S-RNase evolution in self-incompatibility : Phylogenomic insights into synteny with Class I T2 RNase genes

Plant Physiology

Liu, Yunxiao; Zhang, Yangxin; Han, Songxue; Guo, Bocheng; Liang, Jiakai et al

https://doi.org/10.1093/plphys/kiaf072

Plant Physiology®

Plant Physiology, 2025, **197**, kiaf072
https://doi.org/10.1093/plphys/kiaf072
Advance access publication 20 February 2025
Research Article

S-RNase evolution in self-incompatibility: Phylogenomic insights into synteny with Class I T2 RNase genes

Yunxiao Liu,[1] Yangxin Zhang,[1] Songxue Han,[1] Bocheng Guo,[1] Jiakai Liang,[1] Ze Yu,[1] Fan Yang,[1] Yaqiang Sun,[1] Jiayu Xue,[2,3] Zongcheng Lin,[4] M. Eric Schranz,[5] Changfei Guan,[1,*] Fengwang Ma,[1,*] Tao Zhao[1,*]

[1]State Key Laboratory for Crop Stress Resistance and High-Efficiency Production/Shaanxi Key Laboratory of Apple, College of Horticulture, Northwest A&F University, Yangling, Shaanxi 712100, PR China
[2]College of Horticulture, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing 210095, China
[3]Center for Plant Diversity and Systematics, Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing 210014, China
[4]Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan 430070, China
[5]Biosystematics Group, Wageningen University and Research, 6708 PB Wageningen, The Netherlands
*Author for correspondence: tao.zhao@nwafu.edu.cn (T.Z.), fwm64@nwafu.edu.cn (F.M.), guanchangfei@nwafu.edu.cn (C.G.)
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is: Tao Zhao (tao.zhao@nwafu.edu.cn).

Abstract

S-RNases are essential in the gametophytic self-incompatibility (GSI) system of many flowering plants, where they act as stylar-S determinants. Despite their prominence, the syntenic genomic origin and evolutionary trajectory of *S-RNase* genes in eudicots have remained largely unclear. Here, we performed large-scale phylogenetic and microsynteny network analyses of T2 RNase genes across 130 angiosperm genomes, encompassing 35 orders and 56 families. *S-like RNase* genes in Cucurbitaceae species phylogenetically grouped with functionally characterized *S-RNases* in various species. Additionally, Cucurbitaceae *S-like RNase* genes showed conserved synteny with Class I T2 RNase genes. From this, we inferred that the well-characterized *S-RNase* genes (belonging to Class III-A genes) and Class I T2 RNase genes (located on duplicated genomic blocks) likely derived from the gamma triplication event shared by core eudicots. Additionally, we identified frequent lineage-specific gene transpositions of *S-RNases* and *S-like RNases* across diverse angiosperm lineages, including Rosaceae, Solanaceae, and Rutaceae families, accompanied by a significant increase in transposable element activity near these genes. Our findings delineate the genomic origin and evolutionary path of eudicot *S-RNase* genes, enhancing our understanding of the evolution of the *S-RNase*-based GSI system.

Introduction

RNases are integral to a variety of cellular processes, including DNA replication, RNA metabolism, plant defense, and self-incompatibility (SI) (Luhtala and Parker 2010; MacIntosh 2011). The T2 RNase family, characterized by 2 conserved active sites, is particularly widespread across organisms and plays essential biological roles (Irie 1999). In plants, the RNase T2 gene family has expanded significantly, with gene duplication and loss leading to varying numbers among species (MacIntosh et al. 2010). Phylogenetic analyses and intron count categorize these genes into 3 distinct groups, Classes I, II, and III (Igic and Kohn 2001), with *S-RNase* genes specifically classified under Class III. Despite structural similarities, they exhibit notable differences in gene expressions and functions. While Class I and Class II T2 RNases are involved in gene expression control and antimicrobial defense responses, *S-RNases* are key determinants in the GSI systems of numerous flowering plant (angiosperm) lineages (Franklin-Tong and Franklin 2003; Takayama and Isogai 2005; Hua et al. 2008; Asquini et al. 2011; Liang et al. 2020; Ramanauskas and Igić 2021).

The GSI mechanism serves as a widespread system that prevents inbreeding and promote outcrosses in bisexual flowering plant species (De Nettancourt 2001). Current knowledge identifies 4 types of SI systems (Zhao et al. 2022). Type I GSI, observed in various plant families like Rosaceae, Solanaceae, Plantaginaceae, Rutaceae, Rubiaceae, and Cactaceae families, operates through *S-RNase*-based mechanisms (Franklin-Tong and Franklin 2003; Takayama and Isogai 2005; Hua et al. 2008; Asquini et al. 2011; Liang et al. 2020; Ramanauskas and Igić 2021). Type II SI, referred to as sporophytic SI, exists in Brassicaceae and relies on *S-locus* receptor kinase and *S-locus* cysteine-rich proteins (Schopfer et al. 1999; Suzuki et al. 1999). Type III GSI, found in *Papaver*, is governed by *Papaver rhoeas* stigma *S* (PrsS) and *P. rhoeas* pollen *S* (PrpS) proteins (Foote et al. 1994; Wheeler et al. 2009). Type IV SI characterizes the sporophytic heterostyly of Primula (Huu et al. 2020; Giacomo et al. 2022). *S-RNase*-based GSI system is prevalent in eudicot species, as listed above, where it is controlled by the *S-locus*, which includes *S-RNase* genes as female determinants and *S-locus F-box* (*SLF*)/*S-haplotype specific F-box* (*SFB*) genes as male determinants, respectively, during SI interactions (Takayama and Isogai 2005). The coevolution of *S-RNase* and *SLF/SFB* genes is central to the GSI system, which is thought to have evolved once in eudicots (Steinbachs and Holsinger 2002; Vieira et al. 2008; Kubo et al. 2010).

