_1*-RNase (upper panel) and *S_2*-RNase (middle panel); Coomassie staining (lower panel) shows loading. The *S_1*-RNase protein (~27 kDa) was detected only in pistil extracts carrying the *S_1*-allele. Likewise, the *S_2*-RNase (~25 kDa) was only detected in pistils carrying the *S_2*-allele and not in those carrying other *S*-alleles. This shows that the antibody is both *S*-RNase-specific (as no other RNases are detected here) and a direct link between the *S-RNase* cloned (through the antibody to the recombinant protein) and *S*-alleles carried by the plant. Experiments were repeated independently twice for panel **c**, **d**, **e** and three times for panel **b**, with similar results obtained for each.

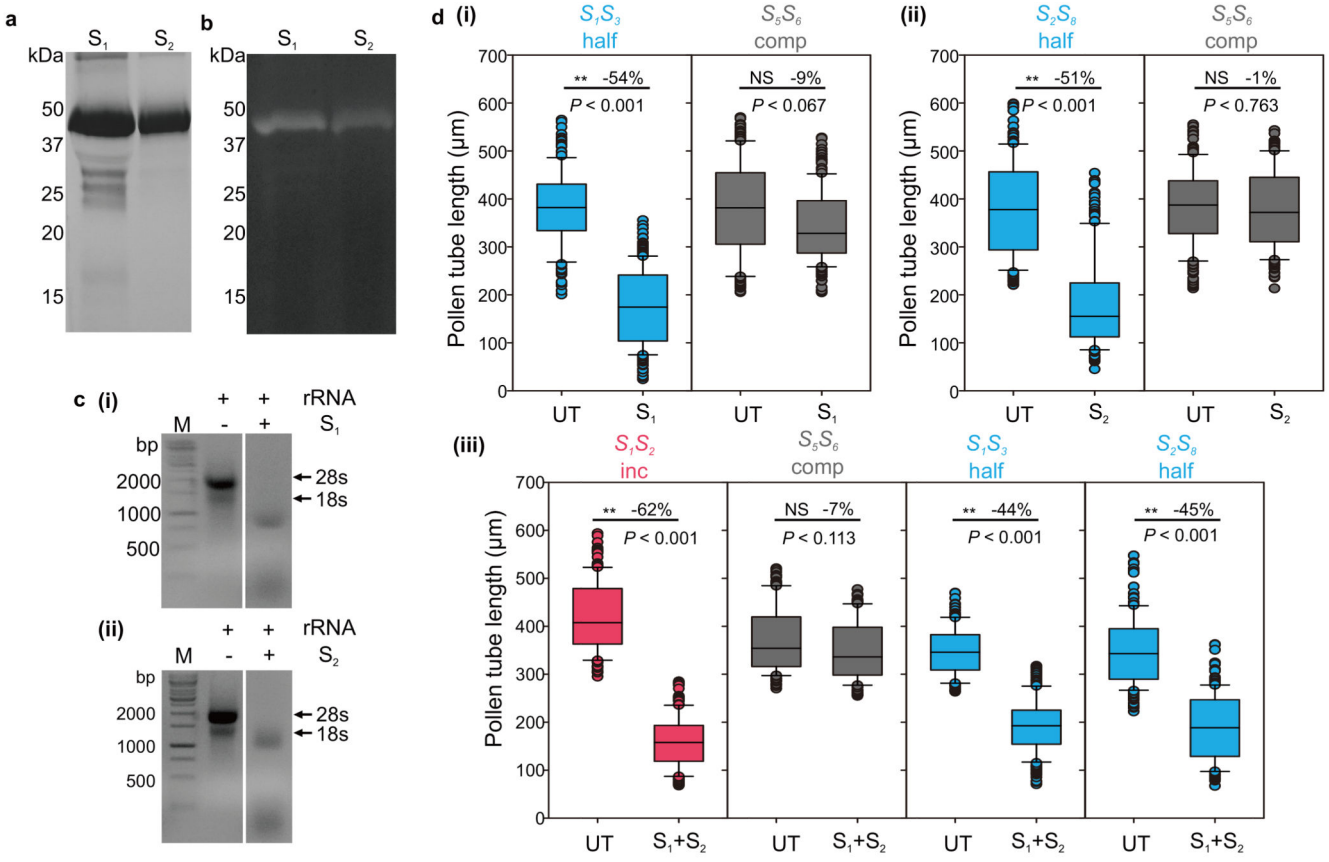

Figure 2. The pummelo *S-RNase*s exhibit RNase activity and can elicit *S*-specific pollen inhibition in an *in vitro* bioassay.

