nature plants

ARTICLES

https://doi.org/10.1038/s41477-020-0597-3

Check for updates

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

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

Self-incompatibility (SI) is an important mechanism that prevents self-fertilization and inbreeding in flowering plants. The most widespread SI system utilizes *S* ribonucleases (*S*-RNases) and *S*-locus F-boxes (SLFs) as *S* determinants. In citrus, SI is ancestral, and *Citrus maxima* (pummelo) is self-incompatible, while *Citrus reticulata* (mandarin) and its hybrids are self-compatible (SC). Here, we identify nine highly polymorphic pistil-specific, developmentally expressed *S*-RNases from pummelo that segregate with *S* haplotypes in a gametophytic manner and cluster with authentic *S*-RNases. We provide evidence that these *S*-RNases function as the female *S* determinants in citrus. Moreover, we show that each *S*-RNase is linked to approximately nine SLFs. In an analysis of 117 citrus *SLF* and SFL-like (*SLFL*) genes, we reveal that they cluster into 12 types and that the *S-RNase*s and intra-haplotypic *SLF* and *SLFL* genes co-evolved. Our data support the notion that citrus have a *S* locus comprising a *S-RNase* and several *SLF*s that fit the non-self-recognition model. We identify a predominant single nucleotide mutation, S_m-*RNase*, in SC citrus, which provides a 'natural' loss of function. We show 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 to facilitate outcrossing. SI is usually controlled by a single *S* locus organized in a haplotype that carries two tightly linked *S* genes: the pollen and pistil *S* determinants[1]. Solanaceae, Rosaceae and Plantaginaceae plants employ gametophytic SI (GSI)[2–5], and the *S* genotype is 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 ref. [5] for a review). The pollen *S* determinants of *S*-RNase-based SI usually comprise multiple *S*-locus F-box (*SLF*) genes[6] (see ref. [7] for a review). Families utilizing the *S*-RNase SI system have a common origin and are the ancestors of ~75% of dicot families; therefore, *S*-RNase-based SI is believed to be the ancestral state for the vast majority of dicots[8,9].

Citrus belong to the Rutaceae family and are 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 them being long-lived perennials; that is, 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 utilize sporophytic apomixis, which is an asexual reproduction process that results in seed formation from somatic nucellar cells[15,16]. Data from crosses show that SI in citrus is controlled by a single co-dominant *S* locus with multiple *S* alleles[17,18]. It has been proposed that citrus may employ a *S*-RNase-based SI system, as several *S-RNase* homologues were identified in citrus accessions[19–21]. However, there is currently no evidence to indicate that these genes function as *S* determinants in citrus.

In a large SI population, the diversification of *S* alleles is maintained by negative frequency-dependent selection because pollen with rare *S* haplotypes are compatible with more potential pistils than those with common *S* haplotypes[22,23]. However, when compatible pollen or pollinators are limited, natural selection favours the 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 *S-RNase* and *SLF* genes or non-*S* determinants (see refs. [25,26] for reviews).

Here, we demonstrate that SI citrus species employ the *S*-RNase-based GSI and harbour a *S-RNase* linked to several *SLF*s at each *S* locus. Notably, we identify a mutant *S*-RNase, S_m-*RNase*, that is 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 that use *S-RNase* SI, our

[1]Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan, P. R. China. [2]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA. [3]Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, P. R. China. [4]State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University, Guangzhou, P. R. China. [5]Institute of Tropical and Subtropical Cash Crops, Yunnan Academy of Agricultural Sciences, Kunming, P. R. China. [6]Guangxi Key Laboratory of Citrus Biology, Guangxi Academy of Specialty Crops, Guilin, P. R. China. [7]Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Plas Gogerddan, Aberystwyth, UK. [8]School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, UK. ✉e-mail: chailijun@mail.hzau.edu.cn

data provide new insights into the evolution of this widespread SI system.

Results

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

Identification of pistil-expressed *S-RNase* genes in pummelo. We constructed 64 RNA sequencing (RNA-seq) libraries of styles and anthers from these Chinese pummelos (Supplementary Table 3). As a previous study[20] had suggested that pummelo had candidate *S-RNase* genes, we investigated this further. Nine candidate *S-RNase* genes with complete open reading frames and homology to previously reported *S-RNase*s were identified. We named these genes S_n*-RNase*, with *n* denoting the *S* haplotype (that is, S_1*-RNase* to S_9*-RNase*). Their full-length complementary DNA clones contained coding regions ranging from 660 bp to 699 bp (Supplementary Fig. 1) and encoded highly polymorphic proteins (38.1–76.7% deduced amino acid identity; Supplementary Fig. 2). Their predicted molecular masses (between 22.96 kDa and 24.47 kDa) and alkaline isoelectric points (7.67–9.39; Supplementary Table 4) were similar to known *S*-RNases[8]. The highly polymorphic citrus sequences contained key features of known functional *S*-RNases[9] (Fig. 1a; Supplementary Fig. 2). However, a comparison of these sequences with known *S*-RNases revealed that although the C2 and C3 domains were relatively well conserved (including the histidine residues implicated in catalysis), other domains were poorly conserved across species (Supplementary Fig. 2). An extra histidine residue was conserved across all nine citrus *S*-RNases, but was not present in the other *S*-RNases. The pummelo *S*-RNases contained five hypervariable regions, whereby two corresponded to the HVa and HVb domains in other species, but three were unique to pummelo (Supplementary Figs. 2 and 3). A phylogenetic analysis revealed that pummelo *S*-RNases clustered together with authentic *S*-RNases, but on a separate branch (Supplementary Fig. 4). This provides good evidence to indicate that these highly polymorphic pummelo sequences may be *S*-RNases.

We investigated the frequency of these nine *S-RNase* genes within natural pummelo populations, which comprised 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].

An analysis of various tissues using quantitative PCR with reverse transcription (qRT–PCR) 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 5 days before anthesis and decreased thereafter, western blots revealed that the protein was not detectable at this stage. However, it was detected 4 days before anthesis, and levels of the protein progressively increased until the pistils were mature (Fig. 1b; Supplementary Fig. 6c). Thus, these citrus RNases display the tissue and developmental specificity expected of a *S* determinant.

The pummelo *S-RNase*s segregate with *S* haplotype in a GSI manner. The *S* genotypes of 15 pummelo accessions were assigned on the basis of pollinations and aniline blue staining (Supplementary Fig. 7; Supplementary Table 5). As many of the examined pummelo accessions contained the nine identified *S-RNase*s, these 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, whereby each accession had a pair of *S*-RNase bands corresponding to that particular genotype in each of 15 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 (T_1 plants; Fig. 1d; Table 1). For a half-compatible cross (for example, 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 (that is, an absence of S_5S_5 and S_5S_6 genotypes). All of the 118 T_1 plants had either the S_2S_5 (56 plants) or S_2S_6 (62 plants) genotypes in the expected 1:1 ratio ($\chi^2 = 0.31$, $P = 0.58$; Table 1). They lacked S_5S_5 or S_5S_6 genotypes, which demonstrates that only S_2 pollen is compatible with S_5S_6 pistils, as expected for a GSI system. Reciprocal crosses (WB × SJ) yielded 59 T_1 progenies with either the S_2S_6 or S_5S_6 genotypes in a 1:1 ratio ($\chi^2 = 1.37$, $P = 0.24$). This half-compatibility was also observed in other tests whereby 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 that segregated at a 1:1:1:1 ratio as expected ($\chi^2 = 2.32$, $P = 0.51$; 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-RNase and S_2-RNase 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-RNase and S_2-RNase as glutathione *S*-transferase (GST) fusion proteins (Fig. 2a) and confirmed that they exhibited RNase activity (Fig. 2b,c). To establish whether these *S-RNase*s function as *S* determinants in citrus, we examined whether these fusion proteins specifically inhibit incompatible pollen tube growth in vitro (Fig. 2d; Supplementary Figs. 8 and 9). We used a bioassay similar to that used for *Papaver* SI[27,28]. While this assay 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 (that is, inhibition of pollen tube growth for 50% of the pollen tubes). The recombinant S_1-RNase-GST protein inhibited pollen tubes from plants with genotype S_1S_3 (half compatible; Fig. 2d, blue bars) by ~54% ($^{**}P < 0.001$) compared with its untreated control, while the compatible pollen genotype S_5S_6 (Fig. 2d, grey bars) was only inhibited by 9% compared with its untreated control ($P = 0.067$ (not significant (NS)); Fig. 2d (i)). Similarly, the S_2-RNase-GST inhibited pollen tubes from plants with genotype S_2S_8 by ~51% ($^{**}P < 0.001$) compared with its untreated control, while the compatible pollen genotype S_5S_6 (Fig. 2d, grey bars) was only inhibited by 1% compared with its untreated control ($P = 0.763$ (NS); Fig. 2d(ii)). Combined recombinant S_1-RNase and S_2-RNase fusion proteins inhibited pollen from plants with genotype S_1S_2 (an incompatible combination) by 62% ($^{**}P < 0.001$) compared with untreated pollen tubes. The same proteins had reduced inhibitory activity against compatible pollen from plants with genotype S_5S_6, with a 7% ($P = 0.113$ (NS)) reduction in length compared with its untreated

