Gene duplication and genetic exchange drive the evolution of S-RNase based self-incompatibility

Ken-ichi Kubo[1], Timothy Paape[2], Masaomi Hatakeyama[2,3], Tetsuyuki Entani[1], Akie Takara[1], Kie Kajihara[1], Mai Tsukahara[1], Rie Shimizu-Inatsugi[2], Kentaro K. Shimizu[2] & Seiji Takayama[1]

[1]Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma 630-0192, Japan.
[2]Institute of Evolutionary Biology and Environmental Studies, University of Zurich, CH-8057 Zurich, Switzerland.
[3]Functional Genomics Center Zurich, CH-8057 Zurich, Switzerland.

Self-incompatibility (SI) systems in flowering plants distinguish self and non-self pollen to prevent inbreeding. While all other SI systems studied to date rely on the self-recognition between each single male- and female-determinants, the Solanaceae plants has a non-self recognition system that functions through the detoxification of non-self female-determinants of S-ribonucleases (S-RNases), expressed in pistils, by multiple male-determinants of *S*-locus F-box proteins (SLFs), expressed in pollen. However, little is known about how many SLF components constitute such a non-self recognition system and how they evolve. Here we conducted large-scale next-generation sequencing and genomic PCR and identified 16–20 *SLF*s in each *S*-haplotype in SI *Petunia*, for a total of 168 *SLF* sequences. We predicted the target S-RNases of SLFs by assuming that a particular *S*-allele must not have a conserved SLF that recognizes its own S-RNase, and validated them by transformation experiments. A simple mathematical model showed that 16–20 *SLF* sequences would be adequate to recognize the vast majority of target S-RNases. We found evidence of gene conversion events, which we suggest are essential to constitute a non-self recognition system and as well as contributed to self-compatible mutations.

SI is a genetically controlled reproductive barrier in angiosperms that allows the pistil to reject self (genetically-related) pollen and accept non-self (genetically-unrelated) pollen[1-4]. In most cases, this self/non-self discrimination is controlled by male- and

female-specificity determinants (pollen-*S*, pistil-*S*) encoded by multi-allelic genes at the *S*-locus. Because *pollen-S* and *pistil-S* are tightly linked to each other at the *S*-locus, combinations of these alleles are considered *S*-haplotypes.

Two main types of SI system exist: self recognition and non-self recognition[3]. Although SI species in Brassicaceae and in Papaveraceae differ in their determinant proteins, both possess self recognition that relies on the interactions between highly polymorphic molecules, a ligand and a kind of receptor, derived from a single *S*-haplotype[1,2]. In such a system, suppression of recombination between *pollen-S* and *pistil-S* results in corresponding shapes of phylogenetic trees of alleles (often called co-evolutionary relationships in a narrow sense)[5].

Self-incompatibility in Solanaceae, Plantaginaceae, and Maloideae of Rosaceae are non-self recognition systems[3]. In these families, pistil-*S* is a secreted ribonuclease termed S-RNase, which exerts cytotoxic effects that inhibit the elongation of self-pollen tubes by degrading RNA[1-4]; consequently, the SI system in these families is referred to as S-RNase–based SI. The pollen-*S* is a set of F-box protein(s), termed *S*-locus F-box (SLF[1-4], also called *S*-haplotype-specific F-box, SFB[6], or *S*-haplotype-specific F-box brothers, SFBB[7] in Rosaceae), and function as a component of the SCF (Skp1–Cullin1–F-box)–type E3 ubiquitin ligase which generally mediates ubiquitination of target proteins for degradation by the 26*S* proteasome[8]. Previously, we proposed that S-RNase–based SI in Solanaceae is a collaborative non-self recognition system, in which the product of each SLF interacts with a subset of non-self S-RNases, and the products of multiple SLF types are required for the entire suite of non-self S-RNases to be collectively recognized and detoxified[9].

In contrast to the co-evolutionary relationships observed in the specificity determinants in Brassicaceae and Papaveraceae, *S-RNase* and *SLF*s derived from Solanaceae and Maloideae exhibited no corresponding allele phylogenies. One possible explanation for this observation is that *S-RNase*s and *SLF*s each proliferate by different mechanisms, giving them the appearance of different evolutionary histories despite tight linkage and co-inheritance as a single haplotype. Increasing the repertoire of *SLF* genes that constitute *pollen-S* would be advantageous, as this would increase the number of potential mating partners by allowing pollen to recognize and detoxify more non-self S-RNases, whereas an increase in diversity of the *S-RNase* genes would have the opposite effect by allowing new S-RNases to escape detoxification by the existing

repertoire of SLF proteins[9]. This pattern more resembles disease recognition and pathogen evasion models than other SI systems[10]. Such factors may have caused the differences in the evolutionary diversification of these genes, but the underlying details remain unclear. In order to characterize the evolutionary history of the entire non-self recognition locus, we conducted a large-scale identification of *SLF*s from many *S*-haplotypes.

RESULTS

Petunia pollen-S consists of approximately 18 *SLF* types

We identified *SLF* genes from eight SI haplotypes (S_5, S_7, S_9, S_{10}, S_{11}, S_{17}, S_{19} and S_{22}) and two self-compatible (SC) haplotypes (S_m and S_{0m}; see Online Methods) using a combination of next-generation sequencing (NGS) and PCR techniques. Initially, we constructed expressed sequence tag (EST) libraries from male reproductive organs of lines homozygous for each *S*-haplotype except S_{10}, S_{22} and S_m, and then identified *SLF*-related sequences from these EST libraries. Next, we conducted RT-PCR and RACE-PCR to fill gaps and obtain full-length sequences. We then cloned whole coding sequences of novel candidate *SLF*s by genomic PCR to confirm the absence of assembly errors in all identified (by Sanger sequencing). PCR reactions were applied to all lines including those homozygous for S_{10}, S_{22} and S_m-haplotypes to identify additional and undetected *SLF*s through NGS. The expression of all identified *SLF*s in anther was confirmed by semi-quantitative RT-PCR. Finally, we identified 16 (in S_{11}-haplotype) to 20 (in S_{19}-haplotype) *SLF*-related sequences per haplotype, for a total of 168 sequences (180 sequences including 12 pseudogenes were listed in Supplementary Table 1). Based on their phylogenetic relationships, we classified them into 18 types (named Type 1 to 18 *SLF*s). When the sequences from more than three SI *S*-haplotypes were collapsed into one clade, we classified them as a novel type. Ungrouped sequences that belonged to none of these 18 types were tentatively named *FBX*s (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1).

Previous analyses demonstrated genetic linkage among the already-known alleles of type 1–6 *SLF*s and their cognate *S-RNase*s[9]. We analyzed the linkage of newly isolated *SLF*s using gene-specific primer pairs (Supplementary Table 2). Examination of 48 plants with segregating *S*-haplotypes revealed no recombination between *SLF*s and their

cognate *S-RNase*s (Supplementary Fig. 2). We also confirmed male reproductive organ–specific expression profiles of newly isolated *SLF*s and *FBX*s using RT-PCR (Supplementary Fig. 3). These results indicated that we have identified strong candidates for novel *pollen-S* components.

Diversity and deletion of *SLF*s predict target S-RNase

Variations among allelic sequences within each type of *SLF*s can be classified into two types of polymorphisms: copy number variation and amino-acid sequence polymorphism. As for copy-number variations, 0–2 genes of each type of *SLF* were identified in each *S*-haplotype: for example, no type-9 *SLF* was identified in the S_{19}-haplotype, whereas two copies of type-1 *SLF*s were detected in the S_7-, S_{17}-, and S_{19}-haplotypes (Fig. 2). Regarding amino-acid sequence polymorphism, we observed alleles with high sequence conservation as well as those with relatively moderate sequence conservation. For example, in type 3, seven alleles have high sequence conservation (99.4–87.4% identities), while two alleles (*S_7-SLF3*, *S_{11}-SLF3B*) have moderate sequence conservation (76.5–72.0%) (Fig. 2, Supplementary Table 3).

Assuming the collaborative non-self recognition model[9], a functional S_x-haplotype must not encode a SLF that recognizes and detoxifies its own S_x-RNase. This can be achieved by having either a diverged or deleted allele of the SLF type that recognizes the S_x-RNase. This logic predicts that, if the S_x-haplotype encodes no highly conserved allele of type-n *SLF* (*SLFn*), the conserved SLFn would recognize the "non-self" S_x-RNase.

As for type-3 *SLF*s, S_7-haplotype has only one relatively diverged allele (*S_7-SLF3*) (Fig. 2a). S_{11}-haplotype also has such relatively diverged allele (*S_{11}-SLF3B*), but it also has conserved one (*S_{11}-SLF3*). Therefore, the model predicts that S_7-RNase is the target of the conserved SLF3. Indeed, our previous transgenic experiments showed that S_{11}-SLF3 targeted S_7-RNase[9]. In our current study, we further verified interactions by a similar transgenic approach showing that another conserved SLF3, S_5-SLF3, recognizes S_7-RNase (for details of transgenic experiments to assay S-RNase–SLF interaction *in vivo*, see Fig. 3, Supplementary Fig. 5, Supplementary Tables 4 and 5). Furthermore, we confirmed that S_7-SLF3 and S_{11}-SLF3B did not recognize S_7-RNase (Supplementary Fig. 5, Supplementary Table 4).