a, Recombinant S_1-RNase-GST (S_1) and S_2-RNase-GST (S_2) proteins migrate to ~45 kDa on SDS-PAGE. **b**, An in-gel RNase assay shows that the recombinant S_1- and S_2-RNase-GST proteins have RNase activity. **c**, Agarose gel showing that the recombinant S_1-RNase **(i)** and S_2-RNase **(ii)** can degrade citrus 28S and 18S rRNA. Experiments shown in panel **a**, **b**, **c** were repeated independently three times with similar results. **d**, Recombinant S_1- and S_2-RNases (S_1 and S_2, respectively) inhibit pollen tube growth differentially. **(i)** Individually, the recombinant S_1-RNase (S_1) partially inhibited pollen tubes from plants with genotype S_1S_3 that were half-compatible (blue bar, half) by ~50% compared to its untreated (UT) control, while the compatible pollen genotype S_5S_6 (grey bars, comp), was only inhibited by ~10% compared to its untreated (UT) control. Similarly, **(ii)** the S_2-RNase (S_2) inhibited pollen tubes from plants with genotype S_2S_8 (also half-compatible, blue bar) by ~50% compared to its untreated (UT) control. Together these data provide evidence that the *S*-RNases have *S*-specific pollen inhibitory activity. **(iii)** Combined recombinant S_1- and S_2-RNases (S_1+S_2) inhibited incompatible pollen from pummelo plants with genotype S_1S_2 (red bars, inc) by 62% compared to untreated pollen tubes (UT). The same proteins had little inhibitory activity against compatible pollen from plants with genotype S_5S_6 (grey bars), with a 7% reduction in length for S_1+S_2 compared to untreated controls (UT); this was considered non-specific activity. Combined S_1- and S_2-RNases (S_1+S_2) had an intermediate effect on half-compatible pollen from plants with genotypes S_1S_3 and S_2S_8 (blue bars) with

a ~45% reduction in pollen tube length compared to their respective untreated (UT) controls. The length of >50 pollen tubes was measured for each replicate (n = 3 biologically independent replicates, >150 in total). Box and whisker plots show the distribution of individual pollen tube lengths in *in vitro* bioassays of recombinant S_1- and S_2-RNases with pollen from plants of different genotypes (box indicates the upper & lower quartile, with median; lines above and below indicate the range; dots indicate the outliers). For each treatment, the mean reduction of pollen tube length (%) compared to its pairwise control is shown within each box. One-way ANOVA analysis was used to compare the pollen tube length of treatment vs untreated control (**, significantly different; NS, non-significant difference).