Extensive exploration of the T2 RNase gene, encompassing *S-RNase* genes, has been conducted across diverse plant species, including representatives of Poaceae, Brassicaceae, Rutaceae, Fabaceae, Rosaceae, and Plantaginaceae (Igic and Kohn 2001; MacIntosh et al. 2010; Morimoto et al. 2015; Liang et al. 2017; Azizkhani et al. 2021; Vieira et al. 2021; Zhu et al. 2023). A considerable number of studies have notably emphasized the origin and

Received August 14, 2024. Accepted January 3, 2025.

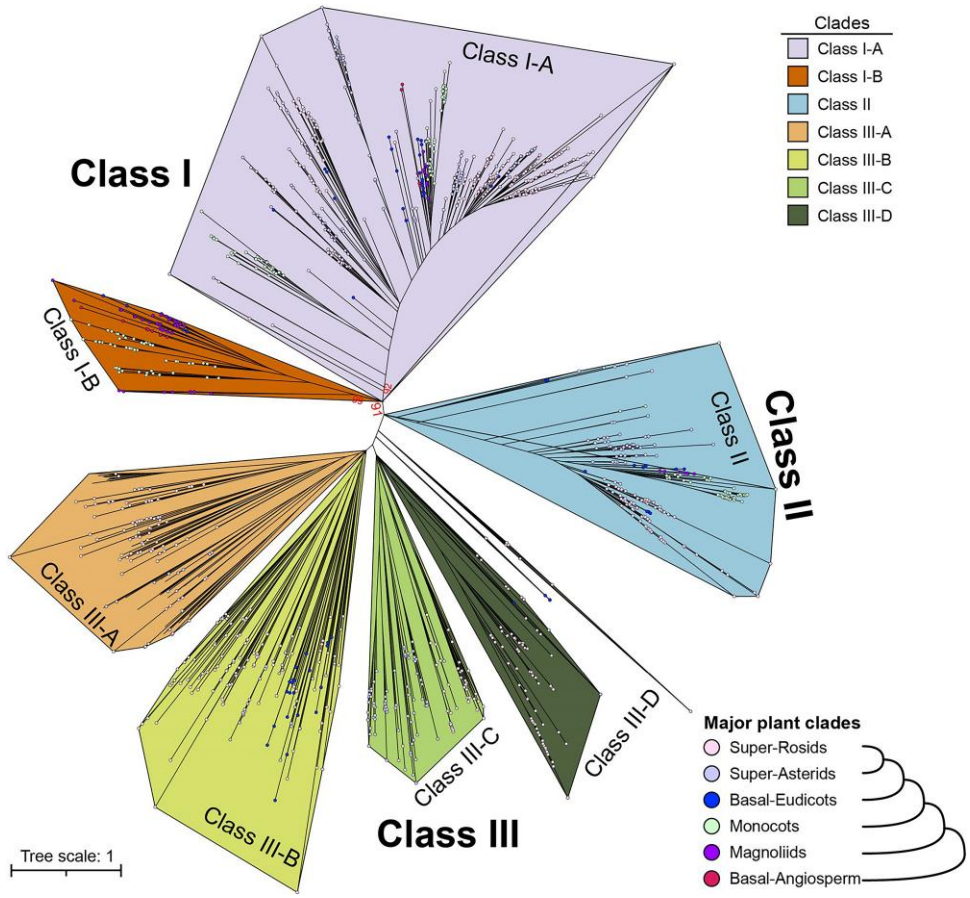

Figure 1. Phylogenetic relationships among subclasses within the 3 major classes of the RNase T2 gene family in 130 angiosperm species. The phylogenetic tree, reconstructed using full-length protein sequences of the RNase T2 gene family and ML method with 1,000 bootstrap replications, delineates the 3 major classes—Class I, Class II, and Class III—further divided into 7 subclasses. The scale bar represents the number of amino acid substitutions per site. Colored dots represent different plant lineages: superrosids (light pink), superasterids (purple), basal eudicots (blue), monocots (light green), magnoliids (purple), and basal angiosperm (magenta). Bootstrap support values are indicated for key nodes in the ML tree.

functional divergence of *S-RNase* from T2-type RNases. These investigations employ phylogenomic approaches that leverage available genomic resources, compare genomic synteny, and infer phylogenetic relationships (Vieira et al. 2008; Lv et al. 2022; Zhao et al. 2022).

While the recent evolution of *S-RNase* genes of SI systems in eudicots is widely acknowledged, the precise evolutionary processes leading to the current phylogenetic grouping of *S-RNases* from the entire T2 RNases family remain elusive. Synteny information holds crucial importance in comparative genomic research, providing insights into gene origins. A microsynteny network, complementing phylogenetic analyses, reveals profound positional relationships between subfamilies at the genomic structure level (Zhao and Schranz 2017; Zhao and Schranz 2019). Here, we leverage genomic data from 130 angiosperm species to conduct a comprehensive synteny analysis of the RNase T2 gene family, with a particular focus on unraveling the genomic origin and evolutionary trajectories of *S-RNases*.