As the case of deletion, we focused on the type-9 *SLF* clade, where S_{10}- and

S_{19}-SLF9 was absent (Fig. 2b). The model predicts that S_{10}- and S_{19}-RNase could be the target of conserved type-9 SLFs. We tested the S_{19}-RNase and found that it was indeed targeted by two alleles of type-9 SLFs that we tested (*S_9-SFL9A* and *S_{11}-SLF9*; Supplementary Fig. 6, Supplementary Tables 4 and 5). Additionally, we found that *S_9*-, *S_5*-, and *S_{10}*-haplotypes lack type-2, type-14, and type-6 *SLFs*, respectively. Among these, predicted interaction between conserved type-2 SLFs and S_9-RNase consisted with our previous results[9].

We should emphasize that the predictive method does not exclude the possibility that conserved *SLF* alleles, can act on additional target S-RNases. In the type-1 *SLF* clade, six out of eight *S*-haplotypes had a highly conserved *SLF* allele, whereas *S_{17}*- and *S_{22}*-haplotypes had only relatively diverged *SLF1*s (Fig. 2c). The model predicts that both S_{17}-RNase and S_{22}-RNase are the targets of conserved SLF1. We previously tested four conserved *SLF1* alleles (*S_5*, *S_7*, *S_9*, *S_{11}*) and showed that all of them targeted S_{17}-RNase[9]. Our new experiment confirmed that S_{22}-RNase is also the target of a conserved SLF1, S_7-SLF1 (Supplementary Fig. 7, Supplementary Tables 4 and 5). In addition, our previous experiments showed that two of the conserved SLF1s (*S_5* and *S_7*) targeted S_9-RNase, whereas two others (*S_9* and *S_{11}*) did not, implying that an evolutionary change in specificity could also occur with very limited amino acid substitutions. Such a change is consistent with maintenance of normal SI function in *S_9*-haplotype.

Collectively, we found the following pattern: S_x-RNase is a target of SLFn if the *S_x*-allele of *SLFn* is diverged or deleted. This predictive approach is very useful for identifying the target S-RNase(s) of each type of SLF. Actually, among eight SI *S*-haplotypes analyzed in this study, we could predict the responsible SLF types for seven S-RNases, and five (S_7, S_9, S_{17}, S_{19}, and S_{22}-RNases) among them are actually shown to interact with the predicted SLFs with experimental evidence. Because there are more than 40 *S*-haplotypes in *Petunia*[4], it is not surprising that a conserved *SLF* was found in all of the eight surveyed *S*-haplotypes in the majority of the *SLF* types (types 4, 5, 10, 11, 13 and 16, Supplementary Fig. 1), and are likely to target S-RNases of other unsurveyed *S*-haplotypes. In comparisons of *SLF* sequences belonging to the same types, most of the allelic sequences were shown to be highly conserved with identities higher than 90% (Supplementary Table 3). The results of the *in vivo* assay described above suggested that most of these highly conserved *SLF*s function as *pollen-S*. This is in stark

contrast to other self-recognition SI systems, such as *Brassica* and *Papaver*, where self/non-self discrimination depends on *S*-haplotype-specific interactions between highly diverged pollen-*S* and pistil-*S*[1,2].

Solanum S-loci contain orthologous *SLF*-like paralogs

It is not obvious how *SLF*s and *S-RNase*s came to constitute a genetic unit at a single *S*-locus during evolution of the S-RNase–based-SI system considering the much lower diversity among *SLF*s relative to S-RNases[11]. We conducted phylogenetic analysis of *SLF*s and *S-RNase*s including those of other genera of Solanaceae by first exploring *SLF*-orthologs in the whole-genome databases of tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*)[12,13]. We identified 37 and 66 *SLF*-like F-box sequences in these two species (Online Methods and Supplementary Table 6). These candidates include both *SLF* and other *SLF*-like sequences. Phylogenetic trees including sequences we identified as well as published *pollen-S*–related sequences in other S-RNase–based SI species (Fig. 4a, Supplementary Fig. 8 and Supplementary Table 7) show the *Petunia SLF*s cluster into a single monophyletic group together with 13 genes from tomato and 14 from potato (this subclade is referred as the Solanaceae *SLF* clade; Fig. 4a, Supplementary Fig 8). All of the tomato and potato genes belonging to the Solanaceae *SLF* clade are specifically located within the repeat-rich, subcentromeric regions in chromosome 1 of the assembled *S. lycopersicum* and *S. tuberosum* genome, consistent with the genetically mapped locations of the *S*-loci in these species (Fig. 4c)[14,15]. These Solanaceae *SLF*s are distributed within 17.9 Mb in tomato and 14.6 Mb in potato, suggesting that the *S*-loci of the Solanaceae are very large, about two to three orders of magnitude larger than the Brassicaceae *S*-locus (28–110 kb)[16].

*Allelic SLF*s are much younger than *S-RNase* alleles and than *SLF* types

The *SLF* types in *Petunia* and other *SLF*s derived from different genera in Solanaceae are distributed throughout the Solanaceae *SLF* clade (Fig. 4a and Supplementary Fig. 8c), suggesting that major diversification of the types predated separation of genera in Solanaceae. This is consistent with the extensive trans-specific polymorphism found in Solanaceae *S-RNase*s[17,18] (Fig. 4b, Supplementary Fig. 9 and Supplementary Table 8). However, there are small subclusters specifically derived from certain genera, *e.g.*, the type-3/-13, type-4/-12 and type-9/-10 clusters, suggesting that generation of new *SLF*

types might have continued after separation of genera. The branch depth of S-RNases and SLFs suggest that the timing of proliferation of *SLF* types rather than individual *SLF* alleles (Figs. 1a and 3b and Supplementary Fig. 8b) is similar to that of the *S-RNase* alleles (Figs. 1b and 3b and Supplementary Fig. 9b). Allelic *SLF*s belonging to each type diversify only at terminal branches of the tree and there is no pair of closely related *SLF* sequences between different genera of Solanaceae, indicating diversification of allelic *SLF*s within each type followed the divergence of genera. These results suggest that each genus has a similar number of *SLF* types but there is no one-to-one correspondence of *SLF* types among these genera possibly due to evolutionary turnover of *SLF* types.

Estimates of synonymous and non-synonymous substitution rates (K_s and K_a, respectively) between *SLF*s and *S-RNase*s showed that inter-allelic K_a and K_s values of each *SLF* type (K_a = 0.000–0.090; K_s = 0.001–0.303) were much lower than the values for the *S-RNases* (K_a = 0.400; K_s = 0.850) in *Petunia* (Supplementary Table 9). However, intra-haplotypic K_a and K_s values of *SLF*s exhibited ranges similar to those of the *S-RNase*s (K_a = 0.321–0.349; K_s = 0.747–0.762). These values were similar in *Solanum*. These results indicate that alleles of each *SLF* type are much younger than the *S-RNase*s alleles and than the *SLF* types.

Genetic exchanges of *SLF*s have occurred repeatedly

Our findings seem to conflict with completely suppressed recombination among *SLF*s and *S-RNase*s thought necessary to maintain all SI systems. While it is clear that linkage at the *S*-locus is generally necessary to maintain individual haplotypes over large genomic regions (*e.g.* 15 Mb), we suspect that sharing of *SLF*s among *S*-haplotypes through genetic exchange has occurred repeatedly, at least until relatively recently. Supporting this speculation, some sets of genes share complete identity among several alleles; *e.g.* the S_7- and S_{19}-alleles of type-1 *SLF*s share completely identical nucleotide sequences whereas the alleles of the corresponding *S-RNase*s are quite different (47.5% amino-acid identity; Supplementary Table 3)[9]. These findings support genetic exchange among *SLF*s and the extremely low level of polymorphism among *SLF* alleles cannot be explained solely by purifying selection on amino-acid replacement as both K_s and K_a are low.

In order to detect a statistical signal of genetic exchange among *SLF*s, we used

LDhat[19] and GENECONV[20]. Genetic exchange was estimated on alignments that contained *SLF*s from self-compatible (SC) mutant haplotypes and again on alignments with these sequences removed. Both approaches detected genetic exchange among alleles in type-3, -9 and -10 *SLF*s (Supplementary Tables 10 and 11). When *SLF*s from SC haplotypes were included, we found many more cases of significant pairwise exchanges between *SLF* alleles, most strongly within type-9 *SLF*s, as well as among type-9 and -10 *SLF*s, which are closely related sister groups (Supplementary Tables 11). Several *SLF* types and pairs of *SLF*s within types exhibit significant influences from genetic exchange. In total we found significant exchange among 36 pairs of *SLF*s. This is a conservative estimate of genetic exchange at the *S*-locus because the described approach focused on recombination breakpoints within the coding sequence of a particular *SLF* type but cannot detect those in the intergenic regions that would result in the exchange of entire *SLF*(s). Overall, these results indicate that genetic exchange might play a role in conservation of *SLF* function.

Gene conversion contributed to evolution of SC haplotypes

In the above analysis, we detected gene conversion in self-compatible S_{0m}- and S_m-haplotypes more prevalently than self-incompatible haplotypes (Supplementary Table 11). To investigate the relationship between breakdown of SI and recombination among *SLF* genes, we compared *SLF*s between the pollen-side SC haplotype and its ancestral SI haplotype.