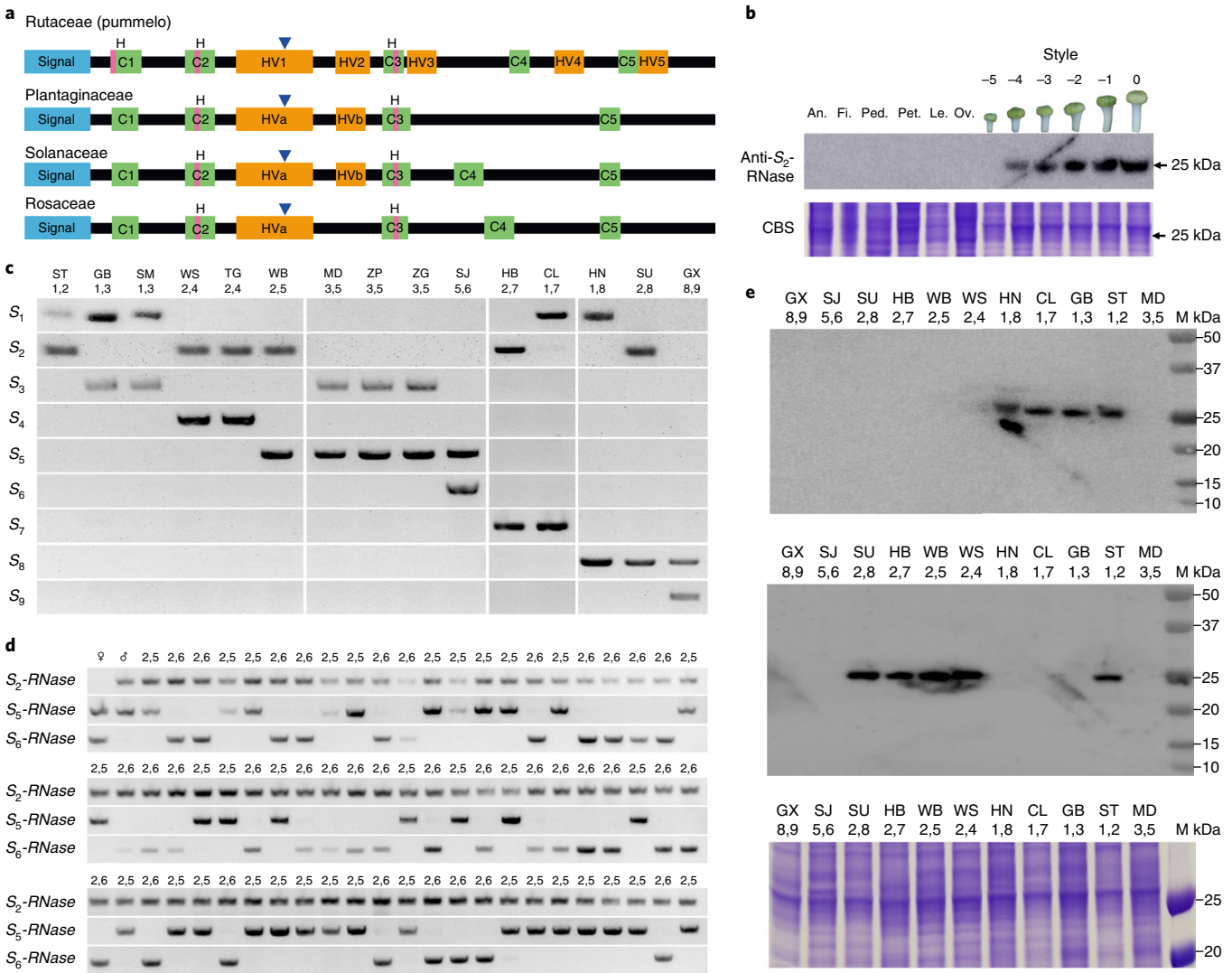

Fig. 1 | The citrus *S*-RNases exhibit key features of *S*-RNases. a, Cartoon showing key features of the pummelo *S-RNase* sequences compared with other *S*-RNases, including five conserved domains (C1–C5, green boxes) and hypervariable domains (HV1–HV5, orange boxes). All of the *S*-RNases have a signal peptide (blue box), two or three conserved histidine residues (H, pink) and a single intron (triangle). **b**, Upper: 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 an ~25 kDa protein in extracts from mature pistils (0, open flower); no protein was detected at −5 days before anthesis, low expression was detected at −4 days and this increased over time as the pistil matured. The S_2-RNase antisera did not cross-react with a protein in other tissues, including anther (An.), filament (Fi.), pedicel (Ped.), petal (Pet.), leaf (Le.) and ovary (Ov.). Lower: Coomassie blue staining (CBS) shows equal loading. **c**, The *S* genotypes of 15 pummelo accessions (indicated above each lane) were assigned using aniline blue staining of pollinated pistils (Supplementary Table 3). PCR of leaf DNA, using *S*-RNase-specific primers (indicated left: S_1 to S_9), showed two *S*-allele-specific transcripts for S_1*-RNase* to S_9*-RNase* (S_1–S_9) amplified from each pistil, which correspond to those assigned by pollination. **d**, *S-RNases* segregate with the *S* locus in F_1 progeny. A pistil (♀) from a pummelo plant (accession SJ) assigned genotype 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 showed 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 70 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 the lanes) using antibody raised against the recombinant S_1-RNase (upper) and S_2-RNase (middle); CBS (lower) 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 that there is a direct link between the *S-RNase* cloned (through the antibody to the recombinant protein) and *S* alleles carried by the plant. M, molecular weight marker. Experiments were repeated independently twice for **c**, **d** and **e** and three times for **b**, with similar results obtained for each.

control. The combined S_1-RNases 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%, respectively, of pollen tube length compared with their respective untreated controls (**$P < 0.001$ for both). Together, these data provide evidence to indicate that the *S*-RNases have *S*-specific pollen inhibitory activity. These data also demonstrate that although pummelo pollen does not grow to the same extent as in vivo (probably because of the absence of key pistil components in vitro), and despite some non-specific inhibitory activity by the recombinant proteins, pollen of

Table 1 | Pummelo *S-RNase*s in F_1 progenies segregate in a gametophytic manner

Phenotype[a]	Genetic cross	No. of progeny	Possible genotypes[b]	Observed ratio[c]	Expected ratio[d]	X^2 value	*P* value
Fully compatible	SJ (S_5S_6) × ST (S_1S_2)	77	S_1S_5: S_2S_5: S_1S_6: S_2S_6	22:17:15:23	1:1:1:1	2.32	0.51 (NS)
					1:1:1:1	2.32	0.51 (NS)
Half compatible	SJ (S_5S_6) × WB (S_2S_5)	118	S_5S_5: S_5S_6: S_2S_5: S_2S_6	0:0:56:62	0:0:1:1	0.31	0.58 (NS)
					1:1:1:1	113.53	1.54×10^{-25}**
	WB (S_2S_5) × SJ (S_5S_6)	59	S_2S_5: S_5S_5: S_2S_6: S_5S_6	0:0:34:25	0:0:1:1	1.37	0.24 (NS)
					1:1:1:1	61.75	2.49×10^{-13}**
	HB (S_2S_7) × ST (S_1S_2)	115	S_2S_2: S_2S_7: S_1S_2: S_1S_7	0:0:53:62	0:0:1:1	0.7	0.4 (NS)
					1:1:1:1	116.41	4.58×10^{-25}**
	ST (S_1S_2) × HB (S_2S_7)	42	S_1S_2: S_2S_2: S_1S_7: S_2S_7	0:0:21:21	0:0:1:1	0	1.0 (NS)
					1:1:1:1	42	4.01×10^{-9}**
	GX (S_8S_9) × SU (S_2S_8)	76	S_8S_9: S_9S_9: S_2S_8: S_2S_9	0:0:39:37	0:0:1:1	0.05	0.82 (NS)
					1:1:1:1	76.11	2.10×10^{-16}**
	GB (S_1S_3) × MD (S_3S_5)	113	S_1S_3: S_3S_3: S_1S_5: S_3S_5	0:0:62:51	0:0:1:1	1.07	0.3 (NS)
					1:1:1:1	115.14	8.58×10^{-25}**