Results

Phylogenetic analysis of the RNase T2 gene family in angiosperms

In our study, we analyzed the RNase T2 gene family across 130 fully sequenced plant species, representing diverse lineages and groups of flowering plants (Supplementary Fig. S1 and Table S1). Our analysis identified a total of 1,366 T2 RNase genes (Supplementary Table S2).

We compared the number of T2 RNase genes identified in our study with those reported in previous publications (Zhu et al. 2020; Lv et al. 2022; Zhao et al. 2022) and found a high degree of consistency (Supplementary Fig. S2). The maximum likelihood (ML) phylogenetic tree was reconstructed using the protein sequences of all identified T2 RNase genes, as well as the protein sequences of the *S-RNase* genes with reported functions (Fig. 1; Supplementary Table S3). This tree classified these genes into 3 main clades: Class I, Class II, and Class III (which includes *S-RNase*), aligning with the current phylogenetic framework (Igic and Kohn 2001; Zhao et al. 2022). Additionally, we further subdivided these classes into 7 major subclasses (Fig. 1; Supplementary Fig. S3A).

The 3 classes of T2 RNase genes showed distinct evolutionary patterns across angiosperm lineages. Class I genes were found in all major angiosperm groups, including eudicots, magnoliids, monocots, and the Amborellales, Nymphaeales, and Austrobaileyales (ANA) grade, suggesting their preservation throughout angiosperm evolution (Fig. 1; Supplementary Fig. S3B). Class II genes were present in core angiosperms, except for the ANA grade. In contrast, Class III genes, including *S-RNases*, were exclusive to eudicots, with some lineages showing gene absence or loss, such as Brassicaceae and Asteraceae (Supplementary Fig. S3B).

The RNase T2 gene family copy numbers varied significantly among angiosperm taxa, even within the same family, ranging

from 3.3 copies in basal angiosperms to 12.9 copies in superasterids (Supplementary Fig. S3B). Notably, some species like Durango root and cotton contained a higher number of genes, while others like jujube and begonia had only 3 genes each (Supplementary Fig. S3B).

The distribution of gene types within the RNase T2 gene family revealed a higher number of Class I (677 genes) and Class III (470 genes) genes compared with Class II (219 genes). This suggests a relatively limited representation of Class II genes and a notable expansion of Class I genes during plant evolution (Supplementary Fig. S3B). Our observations indicate a moderate positive correlation between the gene numbers of Class I and Class II (Supplementary Fig. S4). However, we found weaker correlations between Class I and Class III genes and between Class II and Class III genes (Supplementary Fig. S4).

We classified the tree into 7 major subclasses based on the phylogenetic relationships: Class I-A, Class I-B, Class II, Class III-A (which includes the reported *S-RNase* genes), Class III-B, Class III-C, and Class III-D, (Fig. 1; Supplementary Fig. S5, A and B). Within Class III genes, we analyzed the protein motif differences across subclasses. While no distinct motifs were identified specifically for Class III-A genes (Supplementary Fig. S6), the motif compositions of Rosaceae *S-RNases* differed greatly from those of other reported *S-RNases* (Supplementary Fig S6). Additionally, we highlighted the phylogenetic group containing reported *S-RNases* and *S-like RNases* from various species (Supplementary Fig. S7). To validate the function of representative reported *S-RNases*, we examined their expression and genomic arrangements alongside *SLF* genes (Supplementary Figs. S8 and S9).

Five types of gene duplication events were identified in the RNase T2 gene family across 130 angiosperm species (Supplementary Fig. S10). Different modes of duplication contributed to the amplification of specific gene classes in various plant lineages. Whole-genome duplication (WGD) and tandem duplication (TD) events were major drivers for Class I gene expansion, particularly in Brassicaceae, Cleomaceae, and *Gossypium*. Transposed duplications (TRDs) primarily contributed to the expansion of Class II genes in families like Rosaceae and Fabaceae. Dispersed duplications (DSDs) were the main force in the evolution of Class III genes, including *S-RNases* (Supplementary Fig. S10 and Tables S4 and S5).

Conservation of genomic contexts reveals shared ancestry between Class III-A *S-like RNase* genes in Cucurbitaceae and Class I T2 RNase genes

We extracted the T2 RNase subnetwork from the phylogenomic microsynteny database using all candidate T2 RNase genes. The network consisted of 793 nodes (T2 RNase genes) and 19,873 edges (syntenic relationships) (Supplementary Fig. S11 and Table S6). Using the infomap cluster algorithm, we identified 67 synteny clusters. Class I genes showed more interconnections compared with other classes, suggesting a stronger preservation of synteny. In contrast, Class III genes predominantly formed small-sized clusters, indicating their active evolution of genomic contexts (Supplementary Fig. S12A). Furthermore, Class III displayed the lowest gene syntenic retention rate (number of genes in synteny network relative to genes from the phylogenetic clade) and average clustering coefficient (Supplementary Figs. S12 and S13).

Mapping syntenic relationships of T2 RNase genes onto their phylogenetic tree revealed a general association between synteny and phylogenetic classifications (Fig. 2A). Comparatively, fewer syntenic connections were observed for Class III genes (Fig. 2A), where synteny still supported the classification of the 4 subclasses (Supplementary Fig. S14). The relatively low support values of the phylogenetic clades suggest a complex history of duplication, loss, and sequence divergence within this class (Supplementary Fig. S15). Notably, we observed significant syntenic conservation between some Class III-A genes and Class I T2 RNase genes (Fig. 2A). Delving into the details of the relevant synteny clusters, we identified specific synteny between Class III-A genes in Cucurbitaceae, such as those found in *Cucumis sativus*, *Citrullus lanatus*, and *Cucurbita maxima*, and Class I genes from other angiosperm species (Fig. 2B; Supplementary Fig. S16A).