S_{C2}-haplotype is a pollen-side SC mutant of the S_{17}-haplotype derived from an SC/SI-mixed natural population of *Petunia axillaris*[21] (Supplementary Fig. 10a). Our data showed that the S_{C2}-haplotype shares a *S-RNase* and *SLF*s with the S_{17}-haplotype, but also contains an additional *SLF1* (S_{C2}-*SLF1C*) identical to the "conserved"-type S_7-/S_{19}-*SLF1* (Supplementary Figs. 11a, b), suggesting that this duplication should be the reason for the breakdown of SI. In addition to S_{C2}-haplotype, we newly identified an additional pollen-side SC mutant of S_{22}-haplotype, designated S_{22m}-haplotype (Supplementary Fig. 10b). S_{22m}-haplotype shares a *S-RNase* with S_{22}-haplotype, but also contains an additional *SLF1* (S_{22m}-*SLF1B*) yet again identical to S_7-/S_{19}-/S_{C2}-*SLF1* (Supplementary Figs. 11a, b). Our previous transgenic experiments[9] and newly performed ones (Fig. 2 and Supplementary Fig. 5) indicated that the presence of this common additional *SLF1* was sufficient for the breakdown of SI in S_{C2}- and S_{22m}-pollen.

This suggests that the genetic exchange was responsible for the evolution of self-compatibility (see discussion). The existence of a shared *SLF1* among four different haplotypes further represents evidence of recent inter-haplotypic *SLF*-gene exchange.

Interestingly, in addition to type-1 *SLF*, we found that these four *S*-haplotypes also share one common type-8 *SLF* (*S_7-/S_{19}-/S_{C2}/S_{22m}-SLF8*) (Supplementary Fig. 11c and 12). This allowed us to compare the phylogenies between type-1 and type-8 *SLF*s, and we found that these types showed similar topologies (Supplementary Fig. 13). This result suggests that the *SLF1–SLF8* linkage unit might have been transferred among different *S*-haplotypes over evolutionary time. We conclude that genetic exchange among *S*-haplotypes and some linkage units has occurred repeatedly and contributes to the evolution of both SC and SI *S*-haplotypes.

Mathematical models suggest that 16-20 SLFs would be adequate for non-self recognition

The number of *SLF* types is much less than that of predicted *S-RNase* alleles (40 or more)[4], thus one-to-one interactions between a SLF type and a S-RNase allele is not possible. Rather, some SLF types should interact with multiple S-RNase allelic variants, while some S-RNases can be recognized by multiple SLF types. To estimate whether the 16–20 *SLF* genes we identified here would be adequate for non-self recognition, we compiled the empirical data of the SLF and S-RNase interactions[9,22,23] including data presented in this study (Supplementary Table 12) and developed simple models. Among the 129 tested combinations of SLF and S-RNase, 24 showed positive interactions, thus a SLF would recognize 18.6% S-RNases on average. If we pose a simple assumption that target S-RNase repertories for each SLF are independent and there are *n* SLFs, the proportion of $(1\text{-}0.186)^n$ S-RNases cannot be recognized. Thus, *n* SLFs can recognize $1\text{-}(1\text{-}0.186)^n$ proportion of S-RNase alleles. With *n* = 16–20, the probability of recognition approaches saturation (Fig. 5a), which would be sufficient for this self-incompatible system to work. Next we relaxed the assumption so that each SLF can recognize a different proportion of target S-RNases. Because experimental data are already available from 50% (9/18) of the SLF types (see Online Methods), we used Monte-Carlo simulation with bootstrapping for these data. Again, we found similarly that the interactions become saturated by 16–20 SLF types (Fig. 5b). The models suggest that the previously identified eight SLF types may not constitute an efficient

non-self recognition system, but that 16–20 SLFs on each haplotype would be adequate to recognize the vast majority of S-RNase targets if not all, which is estimated to be about 40 alleles in *Petunia*[4]. The results also support the validity of the number of *SLF*s identified in *Petunia.* We note that the recognition rate from these models may be considered as minimum estimates, because different SLF types may tend to recognize different S-RNases since overlapping targets may not be favored by selection. The upper limit of the number of *SLF* types should be constrained by factors such as the strength of inbreeding depression and the proportion of self-pollen deposited on a stigma in natural population, birth-and-death rate of *SLF* types and effective population size[24-26]. These simple models suggest that we have identified the majority of the genetic components of this non-self recognition system.

DISCUSSION

Co-evolving genes for self/non-self recognition systems in are notoriously difficult to study because they typically involve inter-organism dynamics such as disease resistance and virulence and either involve quantitative phenotypes (with epistatic or pleiotropic effects) or are lethal to one or both organisms. Plant pathogen recognition is governed by hundreds of duplicated *R*-genes that detect numerous pathogens[10]. Autoimmune disease phenotypes by self-recognition may be observed as Dobzhansky-Muller incompatibility in hybrids, because an R-protein from a parent often recognizes self-protein derived from another parent[27,28]. Solanaceae SI genetic systems provide a unique opportunity to study and model the evolutionary dynamics of co-evolving genes that largely resemble disease recognition and detoxification mechanisms, where duplicated SLF proteins need to recognize diverse non-self S-RNases but not the self S-RNase.

Our exhaustive search for new SLF genes indicated that 16 to 20 *SLF*s were present in each of ten haplotypes of *Petunia*. Although only eight *SLF* types were reported previously[9,22,23], phylogenetic analysis showed that these could be classified into 18 major SLF types with occasional allelic duplication or deletion in each type. Our mathematical models (Fig. 5) suggests that 16–20 SLFs on each haplotype would be adequate to recognize the vast majority of about 40 S-RNase targets[4].

Our phylogenetic observations suggested that the origins of the *SLF* types are as ancient as the origin of alleles of *S-RNase*s, whereas alleles of each type of *SLF*s are

much younger than the *S-RNase*s. This is in clear contrast to self-recognition SI systems, in which the *pollen-S* and *pistil-S* show distinct co-evolution in nucleotide substitutions[5]. The generation of new *SLF* types via inter-haplotype genetic exchange provides an explanation for the phylogenetic pattern of *SLF*s in the collaborative non-self recognition system. Three possible consequences are postulated in Fig. 6a.

In the first scenario, a new *SLF* acquired by genetic exchange could recognize more effectively wider range of 'non-self' S-RNases. In such a case, the *SLF* acquisition could confer an advantage for outcrossing, and would be fixed in the population (evolution from S_{1a}- to S_1-haplotypes in Fig. 6a). This *SLF* would spread rapidly over multiple *S*-haplotypes in the population and form a new *SLF* type with short branches in the phylogenetic tree (S_1-*SLFn* in Fig. 6b).

In the second scenario, newly acquired SLF(s) could recognize an endogenous 'self' S-RNase as a specific target, inducing breakdown of SI (evolution from S_{2a}- to S_{2m}-haplotypes in Fig. 6a). Most such SC *S*-haplotypes should be lost by short term disadvantage (inbreeding depression), and show a short branch in a phylogenetic tree (S_{2m}-*SLFn* in Fig. 6b). Occasionally, self-compatible alleles are found in natural SC/SI-mixed populations of *Petunia*[29], and could be fixed by selective forces such as mate limitation or automatic transmission advantage and by escaping rejection from all *S*-haplotypes in outcrossing[24,26,30-33]. Although it has long hypothesized that recombination and/or gene conversion between different *S*-alleles could induce self-compatibility, no clear example has been found. In natural and domesticated populations of Brassicaceae possessing a one-to-one self-recognition system, loss-of-function mutations of pollen-*S*, pistil-*S* or modifiers, rather than recombination, have been shown to be responsible for self-compatibility[34-36]. Here we provide experimental evidence that the acquisition of a *SLF1* by gene conversion has lead to the evolution of two SC haplotypes (S_{C2} and S_{22m}). The self-compatibility in pollen-*S* is also consistent with the theory suggesting that mutations in male components are more likely to spread in natural populations[26,33].

In the third scenario, some SC haplotypes generated in the second scenario could restore SI phenotype by having mutations in the responsible *SLF* or its target *S-RNase* (evolution from S_{3a}- to S_3- or S_4-haplotypes in Fig. 6a). Accumulation of mutations in the acquired *SLF* could regenerate an SI *S*-haplotype. As we used in the prediction of target S-RNases, a particular SI *S*-haplotype must not have the SLF that recognizes its

own S-RNase, thus the *SLF* allele should be lost or diverged, exhibiting a relatively long branch in phylogenetic tree (*S_3-SLFn* in Fig. 6b). Mutations in the target *S-RNase* could also regenerate a novel SI *S*-haplotype. Such an S-RNase mutation that escapes from the recognition by an SLF should have a risk leading to the female sterility. In the collaborative non-self recognition system, the redundant S-RNase recognition by multiple SLFs could reduce the risk and might support the evolution of new *S*-haplotypes (S_4-RNase in Fig. 6b).

In Solanaceae, the *S*-locus is located in a subcentromeric, repeat-rich, and low gene-density region, in which recombination is strongly suppressed[37,38]. Until recently, it has remained unclear how *SLF*s evolved despite being located in such an inactive genomic region. However, recently crossover recombination (*i.e.*, homologous recombination) was shown to be fully suppressed at centromeres, while non-crossover recombination (*i.e.*, gene-conversion) is not[39], therefore, genetic exchange among *SLFs* could also be possible. Other studies suggested that intragenic recombination has also contributed to the diversity of *S-RNase*s[40,41], although the frequency seemed to be rare. Altenatively, RNA-mediated genetic exchange (retrotransposition)[42] may have also contributed because the *Petunia S*-locus is rich in retrotransposons[43] and no *SLF* gene thus far identified contains introns. However, our finding of linkage unit *SLF1-SLF8* in several *S*-haplotypes suggests the former event is more likely based on the distribution of this linkage unit.