Europe PMC Funders Author Manuscripts

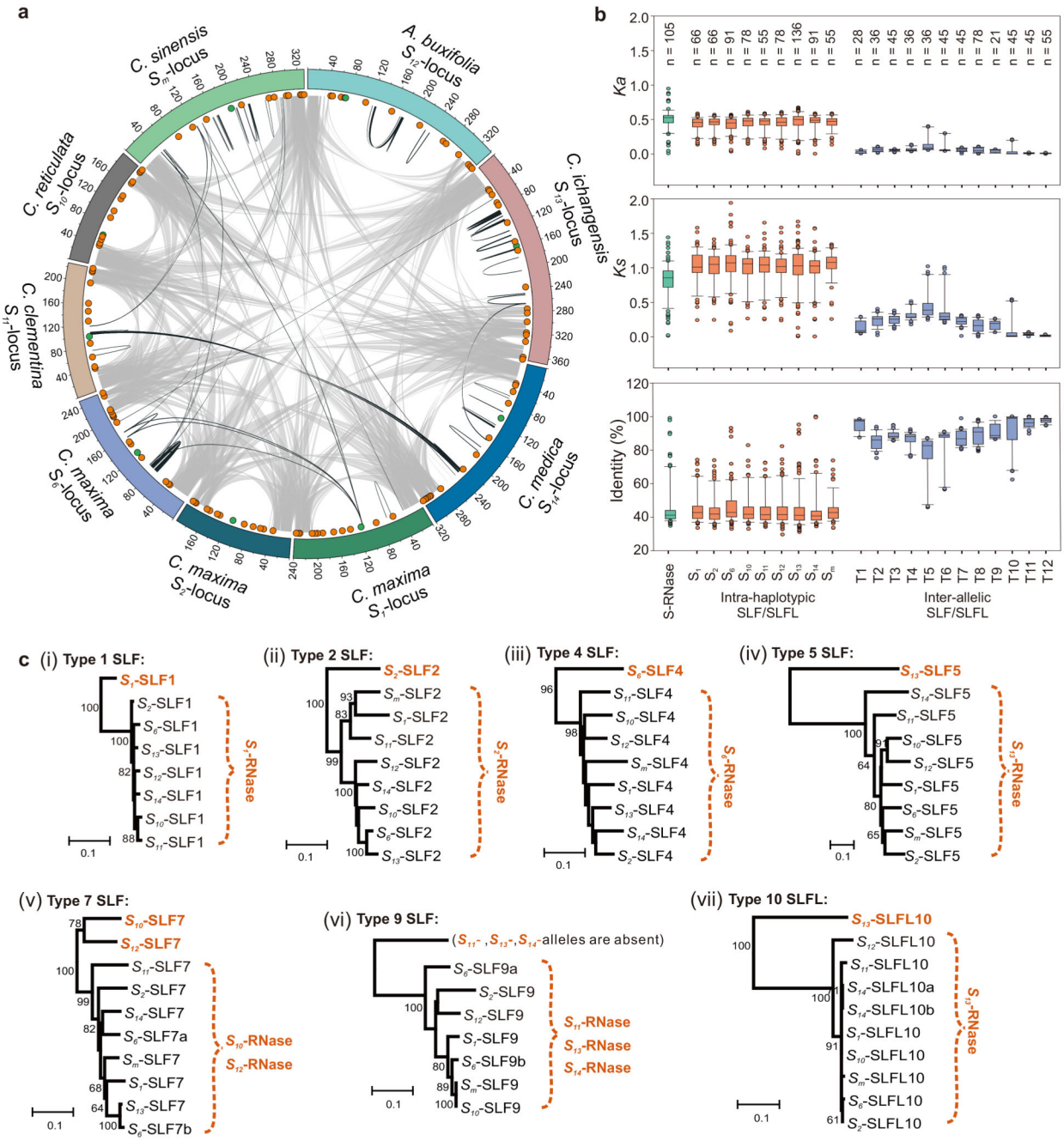

Figure 3. Multiple candidate *SLF/SLFL* genes are located at the citrus *S*-locus.

a, Gene synteny of nine citrus *S*-loci. Grey lines indicate syntenic regions at the end of the *S*-locus. Black lines indicate syntenic sequences in the intergenic regions of the locus. The *S-RNase*s (green dots) and *SLF/SLFL*s (orange dots) are indicated on the ideogram of each locus (see Supplementary Fig. 7 for more detail). **b**, Sequence identity, synonymous (*Ks*) and non-synonymous (*Ka*) substitutions in *S-RNase*s (green box plot), intra-haplotypic *SLF*s/*SLFL*s (S_1 to S_m; orange boxplots) and inter-allelic *SLF*s/*SLFL*s (T1 to T12; blue boxplots). Box and whisker plots show the distribution of the sequence identity, *Ks* and *Ka*

values for pairwise *S-RNase*s and *SLF*s/*SLFL*s. Box indicates the upper & lower quartile, with median; lines above and below indicate the range; dots indicate the outliers. The number of independent comparisons (n) are indicated in at the top of each box. **c**, Phylogenetic trees of *SLF*s designated type 1 (i), type 2 (ii), type 4 (iii), type 5 (iv), type 7 (v), type 9 (vi) and type 10 (vii) predict the SLF/*S*-RNase interaction. For the integrated phylogenetic tree of SLFs, see Supplementary Fig. 15. The SLF types shown here have a diverged or deleted SLF (indicated in orange); several have duplicate copies, indicated by a and b; e.g. *S6-SLF7a, b*. The *S*-RNases (also in orange) that are cognate to the diverged/deleted SLFs are predicted to interact with the conserved SLFs within the brackets under the non-self recognition model.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