The phylogenetic tree of Class III genes showed that all functionally characterized *S-RNase* genes were belong to the Class III-A branch (Fig. 3A; Supplementary Fig. S7). Notably, Cucurbitaceae T2 RNase genes clustered closely with reported Maleae *S-RNase* genes, supported by high bootstrap support values (99%) (Fig. 3B; Supplementary Fig. S7). This finding suggests that these Cucurbitaceae Class III-A T2 RNase genes can be identified as *S-RNase* putative orthologs based on phylogeny. However, as *S-RNase*-based gametophytic SI (GSI) does not apply to Cucurbitaceae species, which predominantly bear unisexual flowers, we refer to these genes as *S-like RNase* genes.

We further investigated the phylogenetic distributions of the Class III genes from all the 29 species with unisexual flowers in our dataset (Fig. 3C). The result showed that only 7 species contain genes from the Class III-A clade (Fig. 3C); however, only the genes from Cucurbitaceae species can be found in the synteny network.

We depicted some representative conserved synteny blocks between Class III-A *S-like RNase* genes in Cucurbitaceae (e.g. cucumber csa_338820) and corresponding Class I T2 RNase genes from various species, including basal angiosperms and eudicots (Fig. 3D; Supplementary Fig. S16B). We then investigated the expression pattern of these Cucurbitaceae *S-like RNase* genes and the syntenic Class I T2 RNase genes. Compared with the *S-RNase* genes from apple, pear, and peach, the nonfunctional putative orthologs of Class III-A *S-like RNase* genes in *C. sativus* (e.g. csa_338820) exhibit low expression, which is similar to its syntenic Class I T2 RNase gene (csa_285040) (Fig. 3E). This suggests little functional divergence occurred between Cucurbitaceae Class III-A *S-like RNase* and Class I T2 RNase genes.

The Class III *S-like RNase* genes of *C. sativus* represent duplicates retained from the gamma triplication event

We exploit phylogenetic analysis and ancestral synteny block reconstruction to study the origin of Cucurbitaceae Class III-A *S-like RNase* genes and Class I T2 RNase genes. Again, Cucurbitaceae T2 RNase genes can be classified into 3 classes, while collinear synteny blocks were found between Class III genes and Class I T2 RNase genes, for example between the gene pairs of csa_338820 and csa_285040 (*C. sativus*) and cla_016996 and cla_010187 (*Ci. lanatus*) (Fig. 4A).

The synonymous nucleotide substitution rate (*Ks*) distributions of intragenomic syntenic gene pairs within tested Cucurbitaceae genomes of *C. sativus*, *Ci. lanatus*, and *Cu. maxima* revealed a consensus pattern of 2 polyploidization events (Fig. 4B). Specifically, the *Ks*~1.5 peaks correspond to the reported cucurbit-common tetraploidization event (CucWGD1) (Fig. 4B) (Ma et al. 2022), while the *Ks*~2.0 peaks caused by the whole-genome triplication (gamma) event shared in core eudicots (Jiao et al. 2012). Considering the *Ks* values of the "Class III-A *S-like RNase* & Class I" syntenic pair we have pinpointed, namely 1.99 for csa_338820 and csa_285040 and 2.21 for cla_016996 and cla_010187, it is plausible