Segregation analysis of *S* haplotypes of F_1 progenies of pummelo accessions for pollinations in half-compatible and fully compatible combinations were assigned using PCR (Fig. 1f). Outcomes show segregation ratios as expected for a GSI system. [a]The pollination phenotype was determined by aniline blue staining (see Fig. 1a–d). [b]The observed genotypes are indicated with an underline. [c]The *S*-genotype ratios observed in all of the progeny. [d]An upper segregation ratio is expected from a GSI system, whereas a lower segregation ratio is expected from simple Mendelian inheritance. All crosses with parents sharing a *S-RNase* haplotype showed a result consistent with GSI, with a non-significant Chi square value for this prediction and a highly significant difference (**$P < 0.001$) for the lower segregation ratio. These data provide clear evidence that pummelo *S-RNase*s segregate with the *S* locus in a GSI manner.

different haplotypes is affected specifically and differentially by the recombinant S_1-RNase and S_2-RNase fusion proteins. Although they may not exactly reflect the in vivo situation, 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 and provide 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 bacterial artificial chromosome (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. A 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 12 F-box genes associated with it, as well as other genes, including transposons (Supplementary Tables 8 and 9). The F-box genes on the S_1-locus have 33.6–74.2% deduced amino acid identity, which is comparable to that of the F-box genes at the S_2-locus (33.5–73.9%). Nine F-box genes exhibited relatively high sequence divergence (78.1–93.7% deduced amino acid identity) between the two *S* loci, and three F-box genes were highly conserved (99.5–99.7% deduced amino acid identity) and may be SFL-like (*SLFL*) genes.

A RNA-seq analysis revealed that all the *SLF*s were specifically expressed in anthers (Supplementary Fig. 11), and qRT–PCR verified this, identifying expression of the *SLF*s in anthers, pollen and pollen tubes (Supplementary Fig. 12). A 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 indicates *SLF*s are 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 spanned 198 kb to 370 kb, and each of these contained one *S-RNase* and 11–17 *SLF* or *SLFL* genes (Fig. 3a; Supplementary Fig. 14). An analysis of 117 *SLF* and *SLFL* genes revealed that they clustered into 12 types. We designated the F-box of each locus as S_n-*SLFx* or S_n-*SLFLx* (where *n* indicates the *S* haplotype and *x* the type; Supplementary Fig. 15). The pollen and pistil *S* determinants should exhibit evidence of co-evolution. Examining the synonymous (K_s) and nonsynonymous (K_a) substitution rates revealed that those of the *S*-RNases ($K_s = 0.814$, $K_a = 0.503$) and each SLF/SLFL type ($K_s = 0.977–1.047$, $K_a = 0.422–0.461$) were similar and much higher than the inter-allelic K_a and K_s values of each SLF/SLFL type ($K_s = 0.015–0.476$, $K_a = 0.009–0.156$; Fig. 3b). These data suggest that the *S-RNase* and intra-haplotypic *SLF* genes co-evolved and are probably similarly ancient.

Similar to *Petunia*[6], the citrus SLF and SLFL proteins showed extensive polymorphism between types (44.24–46.52% identity), while the sequence identities between allelic variants of each type were more highly conserved (74.78–97.49% identity; Fig. 3b). The clustering of the *SLF* sequences, together with intra-haplotypic versus inter-allelic differences is consistent with the non-self-recognition model of *S*-RNase/SLF evolution, which proposes that divergent/deleted *SLF* genes 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, and 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 as the target of the more conserved SLF1 proteins (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 (refs. [6,29]), and we also found missing SLF proteins. 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 to indicate that the *S-RNase* genes and intra-haplotypic *SLF* and *SLFL* genes probably co-evolved and that the divergence of the inter-allelic *SLF* and *SLFL* genes from

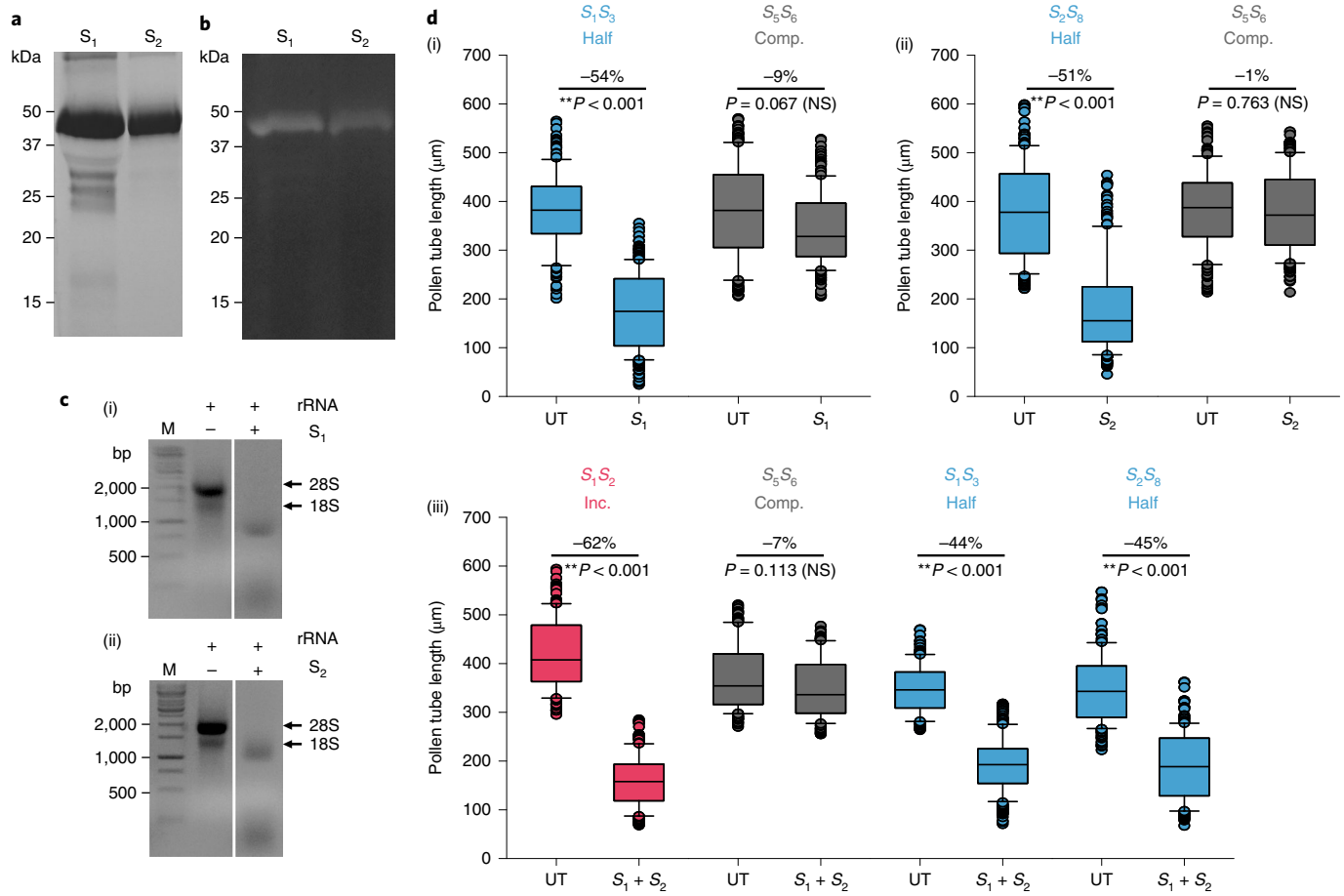

Fig. 2 | The pummelo *S*-RNases exhibit RNase activity and elicit *S*-specific pollen inhibition in an in vitro bioassay. a, Recombinant S_1-RNase-GST (S_1) and S_2-RNase-GST (S_2) proteins migrate to ~45 kDa on SDS–PAGE. **b**, An in-gel RNase assay showed that the recombinant S_1-RNase-GST and S_2-RNase-GST 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 **a**–**c** were repeated independently three times, with similar results obtained. **d**, Recombinant S_1- and S_2-RNases (S_1 and S_2, respectively) inhibit pollen tube growth differentially. (i) Individually, S_1 partially inhibited pollen tubes from plants with genotype S_1S_3 that were half compatible (blue bar, half) by ~50% compared with its untreated (UT) control, while the compatible pollen genotype S_5S_6 (grey bars, Comp.), was only inhibited by ~10% compared with its untreated control. (ii) Similarly, the S_2-RNase inhibited pollen tubes from plants with genotype S_2S_8 (also half compatible, blue bar) by ~50% compared with its untreated (UT) control. Together, these data provide evidence to indicate that the *S*-RNases have *S*-specific pollen inhibitory activity. (iii) Combined recombinant S_1- and S_2-RNases (S_1+S_2, respectively) inhibited incompatible pollen from pummelo plants with genotype S_1S_2 (red bars, Inc.) by 62% compared with untreated pollen tubes. 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 with untreated controls; this was considered nonspecific activity. 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 (blue bars), with an ~45% reduction in pollen tube length compared with their respective untreated 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 (boxes indicate the upper and lower quartiles, with the centre lines representing the median, lines above and below indicate the range, and dots indicate the outliers). For each treatment, the mean reduction of pollen tube length (%) compared with its pairwise control is shown within each box. One-way ANOVA was used to compare the pollen tube length of treatment versus untreated control (two asterisks indicate a significantly different result).