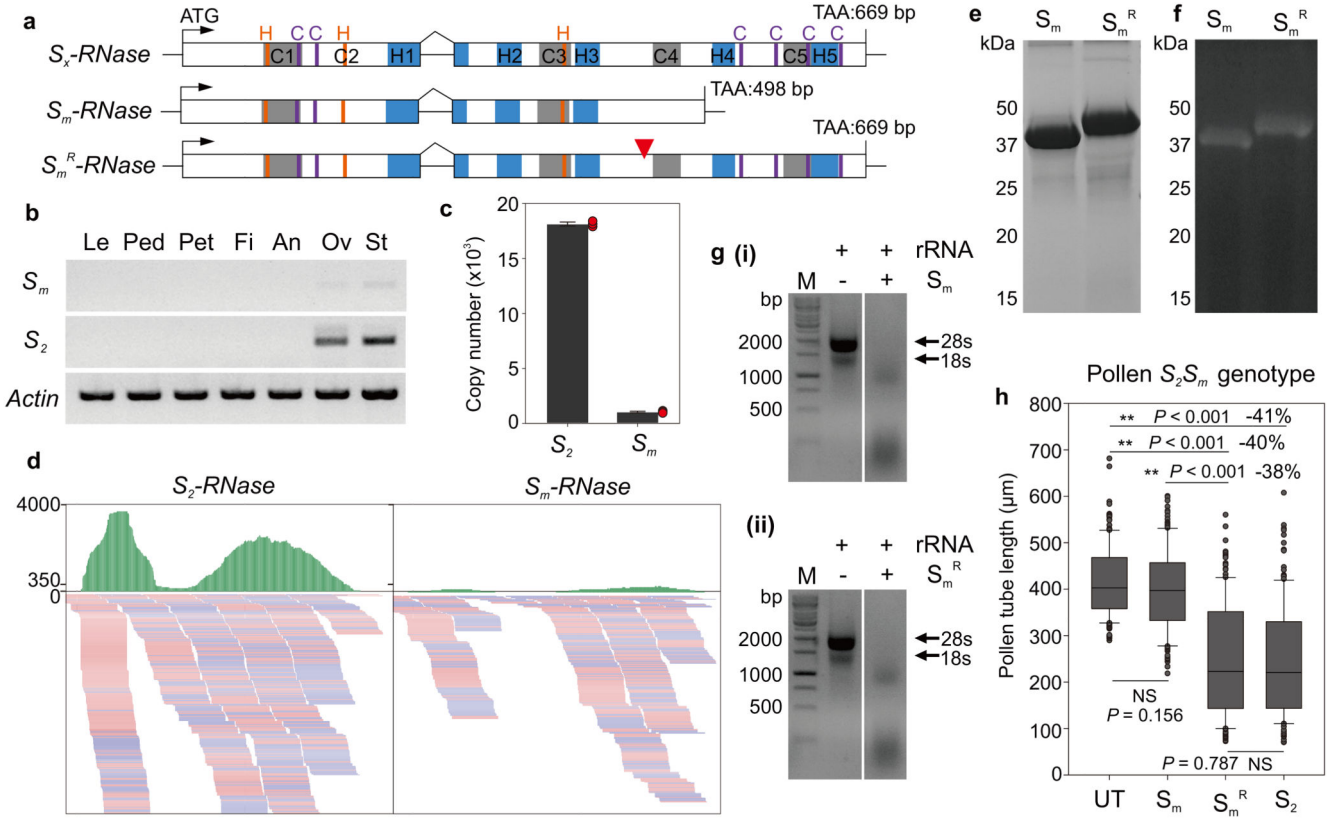

Figure 4. A truncated *S_m-RNase* appears to be responsible for the loss of SI in citrus.