Our comprehensive identification of *SLF*s will provide useful data for characterizing the molecular evolution of S-RNase-based SI, a topic that has been debated for many years. In contrast to collaborative non-self recognition model, traditional models of SI assumed self-recognition by a single gene product[26,44,45]. The findings described here indicate that *pollen-S* and *pistil-S* have undergone different and complex patterns of molecular evolution and apparently depend on the genetic exchange of *SLF*s as well as sequential accumulation of base substitution. In the future, it should be possible to explain the molecular evolution of *S*-specificity in populations using simulations based on the molecular model we proposed here.

METHODS

Plant materials. We used lines of *S_5*-, *S_7*-, *S_9*- *S_{10}*-, *S_{11}*-, *S_{22}*-, *S_{0m}*-, *S_m*-, and

S_{22m}-haplotypes from *Petunia hybrida* and the lines of S_{17}-, S_{19}- and S_{C2}-haplotypes from *P. axillaris* (for details, see Supplementary information).

EST database for male-reproductive organ of *Petunia*. For preparation of cDNA libraries and NGS, see Supplementary information.

Based on the results of next-generation sequencing, we constructed EST databases using GENETYX-MAC (ver. 16.0.6).

Identification of novel *SLF* genes. Local BLAST (NCBI BLAST ver. 2. 2. 24) search against EST databases was carried out with GENETYX-MAC, using *Petunia* type 1–6 *SLF*s[9] as queries. We extracted hits with evaluation (*E*) values less than 10^{-20}, and assembled them using ATSQ (ver. 5.1.3, GENETYX) with the following parameters: matching percentage, 50%; minimum matching number, 10; and group capacity, 5000. All assembled sequences were manually checked. To isolate additional *SLF*s from *S*-haplotypes unanalyzed by NGS and to confirm the full length coding sequences, we performed RACE and genomic PCR. Primers are listed in supplementary table 2 and reaction conditions are described previously[9]. As for newly isolated *SLF*s, to confirm the absence of assembly errors, entire coding regions were amplified by genomic PCRs and cloned into the pGEM-T easy vector (Promega). At least eight independent clones for each *SLF* fragment were sequenced on both strands by Sanger's method using an ABI 3130xl Genetic Analyzer (Applied Biosystems).

Identification of *S*-related genes in *Solanum*. BLAST-searches for *SLF*s and *S-RNase*s in potato and tomato were carried out at Spud DB[12] and at Sol Genomic Network[13] using databases 'PGSC DM1-3 pseudomolecules (v4.03)' and 'ITAG annotation Release 2.4 predicted CDS (SL2.50)', respectively. *Petunia SLF*s and *S-RNase* from S_7-haplotype were used as queries. Hits with *E*-values < 10^{-20} were extracted as candidates.

Preliminary analysis on tomato identified only eight *SLF*-like sequences, including four pseudogenes with premature stop codons, far fewer compared with those in potato and *Petunia*. A supposed reason was that the line used for whole genome analysis of potato is a SI line, *S. tuberosum* Phureja DM[12,46], whereas that of tomato is a SC line, *S. lycopersicum* cv Heinz 1706 (Ref. 13). Considering the possibility that the most *SLF*s

have broken down in tomato, we BLAST-searched against scaffold sequences on chromosome 1, and identified additional five unannotated *SLF*-like pseudogenes, named Solic01_pseudo1–5.

To assess the authenticity of the extracted sequences as *SLF*-related F-box genes, we evaluate the motif composition of deduced products by using Simple Modular Architecture Research Tool (SMART)[47]. Sequences judged to have neither F-box nor FBA motif were eliminated as false positives. Identified *SLF*-related and *S-RNase*-related genes are listed in Supplementary Table 6. As for potato gene name, 'PGSC0003DMG' is omitted from each potato gene ID to simplify, and '-STchx' is attached to indicate the gene location on *S. tuberosum* chromosome x in this paper (see Fig. 4).

Phylogenetic analysis. Deduced amino-acid sequences or coding DNA sequences of *F-box* and *RNase* genes were aligned with the ClustalW algorithm, using MEGA5 (ver. 2.2, Ref. 48). For pseudogenes, frame shifts or premature stops were removed manually. Based on the alignments, phylogenetic trees were constructed by the neighbor-joining method using MEGA5.

Codon-by-codon alignments of coding sequences of each type of *SLF*s or *S-RNase*s were constructed by using MEGA5. Based on these alignments, synonymous and non-synonymous substitution rates (K_s and K_a) were calculated using DnaSP (ver. 5.10.1, Ref. 49), and recombination and gene conversion were evaluated by LDhat (ver. 2.1, Ref. 19) and GENECONV (ver. 1.81, Ref. 20).

Estimation of S-RNase proportion recognized by SLFs. We applied Bernoulli process to the estimation of the proportion of S-RNases recognized by n SLF types, $P_S(n)$, assuming that each SLF recognizes S-RNase allelic variant independently with the same probability P_R as expressed by the following equation (1):

$$P_S(n) = 1 - (1 - P_R)^n \quad (1)$$

where P_R is the overall recognition rate and n is the number of SLF types. The overall recognition rate P_R is defined by the following equation (2):

$$P_R = \frac{N_P}{N_T} \quad (2)$$

where N_P is the number of positive interactions between S-RNase and SLF allelic variants and N_T is the total number of tested interactions. We excluded the positive interaction S_{22}-RNase and S_7-SLF1, because the positive interaction was predicted and S_{22}-RNase was experimentally tested only with S_7-SLF1, thus this could bias the estimation of the overall recognition rate. Remaining dataset would be considered to represent random samples of the overall recognition rate between S-RNase and SLF allelic variants. Note that $N_P = 24$, $N_T = 129$ are actually assigned in this study, so that P_R is nearly equals to 0.186 (Supplementary Table 13a). The confidential interval C_I is calculated as follows:

$$C_I = t' \frac{u}{\sqrt{N_T}} \qquad (3)$$

$$u = \sqrt{\frac{\sum_{i=1}^{N_T} (p_i - P_R)^2}{N_T - 1}}$$

where $p_i \in \{0,1\}$, 0 and 1 indicates the negative and positive interaction, respectively, in the combination of each S-RNase and SLF allelic variant interaction. $t = 1.97867$ is in this calculation assigned as the t value of the Student's t inverse cumulative distribution function at the 95th percentile of both sides of the t-distribution for the degrees of freedom N_T-1.

We also conducted Monte-Carlo simulation in order to consider the difference of recognition rates among SLF types. For details, see Supplementary information.

Accession codes. DNA sequences of newly isolated *SLF*s and *S-RNase*s have been deposited in the DNA Data Bank of Japan (DDBJ) under accessions AB932964 to AB933144. See Supplementary Table 1 and 8 for correspondence between gene names and accession nos.

Note added in proof

During revision of this article, Williams *et al*. reported the independent identification of 17 *SLF* types in two *S*-haplotypes of *Petunia inflata*[50]. Their data supported our observation indicating that 16-20 *SLF*s are sufficient for non-self recognition system in

Petunia SI.

Reference

1. Takayama, S. & Isogai, A. Self-incompatibility in plants. *Annu. Rev. Plant Biol.* **56**, 467–489 (2005).
2. Franklin-Tong, V.E. Self-Incompatibility in Flowering Plants: Evolution, Diversity, and Mechanisms (Springer, Heidelberg, 2008).
3. Iwano, M. & Takayama, S. Self/non-self discrimination in angiosperm self-incompatibility. *Curr. Opin. Plant Biol.* **15**, 78–83 (2012).
4. Sims, T.L. & Robbins, T.P. Gametophytic self-incompatibility in *Petunia*. In *Petunia*. *Evolutionary, Developmental and Physiological Genetics*. (ed. Gerats, T. & Strommer, J.) 85–106 (Springer, New York, 2009).
5. Sato, K. *et al.* Coevolution of the *S*-locus genes *SRK, SLG* and *SP11/SCR* in *Brassica oleracea* and *B. rapa. Genetics* **162**, 931–940 (2002).
6. Ushijima, K. *et al*. Structural and transcriptional analysis of self-incompatibility (*S*) locus of almond (*Prunus dulcis*): identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. *Plant Cell* **15**, 771–781 (2003).
7. Sassa, H. *et al. S locus F-box brothers*: multiple and pollen-specific F-box genes with *S* haplotype-specific polymorphisms in apple and Japanese pear. *Genetics* **175**, 1869–1881 (2007).
8. Entani, T. *et al.* Ubiquitin-proteasome-mediated degradation of S-RNase in a solanaceous cross-compatibility reaction. *Plant J* **78**, 1014-1021 (2014).
9. Kubo, K. *et al.* Collaborative non-self recognition system in S-RNase-based self-incompatibility. *Science* **330**, 796–799 (2010).
10. Bergelson, J., Kreitman, M., Stahl E.A. & Tian, D. Evolutionary dynamics of plant *R*-genes. *Science* **292**, 2281–2285 (2001).
11. Newbigin, E., Paape, T. & Kohn, J.R. RNase-based self-incompatibility: puzzled by pollen *S*. *Plant Cell* **20**, 2286–2292 (2008).
12. The Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189–195 (2011).
13. The Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641 (2012).