each type occurred more recently. Together, our findings are consistent with the non-self-recognition model of *S*-RNase/SLF evolution. It is well established that this mechanism is utilized in species with *S*-RNase and SLFs that were confirmed to function as *S* determinants in SI; therefore, our results contribute to evidence showing that SI in pummelo is probably controlled by *S-RNase* and *SLF* genes that act 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 *Citrus 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, which resulted 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 histidine residues, but lacks the C4 and C5 conserved domains, the HV4 and HV5 hypervariable domains and four conserved cysteine residues (Fig. 4a). Because the non-mutated progenitor of the S_m-*RNase* was not identified in the accessions, we engineered a 'recovered' version (named $S_m{}^R$-RNase) through the 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).

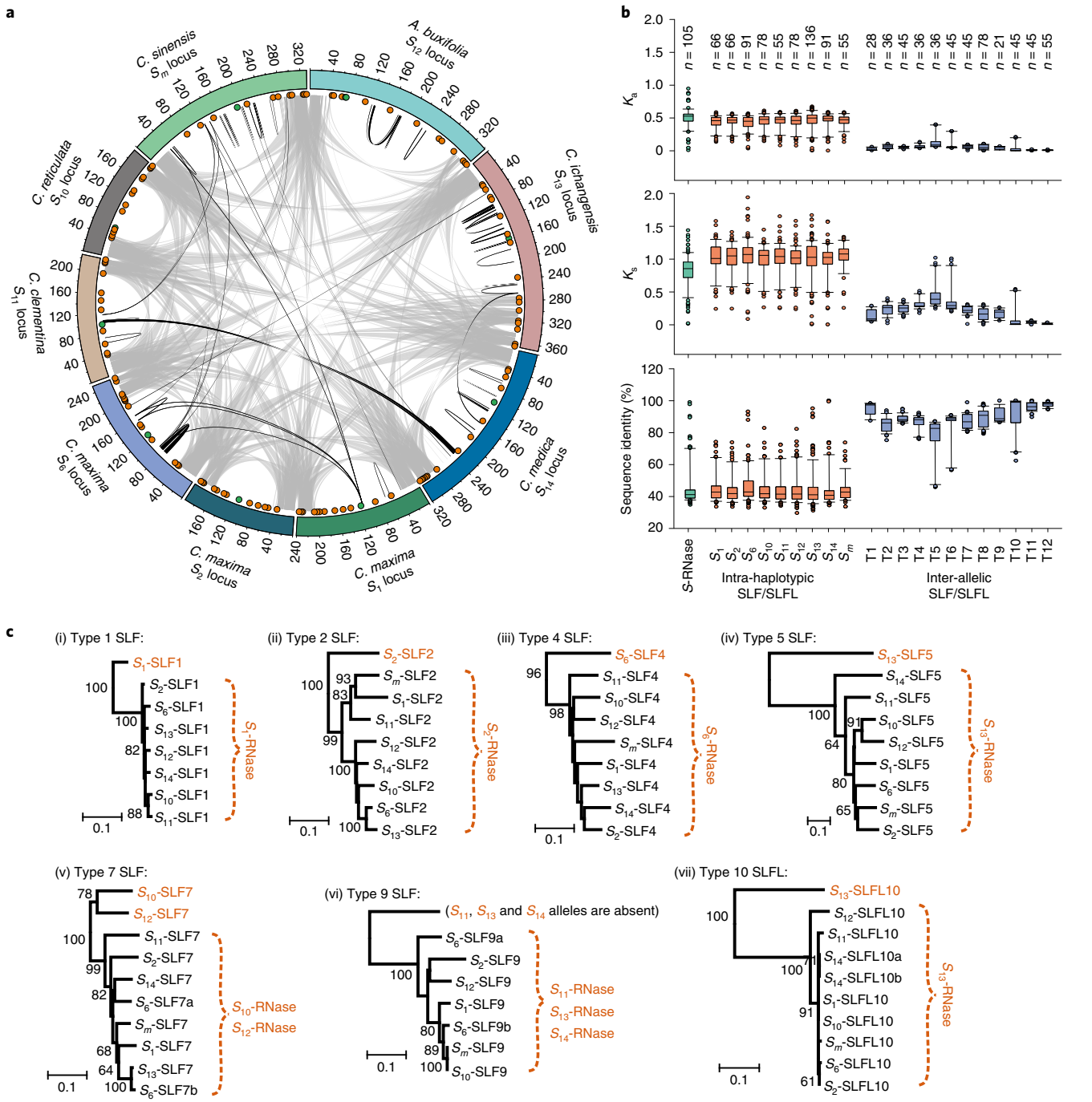

Fig. 3 | Multiple candidate *SLF* and *SLFL* genes are located at the citrus *S* locus. a, Gene synteny of nine citrus *S* loci. Grey lines indicate syntenic regions at the end of the *S* locus. Black lines indicate syntenic sequences in the intergenic regions of the locus. The *S-RNase*s (green dots) and *SLF*s and *SLFL*s (orange dots) are indicated on the ideogram of each locus (see Supplementary Fig. 7 for more detail). **b**, Sequence identity (bottom), synonymous (K_s, middle) and non-synonymous (K_a, top) substitutions in *S-RNase*s (green boxplot), intra-haplotypic *SLF*s and *SLFL*s (S_1 to S_m; orange boxplots) and inter-allelic *SLF*s and *SLFL*s (T1 to T12; blue boxplots). Box and whisker plots show the distribution of the sequence identity, K_s and K_a values for pairwise *S-RNase*s and *SLF*s and *SLFL*s. Boxes indicate the upper and lower quartiles, the centre lines the median, lines above and below indicate the range, and dots indicate the outliers. The number of independent comparisons (*n*) are indicated at the top of each box. **c**, Phylogenetic trees of SLFs 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; for example S_6-SLF7a and S_6-SLF7b. The *S*-RNases (also in orange) that are cognate to the diverged or deleted SLFs are predicted to interact with the conserved SLFs within the brackets under the non-self-recognition model.

We hypothesized that the truncated S_m-RNase is responsible for the loss of functional SI in the SC accessions. We first examined the level of mRNA expression of the S_m-*RNase*, as SC in other species have shown low *S-RNase* expression[30,31]. An analysis of a range of tissues revealed that the expression of S_m-*RNase* was minimal compared with S_2-*RNase* (Fig. 4b). Absolute qRT–PCR confirmed