a, Cartoon showing the gene structure of the *S_x-RNase* (top) and its mutant form (middle). *S_x-RNase* indicates the length and structure of the normal *S-RNase* with five conserved domains (C1-C5, grey boxes) and five hypervariable domains (H1-H5, blue boxes). *S_m-RNase* harbors a single nucleotide deletion resulting in a premature stop codon. The 'recovered' *S_m^R*-RNase (bottom) has a single nucleotide inserted (red triangle) that recovered the full length gene. Conserved histidine and cysteine residues are indicated in orange and purple, based on the deduced amino acid sequences. **b**, Expression of *S_m*- (*S_m*) and *S_2-RNase* (*S_2*) quantified using RT-PCR of different tissues from a *S_2S_m* plant (Le: leaf; Ped: pedicel; Pet: petal; Fi: filament; An: anther; Ov: ovary; St: style). The *S_m-RNase* is expressed at a much lower level than the *S_2-RNase*, in a tissue-specific manner (only in pistils: ovary and style tissues). **c**, Absolute copy number of the *S_2*- and *S_m-RNase* (extrapolated from the equation in Supplementary Fig. 18). The copy number of *S_m-RNase* in 50 μg *S_2S_m* style RNA is much lower than that of *S_2-RNase*. Error bars are shown for mean copy number ± SEM (n = 3 biological replicates; red dots indicate individual samples). **d**, IGV tracks displaying sequencing read clusters of *S_2*- and *S_m-RNase* gene from *S_2S_m* style RNA-seq. The green bars depict the number of the reads mapped to the reference. The reads mapped to *S_2-RNase* are clearly more than the reads mapped to *S_m-RNase*. Partial alignment of the RNA-mapping is shown below, with pink and blue representing the different read strands. **e**, SDS-PAGE analysis of recombinant *S_m*- (S_m) and *S_m^R*-RNase (S_m^R). The Mr of the fusion protein *S_m*-RNase-GST (39 kDa) is lower than the "restored" *S_m^R*-RNase-GST (45 kDa). **f,** RNase-activity gel of the recombinant *S_m*- (S_m)

and S_m^R-RNase (S_m^R). The mutated S_m-RNase and the "recovered" S_m^R-RNase have similar RNase activities. **g**, The recombinant S_m- (i) and S_m^R-RNase (ii) both degrade citrus 28S and 18S rRNA (agarose gel assay). Experiments were repeated independently twice times for panel **b** and three times for panels **e** and **g**, with the similar results for each experiments. **h**, The S_m^R-RNase recombinant protein displays inhibitory activity against pollen. Box plots show the distribution of individual pollen tube lengths in an *in vitro* bioassay of recombinant S_m-, S_m^R- and S_2 RNase against pollen from a plant with S_2S_m genotype. The S_m-RNase did not significantly (NS) inhibit pollen tubes from a S_2S_m plant when compared against untreated (UT) pollen; the S_m^R-RNase exhibited significant inhibitory activity (**) reducing the length of pollen tubes by ~40% and was not significantly different (NS) from that of the S_2-RNase. The length of >50 pollen tubes was measured for each replicate (n = 3 biologically independent replicates, >150 in total). One-way ANOVA analysis was used to compare the pollen tube lengths (treatment vs untreated control). The elements in box and whisker plots are the same as in Fig 2d.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