14. Tanksley, S.D. & Loaiza-Figueroa, F. Gametophytic self-incompatibility is controlled by a single major locus on chromosome 1 in *Lycopersicon peruvianum*. *Proc. Natl. Acad. Sci. USA* **82**, 5093–5096 (1985).
15. Gebhardt, C. *et al.* RFLP maps of potato and their alignment with the homoeologous tomato genome. *Theor. Appl. Genet.* **83**, 49–57 (1991).
16. Goubet, P.M. *et al.* Contrasted Patterns of Molecular Evolution in Dominant and Recessive Self-Incompatibility Haplotypes in *Arabidopsis*. *PLoS Genet.* **8**, e1002495 (2012).
17. Richman, A. Evolution of balanced genetic polymorphism. *Mol. Ecol.* **9**, 1953-1963 (2000).
18. Goldberg, E.E. *et al.* Species selection maintains self-incompatibility. *Science* **330**, 493–495 (2010).
19. McVean, G., Awadalla, P. & Fearnhead, P. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**, 1231–1241 (2002).
20. Sawyer, S. Statistical tests for gene conversion. *Mol. Biol. Evol.* **6**, 526–538 (1989).
21. Tsukamoto, T., Ando, T., Watanabe, H., Marchesi, E. & Kao, T.-h. Duplication of the *S*-locus F-box gene is associated with breakdown of pollen function in an *S*-haplotype identified in a natural population of self-incompatible *Petunia axillaris*. *Plant Mol. Biol.* **57**, 141–153 (2005).
22. Sijacic, P. *et al.* Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature* **429**, 302–305 (2004).
23. Williams, J.S. *et al.* Four previously identified *Petunia inflata S-locus F-box* genes are involved in pollen specificity in self-incompatibility. *Mol. Plant* **7**, 567–569 (2014).
24. Gervais, C.E., Castric, V., Ressayre, A. & Billiard, S. Origin and diversification dynamics of self-incompatibility haplotypes. *Genetics* **188**, 625-636 (2011).
25. Shimizu, K.K., Kudoh, H. & Kobayashi, M.J. Plant sexual reproduction during climate change: gene function in natura studied by ecological and evolutionary systems biology. *Ann Bot.* **108**, 777-787 (2011).
26. Uyenoyama, M.K., Zhang, Y. & Newbigin, E. On the origin of self-incompatibility haplotypes: transition through self-compatible intermediates.

Genetics **157**, 1805-1817 (2001).

27. Bomblies, K. *et al*. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol.* **5**, e236 (2007).
28. Bomblies, K. & Weigel, D. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat. Rev. Genet*. **8**, 382–393 (2007).
29. Kokubun, H. *et al.* Distribution of self-compatible and self-incompatible populations of *Petunia axillaris* (Solanaceae) outside Uruguay. *J. Plant. Res.* **119**, 419–430 (2006).
30. Darwin, C. *The effects of cross and self fertilisation in the vegetable kingdom* (J. Murray, London, 1876).
31. Fisher, R.A. Average excess and average effect of a gene substitution. *Ann. Eugen.* **11**, 53–63 (1941).
32. Busch, J.W. & Schoen, D.J. The evolution of self-incompatibility when mates are limiting. *Trends Plant Sci.* **13**,128–136 (2008).
33. Tsuchimatsu, T. & Shimizu, K.K. Effects of pollen availability and the mutation bias on the fixation of mutations disabling the male specificity of self-incompatibility. *J. Evol. Biol.* **26**, 2221–2232 (2013).
34. Boggs, N.A. *et al*. Expression of distinct self-incompatibility specificities in *Arabidopsis thaliana*. *Genetics* **182**: 1313–1321 (2009).
35. Tsuchimatsu, T. *et al*. Evolution of self-compatibility in *Arabidopsis* by a mutation in the male specificity gene. *Nature* **464**, 1342–1346 (2010).
36. Tsuchimatsu, T., Kaiser, P., Yew, C.L., Bachelier, J.B. & Shimizu, K.K. Recent loss of self-incompatibility by degradation of the male component in allotetraploid *Arabidopsis kamchatica*. *PLoS Genet.* **8**, e1002838 (2012).
37. Wang, Y., Wang, X., McCubbin, A.G. & Kao, T.-h. Genetic mapping and molecular characterization of the self-incompatibility (*S*) locus in *Petunia inflata*. *Plant Mol. Biol.* **53**, 565–580 (2003).
38. Entani, T. *et al.* Centromeric localization of an *S-RNase* gene in *Petunia hybrida* Vilm. *Theor. Appl. Genet.* **99**, 391–397 (1999).
39. Talbert, P.B. & Henikoff, S. Centromeres convert but don't cross. *PLoS Biol*. **8**, e1000326 (2010).
40. Wang X., Hughes, A.L., Tsukamoto, T., Ando, A. & Kao, T.-h. Evidence that

intragenic recombination contributes to allelic diversity of the *S-RNase* gene at the *self-incompatibility* (*S*) locus in *Petunia inflata*. *Plant Physiol.* **125**, 1012–1022 (2001).

41. Vieira, C.P., Charlesworth, D. & Vieira, J. Evidence for rare recombination at the gametophytic self-incompatibility locus. *Heredity* **91**, 262–267 (2003).
42. Kaessmann, H., Vinckenbosch, N. & Long, M. RNA-based gene duplication: mechanistic and evolutionary insight. *Nat. Rev. Genet.* **10**, 19–31 (2009).
43. Wang, Y. *et al*. Chromosome walking in the *Petunia inflata self-incompatibility* (*S*-) locus and gene identification in an 881-kb contig containing S_2*-RNase*. *Plant Mol. Biol.* **54**, 727–742 (2004).
44. Charlesworth, D. How can two-gene models of self-incompatibility generate new specificities? *Plant Cell* **12**, 309–310 (2000).
45. Uyenoyama, M.K. & Newbigin, E. Evolutionary dynamics of dual-specificity self-incompatibility alleles. *Plant Cell* **12**, 310–312 (2000).
46. Cipar, M.S., Peloquin, S.J. & Hougas, R.W. Variability in the expression of self-incompatibility in tuber-bearing diploid *Solanum* species. *Am. Potato J*. **41**, 155–162 (1964).
47. Schultz, J., Milpetz, F., Bork, P. & Ponting, C.P. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. USA* **95**, 5857–5864 (1998).
48. Tamura K. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
49. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
50. Williams, J.S., Der, J.P., de Pamphilis, C.W. & Kao, T.-h. Transcriptome analysis reveals the same 17 *S-locus F-box* genes in two haplotypes of the self-incompatibility locus of *Petunia inflata*. *Plant Cell* **26**, 2873-2888 (2014).

URLs. Sol Genomic Network, http://www.solgenomics.net/; Spud DB, http://potato.plantbiology.msu.edu/integrated_searches.shtml; SMART, http://smart.embl-heidelberg.de/.

Correspondence and requests for materials should be addressed to K.K.S. (kentaro.shimizu@ieu.uzh.ch) or S.T. (takayama@bs.naist.jp).

ACKNOWLEDGMENTS

We thank H. Takatsuji at National Institute of Agrobiological Sciences for *P. hybrida* cv. Mitchell and W138; S. Saha at Cornel University for directions on *Solanum* genomic information; M. Iwano, H. Shiba, H. Shimosato-Asano, Y. Wada, K. Murase, M. Kakita, E. Miura, H. Kakui, T. Tsuchimatsu and M. Robinson for discussion or technical advice; and F. Kodama, M. Okamura, E. Mori, Y. Goto, H. Kikuchi, and Functional Genomics Center Zurich for technical assistance. We thank T.-h. Kao at Pennsylvania State University for adapting a unified nomenclature for the same new *SLF* genes before publication. This work was supported in part by a Grant-in-Aid for Scientific Research on Innovative Areas (23113002) and by Grants-in-Aid for Scientific Research (21248014, 25252021) from the Ministry of Education, Culture, Sports, Science and Technology of Japan awarded to S.T. This work was also supported in part by Swiss National Science Foundation (31003A_140917) and by URPP Evolution in Action of University of Zurich awarded to K.K.S., Marie-Heim Vögtlin grant awarded to R.S.-I. and by Plant Fellows awarded to T.P.

AUTHOR CONTRIBUTIONS

K.-i.K., K.K.S. and S.T. planned and designed the research. K.-i.K. and T.E. constructed pollen cDNA libraries. M.H. and R.S.-I. performed next generation sequencing and construction of pollen EST databases. K.-i.K. and K.K. performed isolation and Sanger sequencing of cDNA and genomic clones. K.-i.K. and A.T. performed construction and analysis of transgenic plants. K.-i.K. and T.P. performed phylogenetic and evolutionary analyses. K.-i.K. and M.T performed linkage analysis and expression profiling. M.H. and K.K.S. designed mathematical models and performed simulations. S.T. initiated and guided the project. K.-i.K., T.P., M.H., K.K.S. and S.T. wrote the manuscript.

COMPLETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

FIGURE LEGENDS

Figure 1 Phylogenies of *SLF*s and *S-RNase*s from *Petunia*. (**a, b**) Neighbor-joining phylogenetic trees of deduced amino-acid sequences of *SLF* (a) and *S-RNase* genes (b) were created with MEGA5 (Ref. 48). Both trees are shown in the same scale; the bar for each tree indicates the number of amino-acid substitutions per site. *PpS_4-Fbox0* (a) or *PpS_4-RNase* (b) was used as an outgroup. Numbers on the branches indicate bootstrap values > 50% with 1,000 trials. To simplify, subgroups (allelic *SLF*s) within single *SLF* type are showed in a compressed representation (black triangles). The full tree of *SLF*s appears in Supplementary Fig. 1.