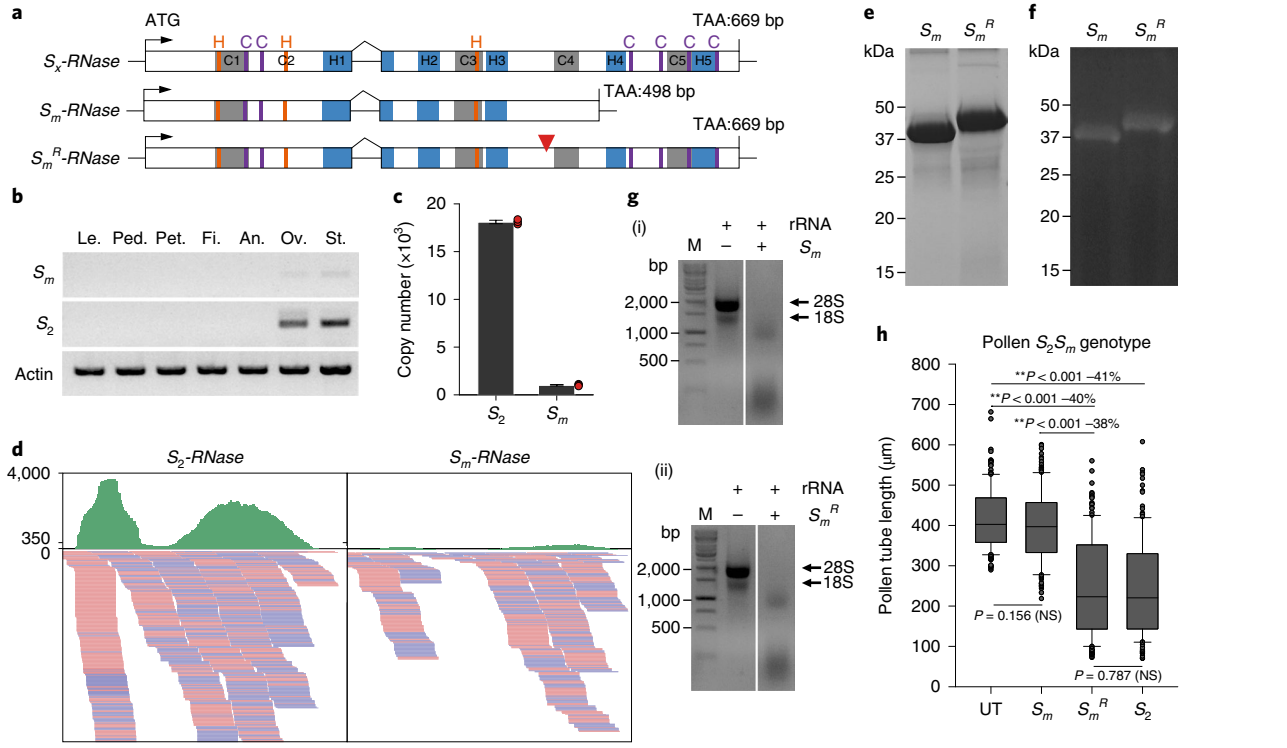

Fig. 4 | A truncated S_m-*RNase* appears to be responsible for the loss of SI in citrus. a, Cartoon showing the gene structure of the S_x-*RNase* (upper) and its mutant form (middle). S_x-*RNase* indicates the length and structure of the normal *S-RNase* with five conserved domains (C1–C5, grey boxes) and five hypervariable domains (H1–H5, blue boxes). S_m-*RNase* harbours a single nucleotide deletion resulting in a premature stop codon. The 'recovered' S_m^R-RNase (lower) has a single nucleotide inserted (red triangle) that recovered the full-length gene. Conserved histidine and cysteine residues are indicated in orange and purple, respectively, based on the deduced amino acid sequences. **b**, Expression of S_m-*RNase* (S_m) and S_2-*RNase* (S_2) quantified using RT–PCR of different tissues from a S_2S_m plant. The S_m-*RNase* is expressed at a much lower level than the S_2-*RNase* in a tissue-specific manner (only in ovary and style tissues in pistils). **c**, Absolute copy number of the S_2-*RNase* and S_m-*RNase* (extrapolated from the equation shown in Supplementary Fig. 18). The copy number of S_m-*RNase* in 50 μg of S_2S_m RNA extracted from styles is much lower than that of S_2-*RNase*. Error bars are shown for mean copy number ± s.e.m. ($n = 3$ biological replicates; red dots indicate individual samples). **d**, Integrative Genomics Viewer tracks displaying sequencing read clusters of S_2-*RNase* and S_m-*RNase* genes from RNA-seq of S_2S_m from styles. The green bars depict the number of the reads mapped to the reference. There are clearly more reads mapped to S_2-*RNase* than for 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-RNase (S_m) and S_m^R-RNase (S_m^R). The molecular weight 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-RNase and S_m^R-RNase. The mutated S_m-RNase and the recovered S_m^R-RNase have similar RNase activities. **g**, The recombinant S_m-RNase (i) and S_m^R-RNase (ii) both degrade citrus 28S and 18S rRNA (agarose gel assay). Experiments were repeated independently twice for **b** and three times for **e** and **g**, with the similar results obtained for each experiment. **h**, The S_m^R-RNase recombinant protein displays inhibitory activity against pollen. Boxplots show the distribution of individual pollen tube lengths in an in vitro bioassay of recombinant S_m-, S_m^R- and S_2-RNases against pollen from a plant with the S_2S_m genotype. The S_m-RNase did not significantly inhibit pollen tubes from a S_2S_m plant compared with untreated pollen; the S_m^R-RNase exhibited significant inhibitory activity, reducing the length of pollen tubes by ~40% and was not significantly different 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 was used to compare the pollen tube lengths (treatment versus untreated control). The elements in box and whisker plots are the same as in Fig. 2d.

that the expression of the S_m-*RNase* transcript was greatly reduced (Fig. 4c; Supplementary Fig. 18), and RNA-seq confirmed this (Fig. 4d). These data suggest that the SC phenotype could be due to the reduced S_m-*RNase* transcript level. We next expressed the recombinant S_m-RNase-GST and S_m^R-RNase-GST fusion proteins (Fig. 4e). Both types of fusion proteins 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 the 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 (Fig. 4h; Supplementary Fig. 19). The recombinant S_m-RNase-GST fusion protein did not significantly inhibit pollen tube growth compared with the untreated control ($P = 0.156$ (NS), analysis of variance (ANOVA)). This lack of pollen inhibitory activity for the S_m-RNase suggests that this mutation could be responsible for the SC phenotype. As the S_m-*RNase* does not contain the hypervariable domains HV4 and HV5, it is possible that specificity may reside here. In support of this idea, a predicted structural analysis suggests that these regions reside at the surface of the protein (Supplementary Fig. 20). In contrast, treatment of pollen from a S_2S_m genotype plant with the recovered S_m^R-RNase-GST protein resulted in inhibition of growth, with pollen tube lengths significantly reduced compared with the S_m-RNase (**$P < 0.001$, ANOVA) and was not significantly different from the (half-compatible) pollen inhibitory activity displayed by the S_2-RNase-GST fusion protein ($P = 0.787$ (NS), ANOVA). Moreover, as the recovered S_m^R-RNase exhibited a gain of pollen inhibitory activity, this is consistent with the explanation that truncation of this gene may be responsible for loss of activity and the SC

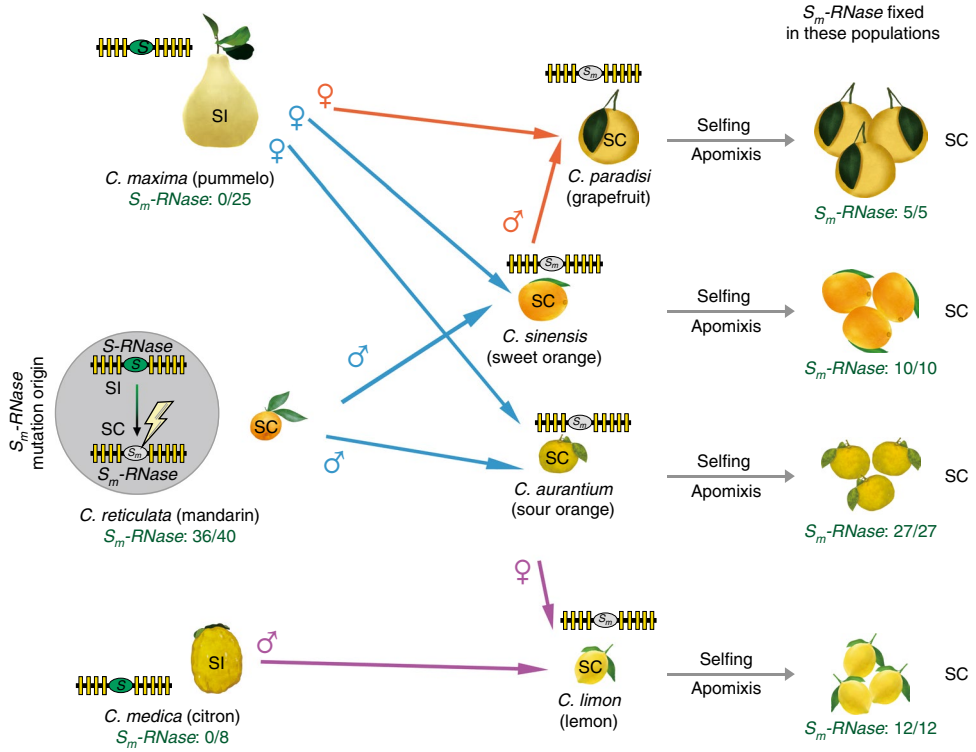

Fig. 5 | Postulated spread of the S_m-*RNase* and SC in *Citrus*. Left: pummelo (upper) is SI (no S_m-RNases were identified in 25 accessions). Citron (lower) had a previously uncharacterized reproduction strategy, and we propose it is SI, as we identified a single *S* locus with one *S-RNase* and approximately nine *SLF*s (and no S_m-RNases detected) in eight accessions. We show a cartoon of the *S* locus, with one *S-RNase* (green ellipse) and approximately nine *SLF*s (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 out of 40 of mandarin accessions, 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, as previously proposed[34], and the crosses are probably 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: 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 examined accessions being SC and containing the S_m-*RNase* (5/5, 10/10, 27/27 and 12/12, respectively); none are SI. We propose that selfing and apomixis allowed the S_m-*RNase* to become fixed in these populations.