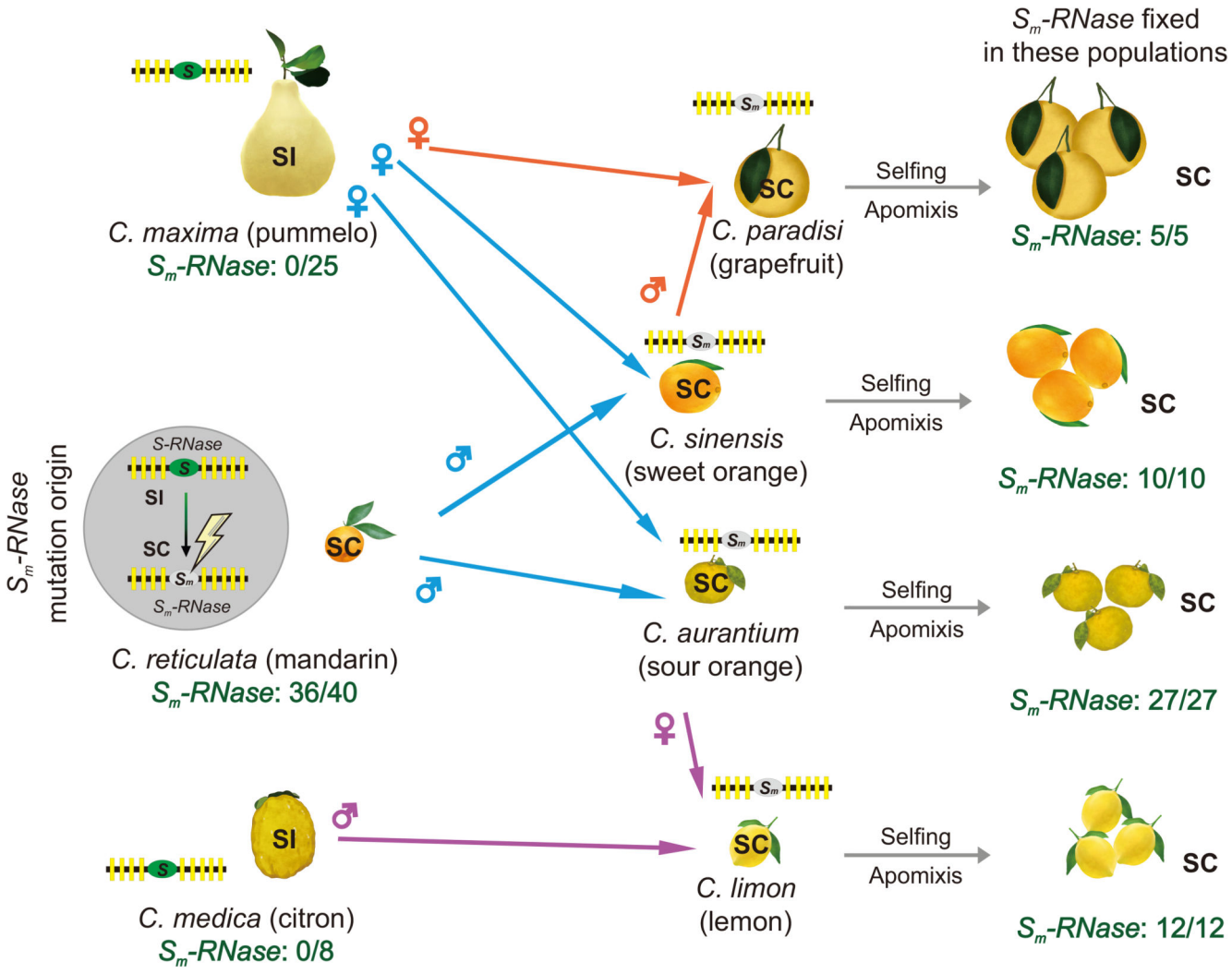

Figure 5. Postulated spread of the *S_m-RNase* and SC in *Citrus*.

Left hand side: Pummelo (top) is SI (no S_m-RNases were identified in 25 accessions). Citron (bottom) had a previously uncharacterized reproduction strategy, and we propose it is SI as we identified a single *S*-locus with one *S-RNase* and ~nine *SLFs* (and no S_m-RNases detected) in 8 accessions. We show a cartoon of the *S*-locus, with one *S-RNase* (green ellipse) and ~nine *SLFs* (yellow rectangles) by each citrus to indicate the status of its S-locus. Mandarin (middle) is SC. We observed the mutant *S_m-RNase* in 36/40 of mandarin accessions examined, and postulate that the mutant *S_m-RNase* arose spontaneously as an ancient event in wild mandarin (indicated by the lightning strike symbol) which converted mandarin from SI to SC (indicated by the grey ellipse). *Middle:* We propose that the *S_m-RNase* was subsequently transferred to other citrus species through pollination. The coloured arrows indicate the crosses between the parental genotypes proposed by Wu et al[34] likely responsible for the origin of these citrus accessions: sweet orange and sour orange originated from crosses between pummelo and mandarin, with both acquiring SC from mandarin (blue arrows); grapefruit originated from a cross between sweet orange and pummelo, acquiring SC from sweet orange (orange arrows) and lemon acquired SC from a

cross between sour orange and citron (purple arrows). *Right hand side:* Here we indicate that the S_m-*RNase* is responsible for the SC phenotype in grapefruit, sweet orange, sour orange and lemon, and that it is fixed in these populations, with all accessions examined being SC and containing the S_m-*RNase* (5/5, 10/10, 27/27, 12/12 respectively); none are SI. We propose that selfing and apomixis allowed the S_m-*RNase* to become fixed in these populations.