Figure 2 Relationships between phylogenies of *SLF*s and SLF/S-RNase interactions. (**a–c**) Neighbor-joining phylogenetic trees of deduced amino acid sequences of type-3 (a), type-9 (b) and type-1 (c) *SLF*s. For creation and explanation of phylogenetic trees, see legend of Fig. 1. Two-headed arrows indicate positive interactions between SLFs and S-RNases, which lead to pollen acceptance, demonstrated by *in vivo* assay. Gray dotted lines indicate negative interactions, which don't lead to pollen acceptance. Red and blue characters indicate relatively diverged SLFs and their cognate S-RNases, which are targeted by conserved SLFs.

Figure 3 Target S-RNase analysis of S_5-SLF3. Compatibility was judged by monitoring the pollen tubes (arrowhead)[9]. (**a**) *S_5S_9*-heterozygote with *S_5–SLF3* retained SI. Similar results were obtained for heterozygotes with *S_{11}*-, *S_{17}*-, and *S_{19}*-haplotypes (Supplementary Fig. 5), suggesting S_5-SLF3 recognizes none of these S-RNases. (**b**) Transformants *S_5S_7/S_5-SLF3* (T) exhibited breakdown of SI. (**c, d**) Reciprocal crosses with *S_5S_7* (WT) suggested that the pollen lost SI. (**e**) PCR-genotyping of F_1-progeny from WT × T. (**f**) Schematic explanation of the results in (e). Among four genotypic pollen from T, only *S_7*-pollen with *S_5-SLF3* successfully fertilized, suggesting that S_5-SLF3 detoxifies S_7-RNase. Bars = 200 μm.

Figure 4 Phylogenies and pericentromeric localization of Solanaceae *SLF*s and *S-RNase*s. (**a, b**) Phylogenies of *SLF*s (a) and *S-RNase*s (b). Symbols indicate the gene is derived from genus: *Petunia* (pink-circle), *Solunum* (blue-triangle), and *Nicotiana*

(green-square). Allelic *SLF*s collapsed into each type clades are denoted by black triangles. (**c**) Chromosome 1 of *Solunum*. Red arrows, *S*-locus regions; thick blue lines, pericentromeric regions; red lines, *SLF*s; blue line, S-RNase; black lines, F-box genes outside of *SLF* clade; ψ, pseudogenes. Potato *S-RNase* is localized to chromosome 1, but is not mapped due to a probable assembly error. For gene name abbreviation, see Methods.

Figure 5 Estimation of the proportion of S-RNase recognized by multiple SLF types. (**a**) The result of a Bernoulli process applied to the estimation, based on the assumption that each SLF type recognizes allelic S-RNases independently with the same probability. Error bars indicate confidential intervals. (**b**) The result of a Monte-Carlo simulation conducted to consider the difference of proportion of S-RNase recognition among SLF types. Error bars indicate standard deviation. Blue and purple horizontal lines indicate 95% and 100% of S-RNase allelic products, respectively. Dotted vertical lines indicate n = 16, 18, and 20.

Figure 6 Model for the evolution of *SLF*s. (**a**) Recombination events and their expected consequences. Possible evolutionary scenarios for three ancient haplotypes (S_{1a}, S_{2a} and S_{3a}) are indicated. Ovals and boxes indicate *SLF*s and *S-RNase*s, respectively. Red ovals indicate genetic-exchanged or integrated *SLFn* whose product recognizes S_{2a}- and S_{3a}-RNases but not S_1-RNase. Lighter colors indicate extinct haplotypes. (**b**) Predicted phylogenies of allelic *SLFn*s (upper) and *S-RNase*s (lower) of extant SI *S*-haplotypes (S_1, S_2, S_3 and S_4) and SC *S*-haplotype (S_{2m}), whose evolutionary scenarios are simulated in (a).

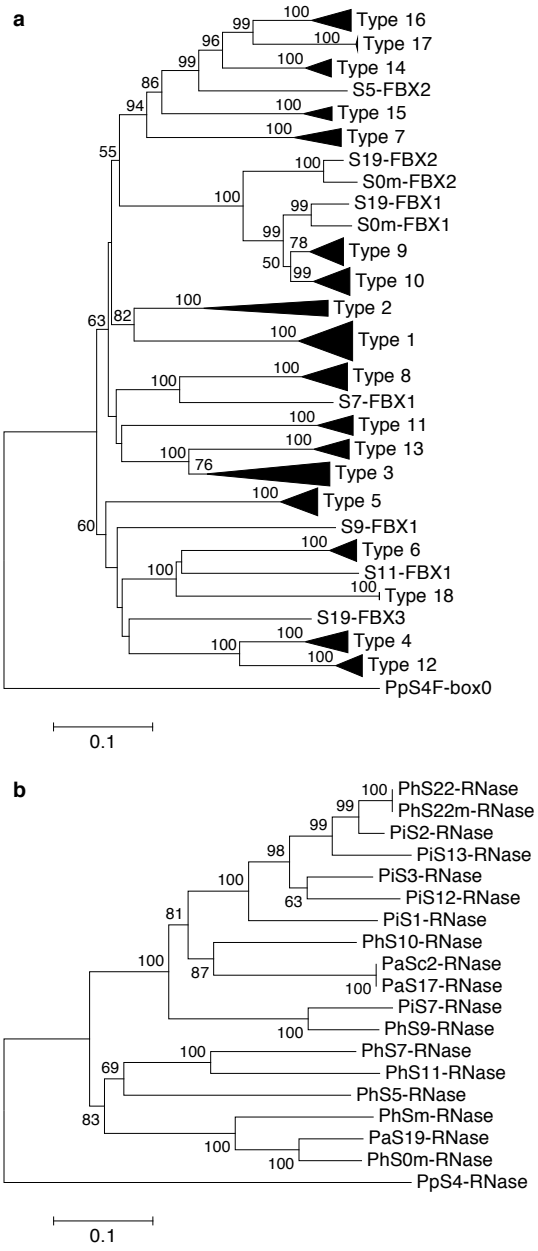

a
Type 16
Type 17
Type 14
S5-FBX2
Type 15
Type 7
S19-FBX2
S0m-FBX2
S19-FBX1
S0m-FBX1
Type 9
Type 10
Type 2
Type 1
Type 8
S7-FBX1
Type 11
Type 13
Type 3
Type 5
S9-FBX1
Type 6
S11-FBX1
Type 18
S19-FBX3
Type 4
Type 12
PpS4F-box0
0.1
b
PhS22-RNase
PhS22m-RNase
PiS2-RNase
PiS13-RNase
PiS3-RNase
PiS12-RNase
PiS1-RNase
PhS10-RNase
PaSc2-RNase
PaS17-RNase
PiS7-RNase
PhS9-RNase
PhS7-RNase
PhS11-RNase
PhS5-RNase
PhSm-RNase
PaS19-RNase
PhS0m-RNase
PpS4-RNase
0.1