phenotype. Thus, although 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 identified 15 *S-RNase*s (Supplementary Table 11). These 15 *S-RNase*s were present in 132 of these accessions. Each *S-RNase* occurred at a low frequency, which is in keeping with it being maintained by negative frequency-dependent selection (Supplementary Fig. 21). An analysis of the relationships of the 15 *S-RNase*s to investigate how they spread through citrus species revealed that the phylogeny of the *S-RNase*s did not fit the phylogeny of citrus species as previously described[32] (Supplementary Fig. 22), which suggests 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 self-incompatible (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 initially arose 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; however, 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 identified several polymorphic pistil-expressed *S-RNase*s from pummelo and showed that they segregate with *S* haplotypes. We provided strong evidence that citrus utilize the *S*-RNase-based SI system and that *S*-RNases function as pistil *S* determinants, inhibiting pollen in a *S*-specific manner. Phylogenetically, *S-RNase*s are found in several divergent families; however, 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 past 25 years, although putative *S*-RNases have been identified in Rubiaceae[35,36]. Our identification of a functional *S*-RNase SI system in citrus, which diverged

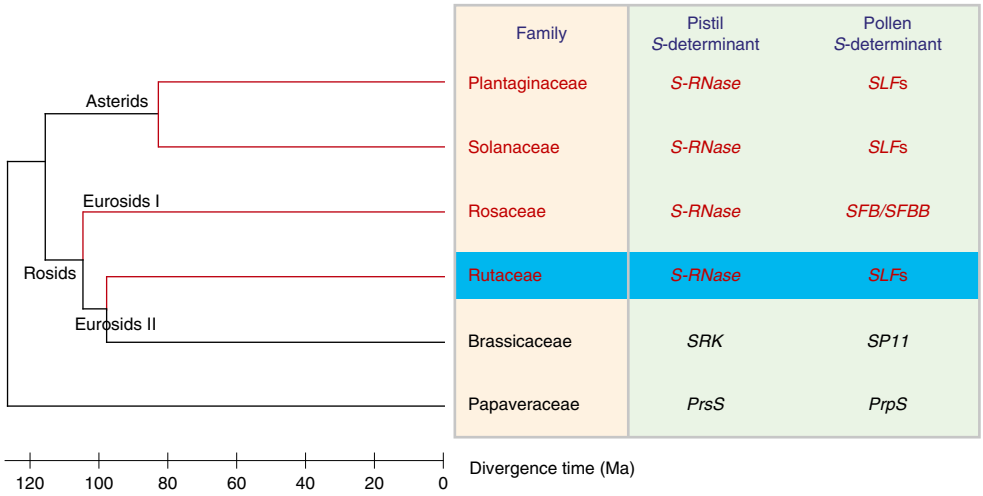

Fig. 6 | Phylogenetic relationship of different SI systems in flowering plants. Phylogenetic tree showing relationships between the SI families (yellow box) with established *S* determinants (green box). Bold text indicates the *S-RNase* and *SLF*s (blue box) identified in this study. The *S*-RNase-based SI systems (red text) all use *S-RNase* as the pistil *S* determinants and *SLF*s, *S*-haplotype-specific F-boxes (*SFB*s) or *S*-locus F-box brothers (*SFBB*s) as pollen *S* determinants, 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 Ma). The most plausible interpretation for this is that the *S-RNase*s in the core eudicots have a single origin and evolved only once, before the divergence of these families >100 million years ago (Ma); this is more likely than several parallel gains of *S-RNase* at least three times. The tree is based on APG III[61] and the divergence time is based on TimeTree (http://timetree.org/).

~110 million years ago (Ma) from the Solanaceae (Fig. 6), supports the idea that *S*-RNases have a single origin, before 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 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. This is because a functional *S* haplotype cannot encode a *SLF* that recognizes its own *S*-RNase; therefore, either a diverged or deleted allele of that *SLF* type is utilized. Thus, within a *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-boxes (*SLF*s) tightly linked to each *S*-RNase suggests that the *S* locus in citrus fits this model. An 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 probably involved in SI.

For many species, the evolutionary history of the SI–SC transition (or transitions) is unclear[37]. Here, we began 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, we identified a frameshift mutation in the female *S*-determinant S_m-*RNase*, which yields a truncated *S-RNase*, and provided evidence that it is responsible for the loss of SI. The prevalence of this mutant in citrus populations suggests 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, whereby the loss of SI is often accompanied by the complete deletion of the *S-RNase* from the *S* locus[38,39], although an exception has been reported[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 provided evidence that SI in citrus utilizes a *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 separated >100 Ma from the nearest *S*-RNase family, our data will help clarify the distribution of *S*-RNase-based SI systems and their evolution. We provided evidence that SI is ancestral and showed that a truncated S_m-*RNase* is responsible for the loss of SI. This allowed us to decipher the evolutionary history of the SI–SC transition in >150 citrus accessions. Selfing, combined with apomixis and selection of SC by breeders, makes this an interesting example of the evolution of plant reproductive strategies.

Methods

Plant materials. To analyse the *S* allele that controls SI in citrus, a natural population of 391 pummelo accessions were collected in the wild (Supplementary Fig. 3a). Among them, 15 pummelo cultivars were used to perform the pollination assay and 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-pollinations, self-pollinations 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] (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 using a previously described method[41]. The RNA was used for high-throughput RNA-seq library construction and sequenced using the Illumina Hiseq 2500 platform (Supplementary Table 3). Approximately 13 Gb reads per sample (read length of 150 bp) were generated. Clean data were de novo assembled separately for each citrus accession using Trinity (v.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 (ref. [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*-RNases involved in SI, six nucleotide sequences encoding *S-RNases* from species with *S*-RNase-based SI were downloaded from NCBI (HE805271.1 and AJ315593.1 from Antirrhinum; D63887.1 and AB568389.1 from Solanaceae; and FJ543097.1 and AF327223.1 from Rosaceae) and aligned according to the codons using ClustalW in MEGA7 (ref. [44]). Using this alignment, a *S-RNase* hidden Markov model profile was built using the Hmmbuild subprogram in HMMER[45]. The Trinity transcripts were queried with this profile using nhmmer.

***SLF* identification.** A 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*s were sequenced using an 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*, *Citrus medica*, *Citrus ichangensis*, *Atalantia buxifolia* and *C. reticulata* were downloaded from http://211.69.140.136/orange/index.php; the genome of *Citrus clementina* was downloaded from https://phytozome.jgi.doe.gov/pz/portal.html. Based on the conserved sequences at each end of the 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 a F-box domain and a F-box associated motif were designated as *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-*RNase* 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 library (Supplementary Table 6). The cDNA fragments were amplified using RT–PCR per standard PCR protocols. All PCR products were cloned into a pEASY-Blunt Cloning Vector (TransGen Biotech) and sequenced using Sanger sequencing technology. Deduced amino acid sequences were aligned using ClustalW in MEGA7 (ref. [44]), and sequence similarity was illustrated by shading with GeneDoc 2.602. The normalized variability index value for *S-RNase* genes was calculated with a sliding window of seven residues as previously described[47].

***S_m-RNase* identification and sequence cloning.** The mutated S_m-*RNase* was identified within the S_m-locus from the sweet orange genome. Primers for the amplification of S_m-*RNase* were designed based on the genomic sequence. The cDNA fragment of S_m-*RNase* was amplified as described above. To examine the function of the non-mutated 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-RNases, S_m-RNases and S_m^R-RNases were made using the I-TASSER server (https://zhanglab.ccmb.med.umich.edu/)[48] and the PyMol molecular visualization package v.2.0.

Quantitative analyses. qRT–PCR and western blotting were carried out to check gene expression and translation, respectively, in different tissues as previously described[20]. Heatmaps for expression were drawn using TBtools[49]. RNA-seq reads were aligned to the *S* locus using TopHat2, and the alignment result was visualized using Integrative Genomics Viewer[50,51]. The uniquely mapped reads without any mismatch were used to calculated the FPKM of the genes on the *S* locus with Cufflinks[52].

An absolute quantification method was employed to analyse S_2-*RNase* and S_m-*RNase* expression levels. The plasmids inserted into full-length S_2-*RNase* and S_m-*RNase* were used to make tenfold serial dilutions of DNA template from 15 ng μl^{-1} to 1.5 fg μl^{-1}. The PCR system and thermocycler conditions were same as that for qRT–PCR. The Ct values (*y* axis) and the log gene copy number (*x* axis) were used to generate a standard curve, and the PCR efficiency was calculated as previously described[53]. Plasmid DNA standard curve equations (Supplementary Fig. 18) were used to calculate the absolute copy number of S_2-*RNase* and S_m-*RNase* within 50 ng of pistil mRNA from the S_2S_m plant.