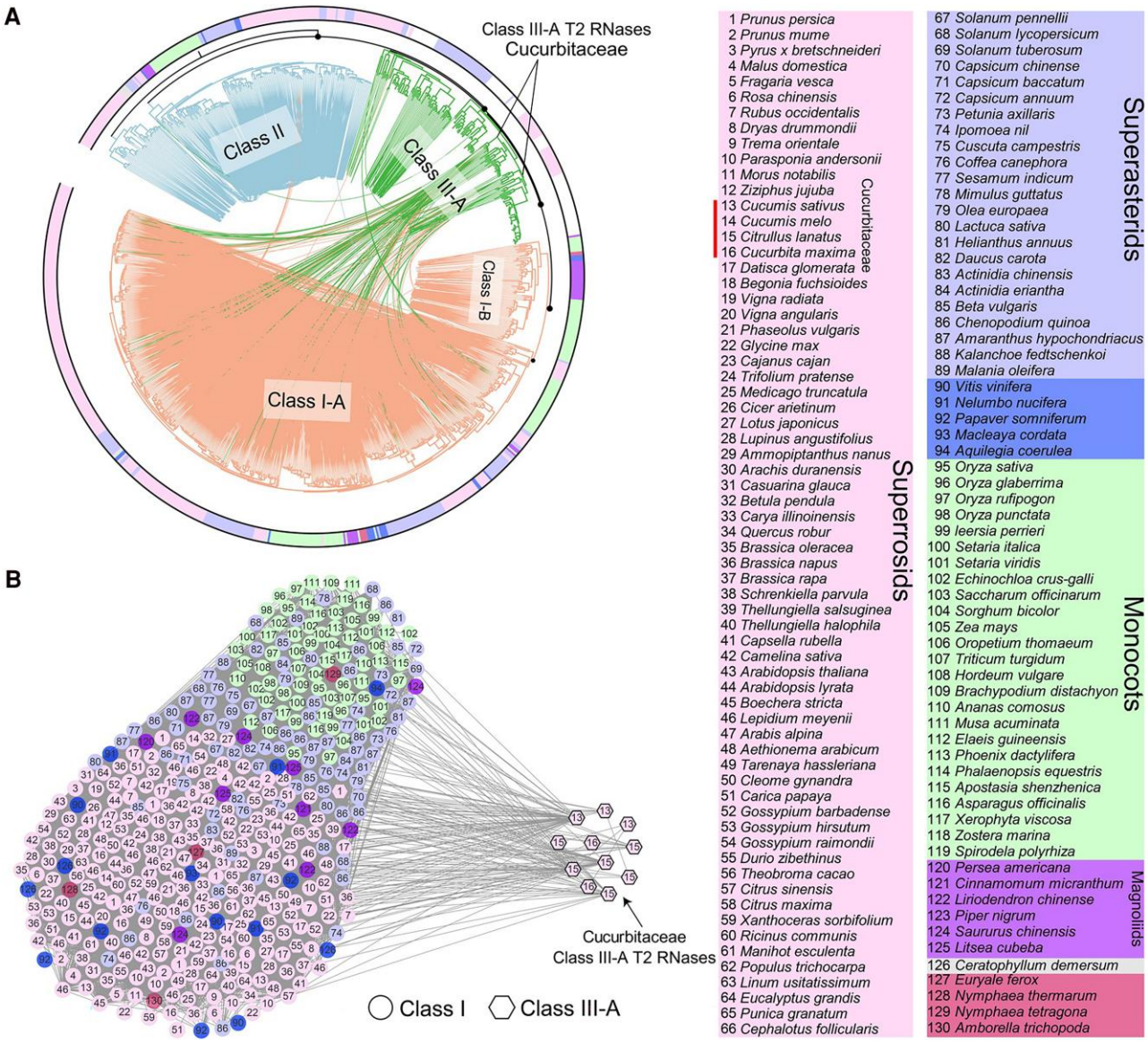

Figure 2. Phylogenomic synteny network analysis of the RNase T2 gene family in 130 angiosperm genomes. **A)** The ML gene tree of the RNase T2 gene family with syntenic relationships between the genes. Notably, synteny relationships show links between Cucurbitaceae Class III-A genes and Class I genes. **B)** Syntenic relationships for Class I and Class III-A T2 RNase genes within the phylogenomic context. The nodes are labeled and colored according to species IDs and clades on the right panel. Node shapes denote different phylogenetic clades or classes.

to conclude that the origin of Cucurbitaceae Class III-A *S-like RNase* genes can be traced back to the gamma triplication event.

Moreover, when mapping *C. sativus* genome to Ancestral Core Eudicots Karyotype (ACEK), *C. sativus* gene csa_338820 (positioned on Chromosome 3) and Class I T2 RNase gene csa_285040 (positioned on chromosome 5) correspond to Chromosomes 3 and 17 of ACEK, respectively (Fig. 4C). This result again suggests their origin from the eudicot gamma triplication event.

Noteworthy lineage-specific gene transpositions of *S-RNase* and *S-like RNase* genes, which are accompanied by transposable element activities across angiosperms

We identified 67 synteny communities from the synteny network of the RNase T2 gene family. Class III genes (including *S-RNases*) from Rosaceae, Rutaceae, Solanaceae, and Malvaceae formed distinct synteny clusters specific to their family, as did Class II (*S-like RNase*) genes observed in Brassicaceae, Cucurbitaceae, and Poaceae (Fig. 5A; Supplementary Table S7).

As an example, a detailed synteny network analysis was performed for the T2 RNase genes in 11 representative Rosaceae species. The synteny network is comprised of 138 nodes and 239 edges, and 12 clusters were identified (Supplementary Tables S8 to S10). Phylogenetic and synteny analysis was performed to elucidate their evolutionary relationships (Fig. 5B; Supplementary Figs. S17 and S18). Of note, *S-RNase* genes from Maleae species (*Malus domestica*, *Pyrus communis*, and *Eriobotrya japonica*), comparing with other Rosaceae species (*Gillenia trifoliata*, Prunoideae, and Rosoideae species), cluster in a specific synteny context and orthologous group (Fig. 5B; Supplementary Figs. S19 to S21 and Tables S11 and S12). Besides Rosaceae, we also find such lineage-specific gene transpositions of the *S-like RNase* genes within Class III-A in Solanaceae (Fig. 5C) and Rutaceae species (Fig. 5D) as well.

These findings suggest a frequent gene transposition of Class III-A *S-RNase* and *S-like RNase* genes in eudicot lineages. Upon