Fig. 1

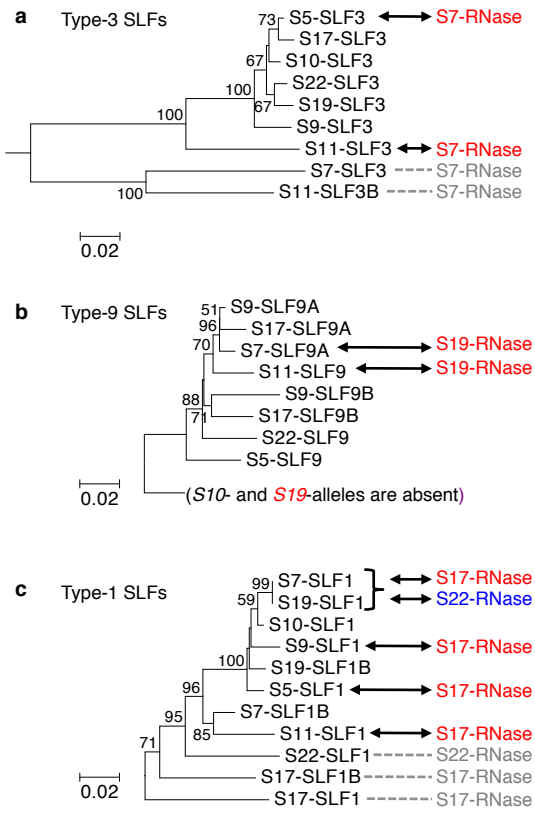
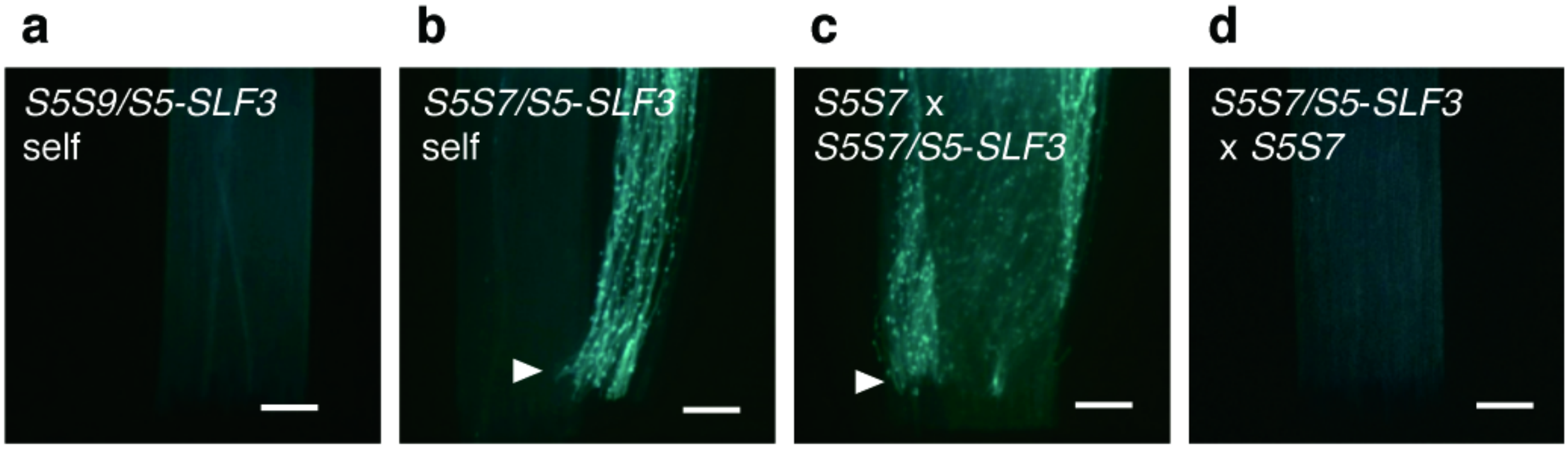
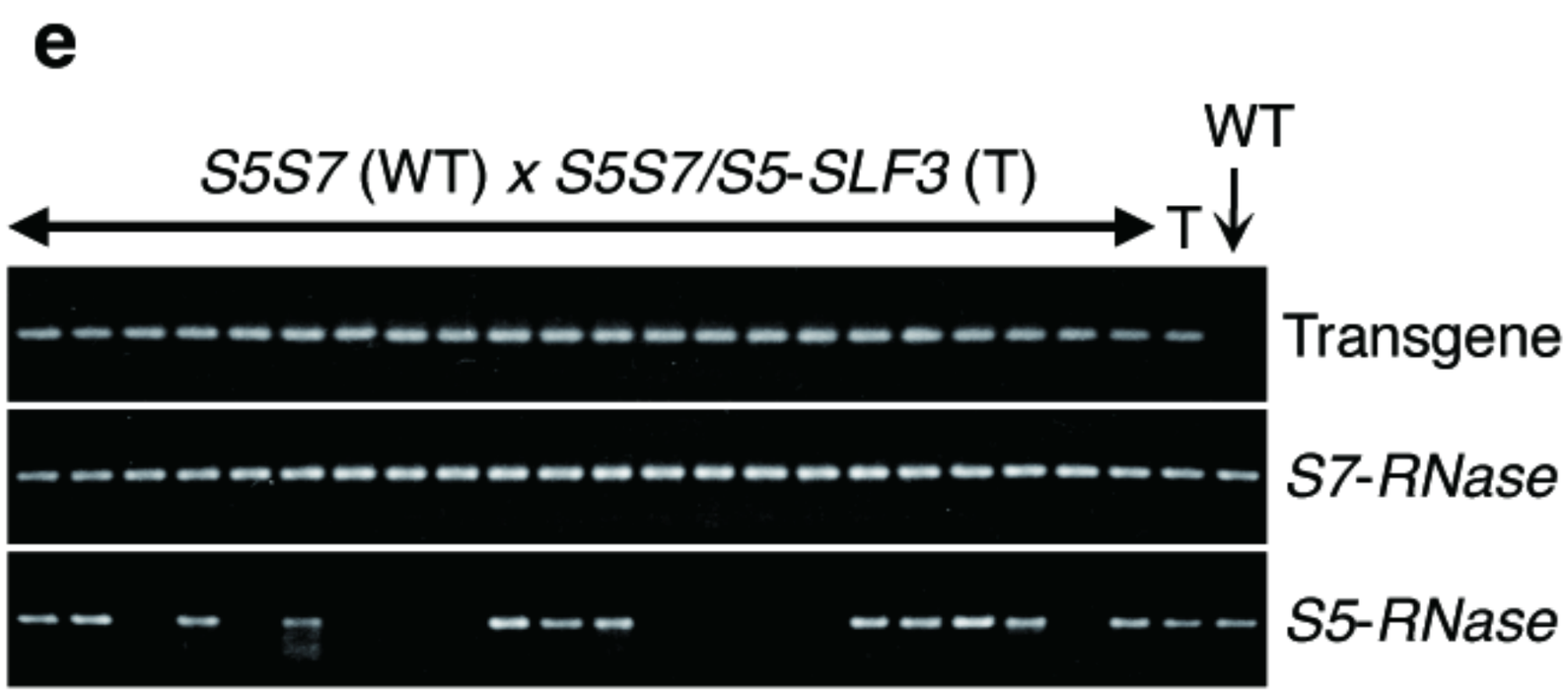
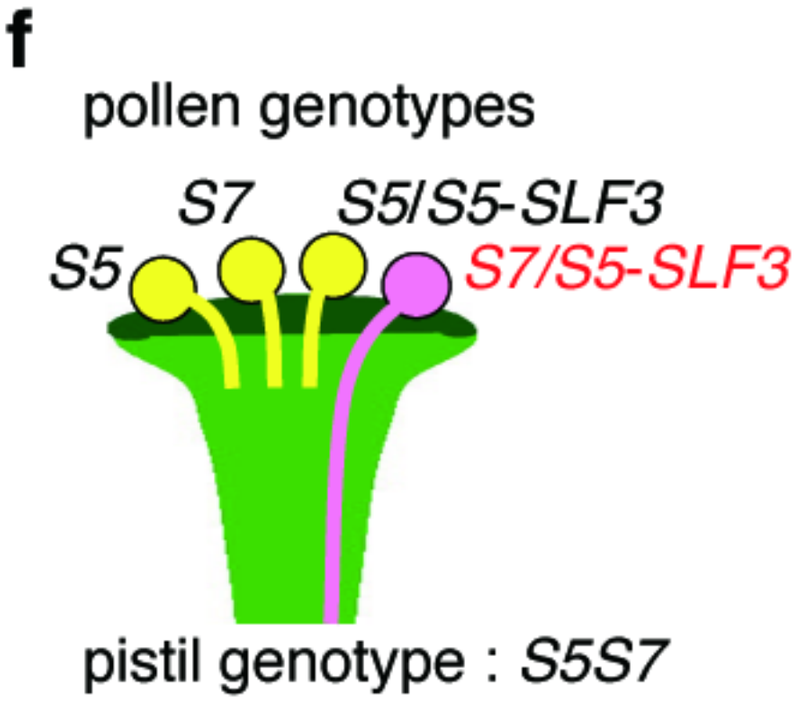

a
Type-3 SLFs
73
S5-SLF3
S7-RNase
S17-SLF3
67
S10-SLF3
100
S22-SLF3
67
S19-SLF3
100
S9-SLF3
S11-SLF3
S7-RNase
S7-SLF3
S7-RNase
100
S11-SLF3B
S7-RNase
0.02
b
Type-9 SLFs
51
S9-SLF9A
96
S17-SLF9A
70
S7-SLF9A
S19-RNase
S11-SLF9
S19-RNase
88
S9-SLF9B
71
S17-SLF9B
S22-SLF9
S5-SLF9
0.02
(S10- and S19-alleles are absent)
c
Type-1 SLFs
99
S7-SLF1
S17-RNase
59
S19-SLF1
S22-RNase
S10-SLF1
S9-SLF1
S17-RNase
100
S19-SLF1B
96
S5-SLF1
S17-RNase
95
S7-SLF1B
85
S11-SLF1
S17-RNase
71
S22-SLF1
S22-RNase
S17-SLF1B
S17-RNase
0.02
S17-SLF1
S17-RNase

Fig. 2


a
S5S9/S5-SLF3
self
b
S5S7/S5-SLF3
self
c
S5S7 x
S5S7/S5-SLF3
d
S5S7/S5-SLF3
x S5S7