Phylogenetic analyses. The deduced amino acid sequences were aligned using MAFFT[54] and manually adjusted by removing spurious alignments and long gaps. RAxML[55] was used to construct a maximum likelihood tree under the substitution model PROTGAMMAWAG with 1,000 bootstrap replications. To estimate synonymous (K_s) and nonsynonymous (K_a) substitution rates in DnaSP[56], the aligned protein sequences were converted to nucleotide alignments. Species divergence was obtained from the mean estimate time in TimeTree[57].

Expression of *S*-RNase recombinant proteins. The open reading frames from S_1-, S_2-, S_m- and S_m^R-RNases without signal peptide regions were expressed in *Escherichia coli* strain BL21 (DE3) (TransGen) as GST fusion proteins using pGEX-6P-1 (GE Healthcare). *E. coli* strains were induced by 0.2 mM isopropyl-1-thio-β-galactoside for 16 h at 18 °C, and glutathione Sepharose 4B bead (GE Healthcare) protein was used for protein purification according to the manufacturer's protocol. These GST fusion proteins were analysed using SDS–PAGE and western blotting. Anti-S_1-RNase and anti-S_2-RNase antibodies were raised against the S_1-RNase-GST and S_2-RNase-GST fusion proteins, respectively, in rabbit and used at a 1:2,000 dilution (anti-S_1-RNase) and 1:1,000 dilution (anti-S_2-RNase), with a goat anti-rabbit IgG-horseradish peroxidase 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 as previously described[58], but with slight modifications. Recombinant *S*-RNase protein (20 μg) 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 at pH 8.0.

We also performed an RNase activity assay of the recombinant *S*-RNases using citrus RNA from pistils as a target. *S*-RNase (10 μg) and total RNA (2 μg) isolated from citrus pistil were incubated at 37 °C for 1 h in a 20-μl reaction mixture[59]; rRNA was then separated on a 1% agarose gel and stained with ethidium bromide and examined for degradation.

In vitro pollen bioassays to assess *S*-RNase pollen inhibitory activity. 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-RNase-GST and S_2-RNase-GST fusion proteins 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 the genotype S_1S_2. Half-compatible combinations were achieved with recombinant S_1-RNase versus pollen from plant genotypes S_1S_3 or S_1S_2, and recombinant S_2-RNase versus pollen from plant genotypes S_1S_2 or S_2S_8. Fully compatible tests used pollen from plant genotype S_5S_6. For functional examination of the pollen inhibitory activity of the recombinant S_m-RNase-GST and the recovered S_m^R-RNase-GST, pollen from plants with the genotype S_2S_m provided a half-compatible test; recombinant S_2-RNase-GST provided a positive control for maximal inhibitory activity (half compatible). GST was used as the untreated control for all tests.

A germination medium (GM)[20] was used to grow pollen tubes in vitro. Before each bioassay, the recombinant GST fusion proteins were dialysed 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; that is, to obtain maximal inhibitory activity with minimal nonspecific activity (Supplementary 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, $n \geq 150$ in total) were measured using Image-Pro Plus v.6.0 (Media Cybernetics). Because we show actual pollen tube lengths, we display the data in pairwise comparisons, and each test had its appropriate control. Data are 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[60] was used for statistical analyses of the nonzero coverage (≥1 reads) of each alignment. For the incomplete alignments with nonzero coverage $>0.9<1.0$, we cloned the full-length sequence used for these alignments and analysed these sequences using Sanger sequencing. The *S* haplotypes of 153 citrus accessions are summarized in Supplementary Table 11.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The RNA-seq data shown in Supplementary Table 3 (for pummelo and grapefruit) are available at NCBI BioProject ID under accession codes PRJNA526584 and PRJNA573625. The sequence data shown in Supplementary Table 7 of the pummelo S_1-locus and S_2-locus BAC clones are available at NCBI BioProject ID under accession codes PRJNA573817 and PRJNA573818, respectively. The DNA sequencing data shown in Supplementary 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*) and PRJNA573624 (*C. reticulata*). The sequence data of the 15 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 and MN652912. Any other raw data are available from the corresponding author upon request.

Received: 24 May 2019; Accepted: 8 January 2020;
Published online: 13 February 2020

References

1. de Nettancourt, D. *Incompatibility and Incongruity in Wild and Cultivated Plants* Vol. 3 (Springer, 2001).
2. Xue, Y., Carpenter, R., Dickinson, H. G. & Coen, E. S. Origin of allelic diversity in antirrhinum *S* locus RNases. *Plant Cell* **8**, 805–814 (1996).
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.* **250**, 547–557 (1996).
4. McClure, B. A. et al. Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. *Nature* **342**, 955–957 (1989).
5. Takayama, S. & Isogai, A. Self-incompatibility in plants. *Annu. Rev. Plant Biol.* **56**, 467–489 (2005).
6. Kubo, K. et al. Collaborative non-self recognition system in S-RNase–based self-incompatibility. *Science* **330**, 796–799 (2010).
7. Fujii, S., Kubo, K.-I. & Takayama, S. Non-self- and self-recognition models in plant self-incompatibility. *Nat. Plants* **2**, 16130 (2016).
8. Ramanauskas, K. & Igić, B. The evolutionary history of plant T2/S-type ribonucleases. *PeerJ* **5**, e3790 (2017).
9. Igić, B. & Kohn, J. R. Evolutionary relationships among self-incompatibility RNases. *Proc. Natl Acad. Sci. USA* **98**, 13167–13171 (2001).
10. Krajewski, A. J. & Rabe, E. Citrus flowering: a critical evaluation. *J. Hortic. Sci.* **70**, 357–374 (1995).
11. Ngo, B. X., Wakana, A., Kim, J. H., 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.* **55**, 67–72 (2010).
12. Ngo, B. X., Wakana, A., Park, S. M., Nada, Y. & Fukudome, I. Pollen tube behaviors in self-incompatible and self-compatible *Citrus* cultivars. *J. Fac. Agr. Kyushu Univ.* **45**, 443–457 (2001).
13. Yamamoto, M., Kubo, T. & Tominaga, S. Self- and cross-incompatibility of various citrus accessions. *J. Jpn. Soc. Hort. Sci.* **75**, 372–378 (2006).
14. Miller, A. J. & Gross, B. L. From forest to field: perennial fruit crop domestication. *Am. J. Bot.* **98**, 1389–1414 (2011).
15. Zhang, S. et al. Reproduction in woody perennial *Citrus*: an update on nucellar embryony and self-incompatibility. *Plant Reprod.* **31**, 43–57 (2018).
16. Hand, M. L. & Koltunow, A. M. G. The genetic gontrol of apomixis: asexual seed formation. *Genetics* **197**, 441–450 (2014).
17. Soost, R. K. The incompatibility gene system in citrus. In *Proc. First Int. Citrus Symp.* (ed. Chapman, H.) 189–190 (Univ. California, Riverside, 1968).
18. Kim, J. H. 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. Jpn. Soc. Hort. Sci.* **80**, 404–413 (2011).
19. Miao, H., Ye, Z., Silva, J. A., Qin, Y. H. & Hu, G. Identifying differentially expressed genes in pollen from self-incompatible "Wuzishatangju" and self-compatible "Shatangju" mandarins. *Int. J. Mol. Sci.* **14**, 8538–8555 (2013).
20. Liang, M. et al. Genome-wide identification and functional analysis of S-RNase involved in the self-incompatibility of citrus. *Mol. Genet. Genomics* **292**, 325–341 (2017).
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* **290**, 365–375 (2015).
22. Castric, V. & Vekemans, X. Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. *Mol. Ecol.* **13**, 2873–2889 (2004).
23. Wright, S. The distribution of self-sterility alleles in populations. *Genetics* **24**, 538–552 (1939).
24. Goodwillie, C., Kalisz, S. & Eckert, C. G. The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* **36**, 47–79 (2005).
25. Stone, J. L. Molecular mechanisms underlying the breakdown of gametophytic self-incompatibility. *Q. Rev. Biol.* **77**, 17–32 (2002).
26. McClure, B., Cruz-García, F. & Romero, C. Compatibility and incompatibility in S-RNase-based systems. *Ann. Bot.* **108**, 647–658 (2011).
27. Franklin-Tong, V. E., Lawrence, M. J. & Franklin, F. C. H. An in vitro bioassay for the stigmatic product of the self-incompatibility gene in *Papaver rhoeas* L. *New Phytol.* **110**, 109–118 (1988).
28. Franklin-Tong, V. E. in *Self-Incompatibility in Flowering Plants: Evolution, Diversity, and Mechanisms* (ed. Franklin-Tong, V. E.) 237–258 (Springer, 2008).
29. Kubo, K. et al. Gene duplication and genetic exchange drive the evolution of S-RNase-based self-incompatibility in *Petunia*. *Nat. Plants* **1**, 14005 (2015).
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*. *Proc. Natl Acad. Sci. USA* **91**, 6511–6514 (1994).
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.* **12**, 223–227 (1997).
32. Wang, X. et al. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nat. Genet.* **49**, 765–772 (2017).
33. Yang, X. M. et al. Molecular phylogeography and population evolution analysis of *Citrus ichangensis* (Rutaceae). *Tree Genet. Genomes* https://doi.org/10.1007/s11295-017-1113-4 (2017).
34. Wu, G. A. et al. Genomics of the origin and evolution of *Citrus*. *Nature* **554**, 311–316 (2018).
35. Nowak, M. D., Davis, A. P., Anthony, F. & Yoder, A. D. Expression and trans-specific polymorphism of self-incompatibility RNases in *Coffea* (Rubiaceae). *PLoS ONE* **6**, e21019 (2011).
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.* **4**, 237–249 (2011).
37. Igić, B., Lande, R. & Kohn, J. R. Loss of self-incompatibility and its evolutionary consequences. *Int. J. Plant Sci.* **169**, 93–104 (2008).
38. Li, M. et al. Genome structure and evolution of *Antirrhinum majus* L. *Nat. Plants* **5**, 174–183 (2019).
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.* **66**, 389–400 (2008).
40. Golz, J. F., Clarke, A. E., Newbigin, E. & Anderson, M. A relic S-RNase is expressed in the styles of self-compatible *Nicotiana sylvestris*. *Plant J.* **16**, 591–599 (1998).
41. Liu, Y., Liu, Q., Tao, N. & Deng, X. Efficient isolation of RNA from fruit peel and pulp of ripening navel orange (*Citrus sinensis Osbeck*). *J. Huazhong Agr. Univ.* **25**, 300–304 (2006).
42. Grabherr, M. G. et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **29**, 644–654 (2011).
43. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
44. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
45. Eddy, S. R. A new generation of homology search tools based on probabilistic inference. *Genome Inform.* **23**, 205–211 (2009).
46. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18 (2012).
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.* **3**, 88–97 (1990).
48. Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
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. Preprint at https://www.biorxiv.org/content/10.1101/289660v2 (2018).
50. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
51. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
52. Trapnell, C. et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
53. Workenhe, S. T., Kibenge, M. J. T., Iwamoto, T. & Kibenge, F. S. B. Absolute quantitation of infectious salmon anaemia virus using different real-time reverse transcription PCR chemistries. *J. Virol. Methods* **154**, 128–134 (2008).
54. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
55. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
56. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).