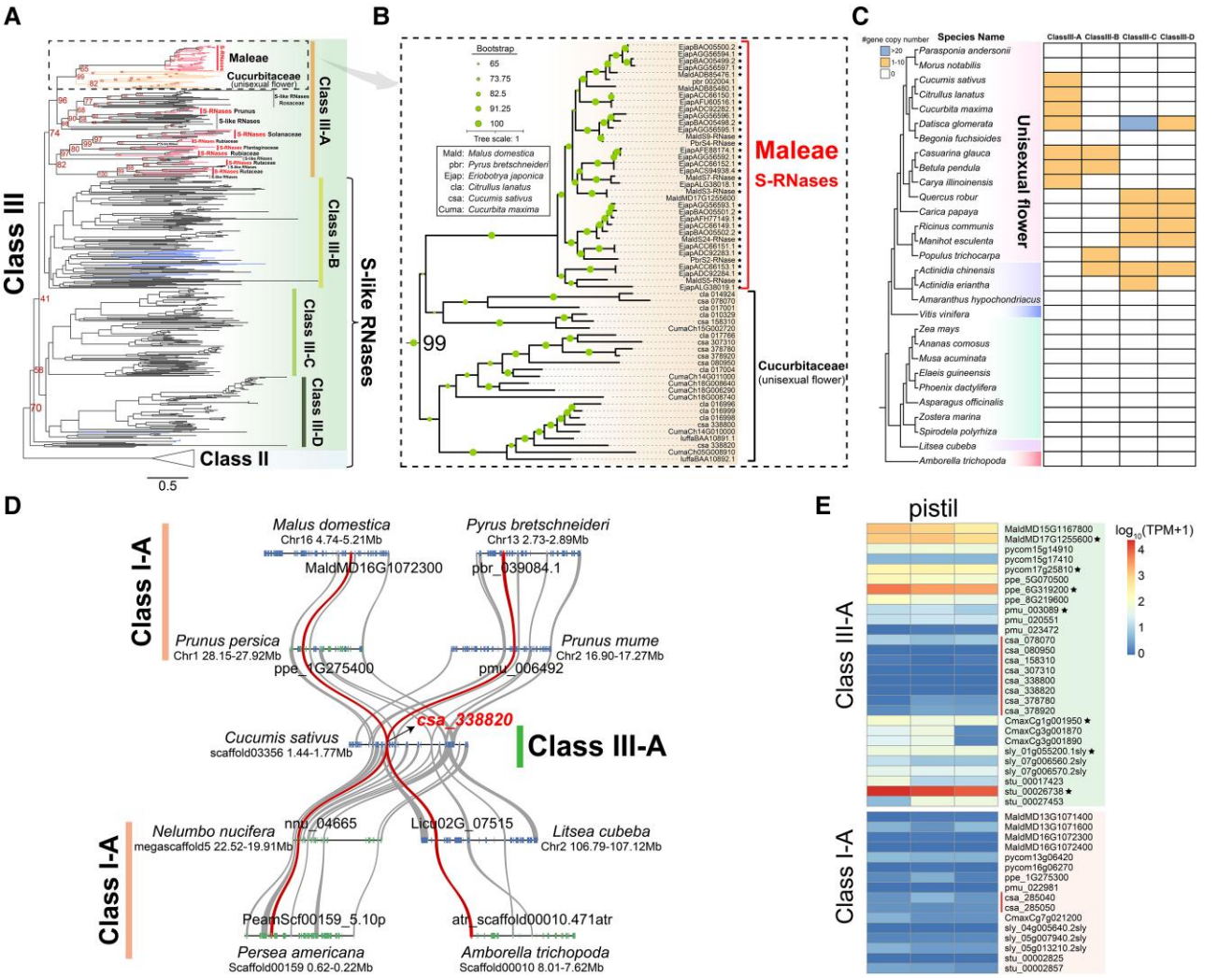

Figure 3. Phylogenetic and microsynteny analysis demonstrating that Class III-A T2 RNase genes from Cucurbitaceae are orthologous to *S-RNases* and syntenic to Class I T2 RNase genes in angiosperms. **A)** The ML tree illustrating Class III clades, derived from the broader T2 RNase phylogeny presented in Fig. 1. The scale bar represents the number of amino acid substitutions per site. Functional characterized *S-RNase* genes are marked in red, while genes from Cucurbitaceae and basal eudicots (e.g. *Aquilegia coerulea* and *Papaver somniferum*) are labeled in orange and blue, respectively. Ultrafast bootstrapping values are provided for key nodes. **B)** Enlarged view of the phylogenetic clade of Class III-A genes from **A)**, highlighting the orthology of Class III-A genes in Cucurbitaceae to functional *S-RNases* in Rosaceae. The scale bar represents the number of amino acid substitutions per site. Abbreviations of species names are displayed in the box in the upper left corner for reference. **C)** Gene copy numbers of Class III subclasses among the plant species with unisexual flowers. **D)** Microsynteny relationships of *C. sativus* Class III-A *S-like RNase* gene (csa_338820) and its syntenic genes in basal angiosperms (*A. trichopoda*), basal eudicots (*Nelumbo nucifera*), magnoliids (*Litsea cubeba* and *Persea americana*), and Rosaceae species (*M. domestica*, *Pyrus bretschneideri*, *P. persica*, and *P. mume*). Curves linking the syntenic T2 RNase genes are highlighted in dark red. **E)** Expression levels of Class III-A and Class I-A T2 RNase genes in pistil.

closer inspection, long terminal repeat (LTR) retrotransposons and terminal inverted repeats elements were found adjacent to *S-RNase* genes in several plant genomes, including *M. domestica* (Rosaceae), *Citrus maxima* (Rutaceae), *Solanum lycopersicum (Solanaceae)*, *Antirrhinum hispanicum* (Plantaginaceae), and *Coffea canephora* (Rubiaceae) (Supplementary Figs. S22 and S13). These findings highlight the impact of transposable elements (TEs) on lineage-specific transposition of *S-RNases* and *S-like RNases*, as well as the decreased synteny for Class III genes.