Europe PMC Funders Author Manuscripts

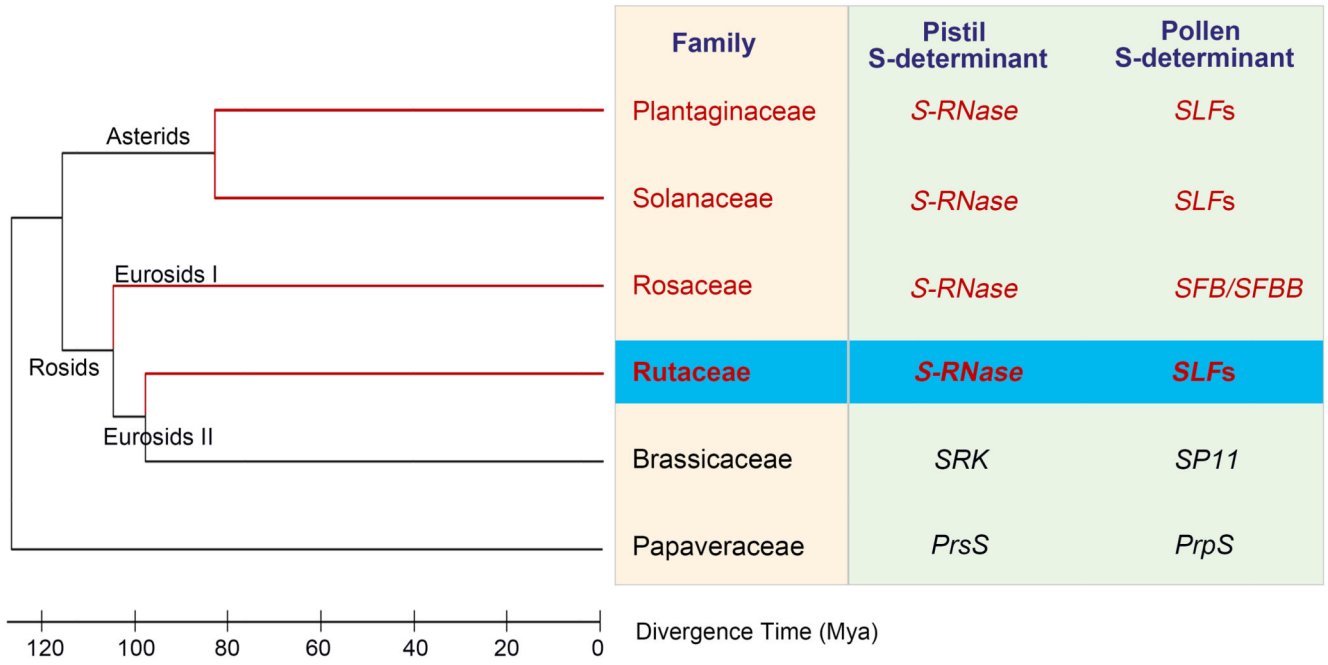

Figure 6. Phylogenetic relationship of different SI systems in flowering plants.