57. Hedges, S. B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.* **32**, 835–845 (2015).
58. Yen, Y. & Green, P. J. Identification and properties of the major ribonucleases of *Arabidopsis thaliana*. *Plant Physiol.* **97**, 1487–1493 (1991).
59. Besbes, F., Franz-Oberdorf, K. & Schwab, W. Phosphorylation-dependent ribonuclease activity of Fra a 1 proteins. *J. Plant Physiol.* **233**, 1–11 (2019).
60. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
61. The Angiosperm Phylogeny Group. An update of the angiosperm phylogeny group classifcation for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* **161**, 105–121 (2009).

Acknowledgements

We are grateful to J. Wu from Nanjing Agricultural University for providing the sample of *Pyrus bretschneideri*. This research was financially supported by the National Key Research and Development Programme of China (grant no. 2018YFD1000107), the National Natural Science Foundation of China (grant nos. 31772259, 31630065 and 31521092), the Fundamental Research Funds for the Central Universities (grant no. 2662019PY044) and the China Agriculture Research System (grant no. CARS-27). The Biotechnology and Biological Sciences Research Council (BBSRC) funds research in the laboratories of M.B. and V.E.F.-T. (grant no. BB/P005489/1). We would like to thank T. Li (China Agricultural University) and C. Franklin (School of Biosciences, University of Birmingham, UK) for their valuable suggestions.

Author contributions

L.C., M.L. and V.E.F.-T. designed the experiments. M.L., Z.C., H.Y., Q.X.Jr and M.T. performed the experiments. J.L. completed the collection and sequencing of sour orange. M.L., A.Z., Y. Liu, Y. Li and R.X. analysed the bioinformatics data. S.W., C.C., Z.X. and C.D. collected the pummelo accessions. J.Y., W.G., Q.X., R.M.L., X.D., M.B. and L.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.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41477-020-0597-3.

Correspondence and requests for materials should be addressed to L.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

nature research

Corresponding author(s): Mei Liang, Lijun Chai

Last updated by author(s): Nov 13, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software/custom algorithm was used to collect data in this study.
Data analysis	The following publicly available software were used: Trinity v2.8.4；MEGA7; HMMER v3.2.1;SOAP denovo((http://soap.genomics.org.cn/soapdenovo.html); Circos v0.69; I-TASSER v5.1; PyMol v2.2.2; TopHat2; TBtools v0.6654 ; MAFFT v7.402; RAxML v8.2.11; DnaSP v5; Bowtie2 v2.3.4.3; Bedtools v2.25.0; GATK -v4.0.3.0; VCFtools v0.1.13; ClustalW v2.1; GeneDoc v2.7.000; Jalview 2.10.5; Sigmaplot v14; GENETYX v14; PyMol v2.2.2; CorelDRAW x8 v18.0.0.448; Adobe Photoshop CS6 v13.0.1; Adobe Illustrator v23.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

For each instance: the RNA-seq data (in Suppl. Table 3), the DNA-seq data (in Suppl. Tables 7 and 11) and the sequence data (in Suppl. Table 4). These data have been deposited in the NCBI Bioproject ID under accession codes PRJNA526584, PRJNA573625, PRJNA573817, PRJNA573818, PRJNA544805, PRJNA544816, PRJNA544866, PRJNA544867, PRJNA573624 and the NCBI Genbank ID under accession codes MN652897, MN652898, MN652899, MN652900, MN652901, MN652902, MN652903, MN652904, MN652905, MN652906, MN652907, MN652908, MN652909, MN652910, MN652911, MN652912.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was chosen to be sufficient based on our past experience or the previous references performing similar analyses. 1. The 64 RNA-seq data were used for the identification of S-RNase. 2. The 391 pummelo accessions were used to investigate the frequency of the S-RNase genes. 3. Approximately 600 pollinations were carried out for assessing S-genotypes. 4. The length of at least 150 pollen tubes were measured in the in vitro pollen bioassay. 5. 157 DNA-seq of citrus accessions were used for the S-RNase mapping.
Data exclusions	No data were excluded during our analyses. All materials generated during this study are available from the corresponding author upon request.
Replication	All of the replications following below obtained same or similar results. 1. qRT-PCR for S-RNase and SLF was carried out two times and each time had three biological replicates. 2. Western blot for developmental expression of pistil S-RNase was carried out at least 5 times to be certain. 3. RNase in gel assay was carried out at least 5 times. 4. The in vitro pollen bioassay was carried at least three biological replicates. 5. The RNA-seq data for each citrus accession was two biological replicates.
Randomization	Experiments did not require randomization of samples. Replicates of each pollen bioassay were carried out independently at different times.
Blinding	The length measurement of the pollen tube in the in vitro assay were performed with a manner blinder to treatment.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access and import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Anti-S1-RNase antibody and Anti-S2-RNase antibody were produced by the Frdbio company in Wuhan, China.
Anti-S1-RNase antibody (1:2,000 dilution), raised against the S1-RNase-GST fusion protein in rabbit
Anti-S2-RNase antibody (1:1,000 dilution), raised against the S2-RNase-GST fusion protein in rabbit

and detected using Goat anti-rabbit IgG antibody (H&L)[HRP], pAb (Company: GenScript; Catalog number: A00098; clone name Dilution: 1:5,000)

Validation

The anti-S1/S2-RNase antibody was validated using GST-tagged S1/S2-RNase proteins using Western blotting in the current paper.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

State the source of each cell line used.

Authentication

Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.

Mycoplasma contamination

Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Palaeontology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.

Wild animals

Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a*

Software | *community repository, provide accession details.*

Cell population abundance | *Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*

Gating strategy | *Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type | *Indicate task or resting state; event-related or block design.*

Design specifications | *Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*

Behavioral performance measures | *State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s) | *Specify: functional, structural, diffusion, perfusion.*

Field strength | *Specify in Tesla*

Sequence & imaging parameters | *Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*

Area of acquisition | *State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software | *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization | *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template | *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal | *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring | *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings | *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested | *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016) | *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction | *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.