Discussion

GSI is a reproductive mechanism in plants that prevents self-fertilization and promotes genetic diversity by enabling pollen recognition and rejection. *S-RNases*, which are key components of this system, act as pistil-expressed genes that degrade incompatible pollen RNA to enforce SI. However, the origin of *S-RNases* has long been enigmatic. Understanding their evolution provides crucial insights into plant reproductive biology by uncovering their genomic history and shedding light on how they have diversified across species. This knowledge clarifies the genomic basis of pollen recognition and rejection systems, enhancing our understanding of species-specific reproductive adaptations and the evolution of alternative mating strategies, particularly in plants where GSI is not applicable, such as those with unisexual flowers.

Conserved synteny found between Class III-A *S-like RNase* and Class I T2 RNase genes

Phylogenomic synteny analysis is becoming a crucial tool for elucidating the conserved gene order across multiple

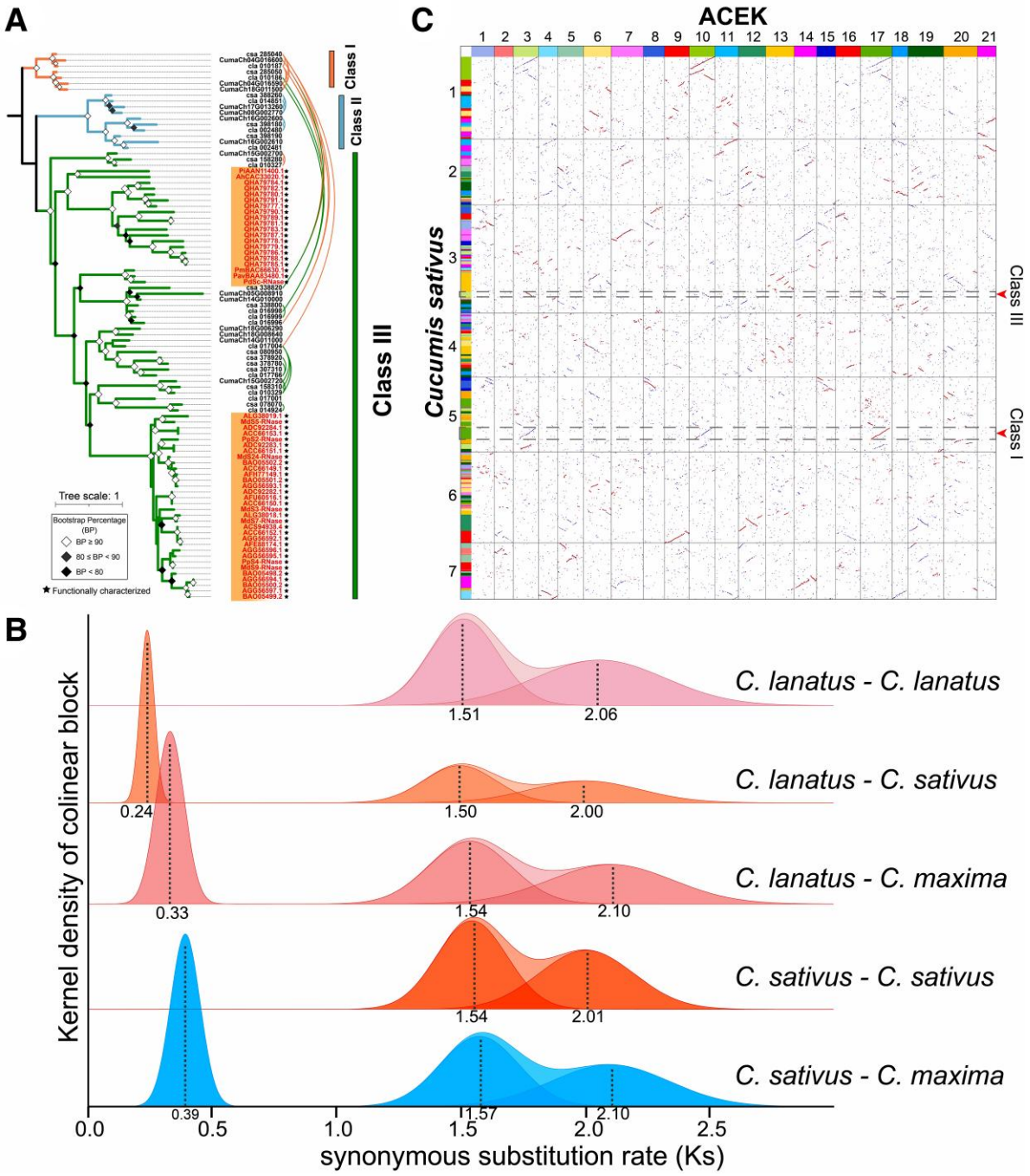

Figure 4. Dating of the Class III-A and Class I T2 RNase syntenic gene pairs in Cucurbitaceae species. **A)** ML gene tree of the RNase T2 gene family, integrated with syntenic relationships among genes in the studied Cucurbitaceae species. The phylogenetic tree is rooted with Class I genes serving as the outgroup. The scale bar represents the number of amino acid substitutions per site. Functional *S-RNase* genes are highlighted with red labels and black stars. **B)** Distributions of synonymous substitution rates (*Ks*) among collinear genes of the compared genomes. The curve is fitted using Gaussian mixture model, with peak values indicated in dashed line. The *Ks* peaks around 1.5 correspond to the cucurbit-common tetraploidization event (CucWGD1), while the peaks around 2.0 correspond to the whole-genome triplication event (gamma) in core eudicots. **C)** Homologous gene dot plots comparing the genome of *C. sativus* with ACEK. Red dots represent the best BLAST-hits, while blue dots represent other BLAST-hits. The syntenic blocks of *C. sativus* Class III-A *S-like RNase* gene (csa_338820) and Class I T2 RNase gene (csa_285040) in ACEK are indicated by gray dashed box.