Phylogenetic tree showing relationships between the SI families (yellow box) with established *S*-determinants (green box). Bold text indicates the *S*-RNase and SLFs (blue box) identified in this study. The *S*-RNase-based SI systems (red text) all use *S-RNase* and *SLF*s/*SFB* (*S-haplotype-specific F-box*) /*SFBB* (*S-locus F-box brothers*) as the pistil and pollen *S*-determinants, respectively, and are highly divergent from the SI systems of the Papaveraceae and Brassicaceae. The *S-RNase*s from Rutaceae, Rosaceae, Solanaceae and Plantaginaceae group into three separate branches (Asterids, Eurosids I and II; see also Supplementary Fig. 3); these families are highly diverged (>100 *m.y.a.*). The most plausible interpretation is that the *S-RNase*s in the core eudicots have a single origin and evolved only once, prior to the divergence of these families >100 *m.y.a.*; this is more likely than several parallel gains of *S-RNase* at least three times. The tree is based on APG III[62] and the divergence time is based on the website of timetree (http://timetree.org/).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Nat Plants. Author manuscript; available in PMC 2020 August 13.

Table 1

Pummelo *S-RNases* in F1 progenies segregate in a gametophytic manner

Segregation analysis of *S*-haplotypes of F1 progenies of pummelo accessions for pollinations in half- and fully-compatible combinations were assigned using PCR (see Fig. 1f). Outcomes show segregation ratios as expected for a GSI system.

Phenotype[a]	Genetic cross	No. of progeny	Possible genotypes	Observed ratio[b]	Expected ratio[c]	Chi square	P-value
Fully-compatible	SJ (S_5S_6) × ST (S_1S_2)	77	$\underline{\mathbf{S_1S_5}}$: $\underline{\mathbf{S_2S_5}}$: $\underline{\mathbf{S_1S_6}}$: $\underline{\mathbf{S_2S_6}}$	22:17:15:23	1:1:1:1	2.32	0.51 (N.S.)
					1:1:1:1	2.32	0.51 (N.S.)
Half-compatible	SJ (S_5S_6) × WB (S_2S_5)	118	S_5S_5 : S_5S_6 : $\underline{\mathbf{S_2S_5}}$: $\underline{\mathbf{S_2S_6}}$	0:0:56:62	0:0:1:1	0.31	0.58 (N.S.)
					1:1:1:1	113.53	1.54^{E}-25**
	WB (S_2S_5) × SJ (S_5S_6)	59	S_2S_5 : S_5S_5 : $\underline{\mathbf{S_2S_6}}$: $\underline{\mathbf{S_5S_6}}$	0:0:34:25	0:0:1:1	1.37	0.24 (N.S.)
					1:1:1:1	61.75	2.49^{E}-13**
	HB (S_2S_7) × ST (S_1S_2)	115	S_2S_2 : S_2S_7 : $\underline{\mathbf{S_1S_2}}$: $\underline{\mathbf{S_1S_7}}$	0:0:53:62	0:0:1:1	0.7	0.4 (N.S.)
					1:1:1:1	116.41	4.58^{E}-25**
	ST (S_1S_2) × HB (S_2S_7)	42	S_1S_2 : S_2S_2 : $\underline{\mathbf{S_1S_7}}$: $\underline{\mathbf{S_2S_7}}$	0:0:21:21	0:0:1:1	0	1.0 (N.S.)
					1:1:1:1	42	4.01^{E}-09**
	GX (S_8S_9) × SU (S_2S_8)	76	S_8S_9 : S_9S_9 : $\underline{\mathbf{S_2S_8}}$: $\underline{\mathbf{S_2S_9}}$	0:0:39:37	0:0:1:1	0.05	0.82 (N.S.)
					1:1:1:1	76.11	2.10^{E}-16**
	GB (S_1S_3) × MD (S_3S_5)	113	S_1S_3 : S_3S_3 : $\underline{\mathbf{S_1S_5}}$: $\underline{\mathbf{S_3S_5}}$	0:0:62:51	0:0:1:1	1.07	0.3 (N.S.)
					1:1:1:1	115.14	8.58^{E}-25**

[a] The pollination phenotype was determined by aniline blue staining (see Fig. 1a-d).

[b] The *S*-genotype ratios observed in all of the progeny.

[c] The upper segregation ratio is expected from a gametophytic self-incompatibility (GSI) system; the lower is expected from simple Mendelian inheritance. All crosses with parents sharing an *S-RNase* haplotype show a result consistent with GSI, with a non-significant (N.S.) Chi square for this prediction and a highly significant difference (**) for the lower segregation ratio. These data provide clear evidence that pummelo *S-RNase*s segregate with the *S*-locus in a GSI manner.