e
S5S7 (WT) x S5S7/S5-SLF3 (T)
WT
T
Transgene
S7-RNase
S5-RNase


f
pollen genotypes
S7
S5/S5-SLF3
S5
S7/S5-SLF3
pistil genotype : S5S7

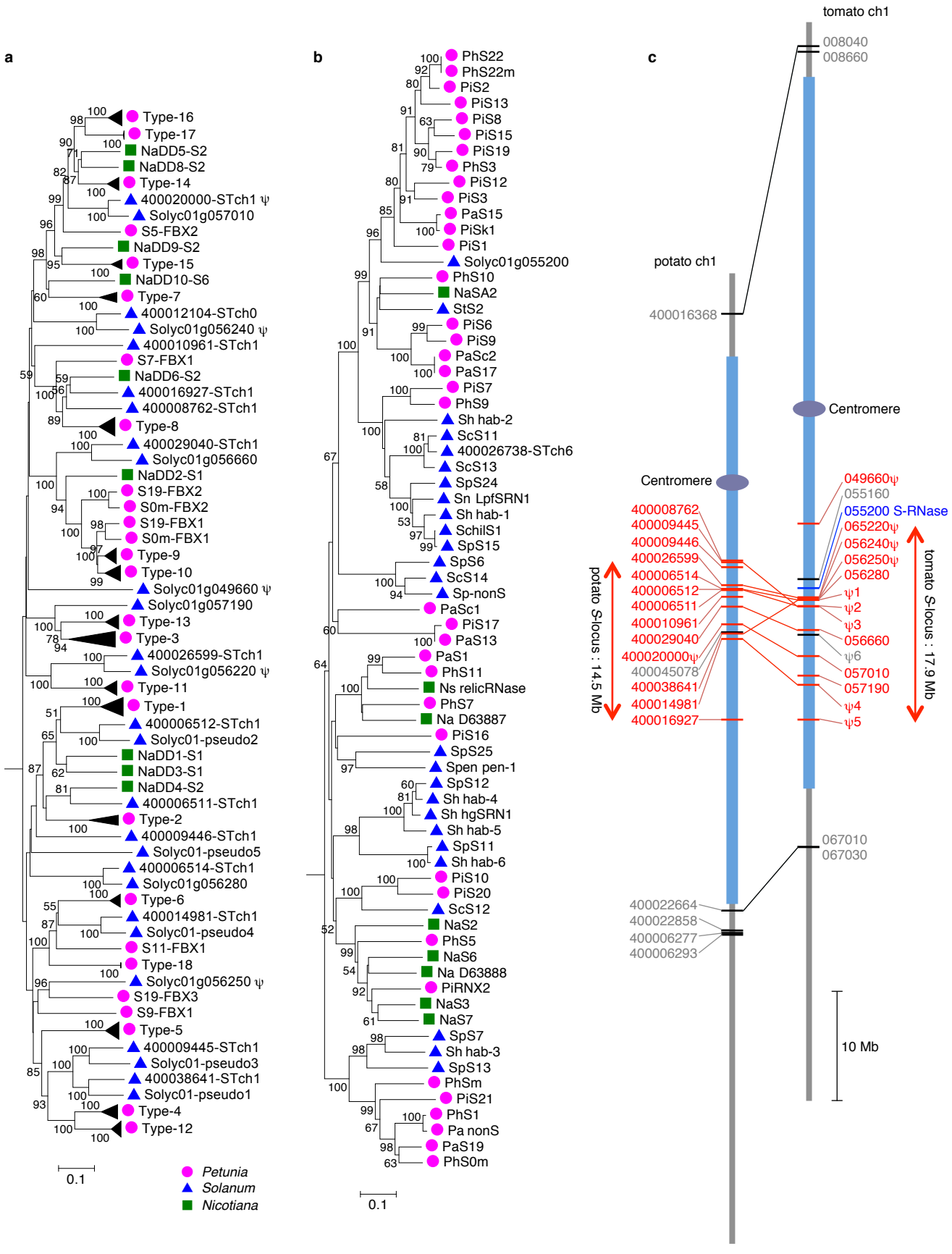

a
Type-16
Type-17
NaDD5-S2
NaDD8-S2
Type-14
400020000-STch1 ψ
Solyc01g057010
S5-FBX2
NaDD9-S2
Type-15
NaDD10-S6
Type-7
400012104-STch0
Solyc01g056240 ψ
400010961-STch1
S7-FBX1
NaDD6-S2
400016927-STch1
400008762-STch1
Type-8
400029040-STch1
Solyc01g056660
NaDD2-S1
S19-FBX2
S0m-FBX2
S19-FBX1
S0m-FBX1
Type-9
Type-10
Solyc01g049660 ψ
Solyc01g057190
Type-13
Type-3
400026599-STch1
Solyc01g056220 ψ
Type-11
Type-1
400006512-STch1
Solyc01-pseudo2
NaDD1-S1
NaDD3-S1
NaDD4-S2
400006511-STch1
Type-2
400009446-STch1
Solyc01-pseudo5
400006514-STch1
Solyc01g056280
Type-6
400014981-STch1
Solyc01-pseudo4
S11-FBX1
Type-18
Solyc01g056250 ψ
S19-FBX3
S9-FBX1
Type-5
400009445-STch1
Solyc01-pseudo3
400038641-STch1
Solyc01-pseudo1
Type-4
Type-12
0.1
Petunia
Solanum
Nicotiana
b
PhS22
PhS22m
PiS2
PiS13
PiS8
PiS15
PiS19
PhS3
PiS12
PiS3
PaS15
PiSk1
PiS1
Solyc01g055200
PhS10
NaSA2
StS2
PiS6
PiS9
PaSc2
PaS17
PiS7
PhS9
Sh hab-2
ScS11
400026738-STch6
ScS13
SpS24
Sn LpfSRN1
Sh hab-1
SchilS1
SpS15
SpS6
ScS14
Sp-nonS
PaSc1
PiS17
PaS13
PaS1
PhS11
Ns relicRNase
PhS7
Na D63887
PiS16
SpS25
Spen pen-1
SpS12
Sh hab-4
Sh hgSRN1
Sh hab-5
SpS11
Sh hab-6
PiS10
PiS20
ScS12
NaS2
PhS5
NaS6
Na D63888
PiRNX2
NaS3
NaS7
SpS7
Sh hab-3
SpS13
PhSm
PiS21
PhS1
Pa nonS
PaS19
PhS0m
0.1
c
tomato ch1
008040
008660
potato ch1
400016368
Centromere
Centromere
400008762
400009445
400009446
400026599
400006514
400006512
400006511
400010961
400029040
400020000ψ
400045078
400038641
400014981
400016927
049660ψ
055160
055200 S-RNase
065220ψ
056240ψ
056250ψ
056280
ψ1
ψ2
ψ3
056660
ψ6
057010
057190
ψ4
ψ5
potato S-locus : 14.5 Mb
tomato S-locus : 17.9 Mb
067010
067030
400022664
400022858
400006277
400006293
10 Mb

Fig. 4

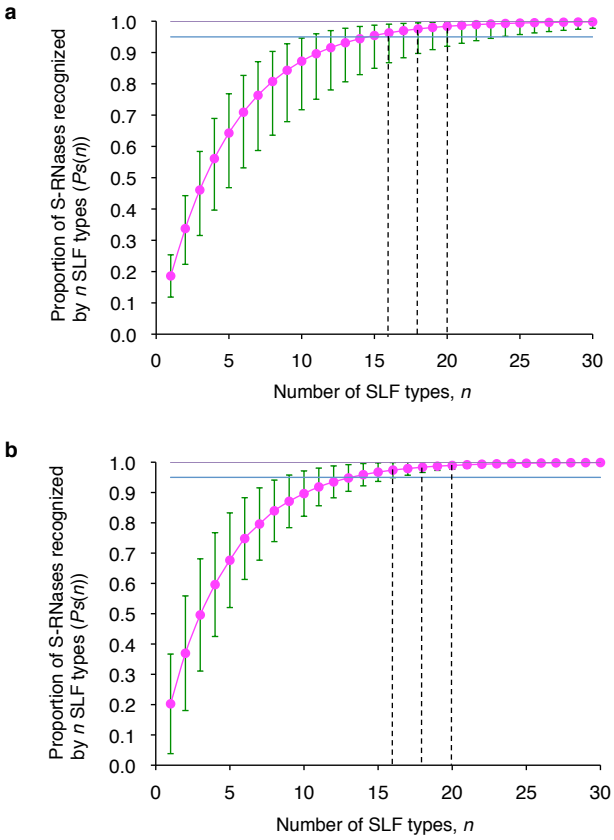

a
Proportion of S-RNases recognized by n SLF types (Ps(n))
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25
30
Number of SLF types, n
b
Proportion of S-RNases recognized by n SLF types (Ps(n))
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25
30
Number of SLF types, n

Fig. 5

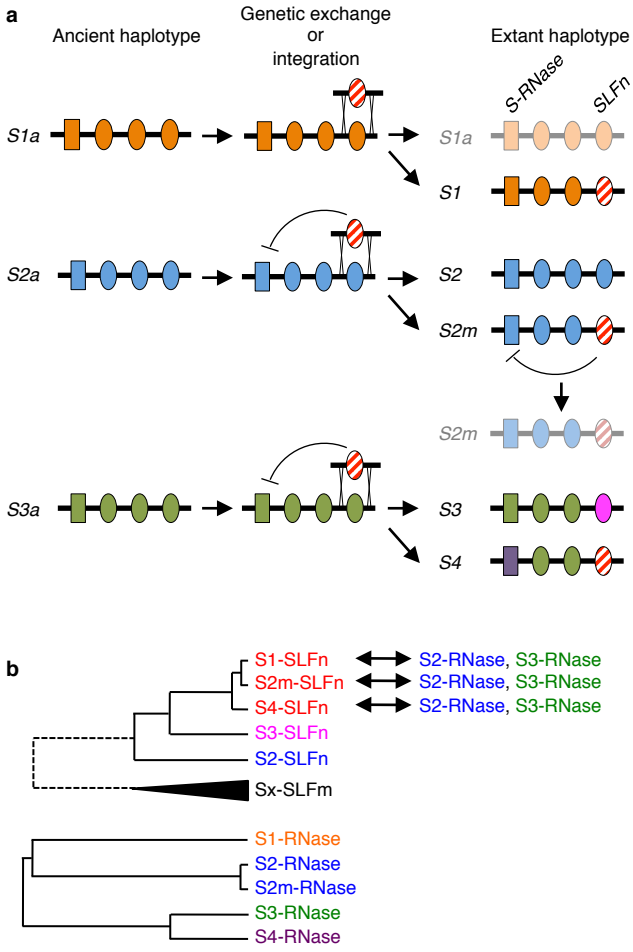

a
Ancient haplotype
Genetic exchange
or
integration
Extant haplotype
S-RNase
SLFn
S1a
S1a
S1
S2a
S2
S2m
S2m
S3a
S3
S4
b
S1-SLFn
S2m-SLFn
S4-SLFn
S3-SLFn
S2-SLFn
Sx-SLFm
S2-RNase, S3-RNase
S2-RNase, S3-RNase
S2-RNase, S3-RNase
S1-RNase
S2-RNase
S2m-RNase
S3-RNase
S4-RNase

Fig. 6