genomes, inferring functional relationships, and investigating the evolutionary lineage of genes (Ruelens et al. 2013; Zhao et al. 2017; Schultz et al. 2023). Despite the abundance of research on *S-RNase* genes in certain plant families such as Rosaceae, Solanaceae, and Rutaceae, our study uncovered conserved syntenic relationships between Cucurbitaceae Class III-A *S-like RNases*, which are orthologous to functionally characterized *S-RNases* (Fig. 3; Supplementary Fig. S7), and Class I T2 RNase genes (Fig. 2). The obtained results have considerably enriched our understanding of complex evolutionary trajectories, shedding light on the existence of genomic "fossils," or remnants from previous evolutionary stages found within the well-conserved synteny in Cucurbitaceae species genomes.

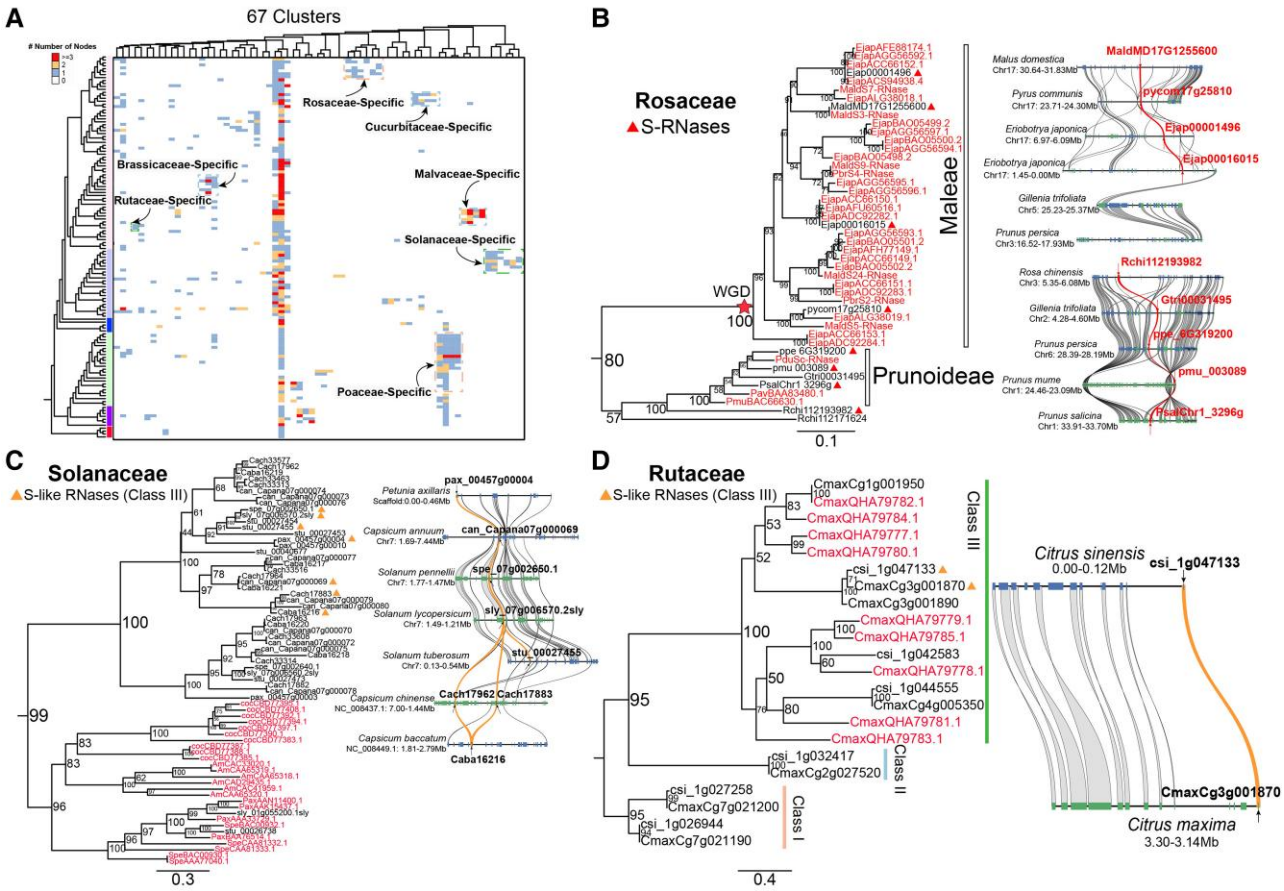

Figure 5. Examples of frequent gene transpositions of *S-RNases* and *S-like RNases* in angiosperms. **A)** Phylogenomic profiling (copy number profiling of microsynteny clusters across a phylogeny) of all synteny clusters of the RNase T2 gene family across 130 angiosperm genomes. Groups of lineage-specific clusters are boxed and labeled. Each column represents the composition of one synteny cluster, and cell colors indicate the number of nodes per species. **B** to **D)** Lineage-specific transpositions of *S-RNases* and/or *S-like RNases* in Rosaceae, Solanaceae, and Rutaceae, respectively. Reported functional *S-RNase* genes are highlighted in red, with syntenic flanking genes connected by gray curves. Genes located on the reverse strand are shown in green, while those on the forward strand are shown in blue. *S-RNases* and *S-like RNases* with transpositions are marked by red and orange triangles in the tree, respectively. Syntelogs of *S-RNases* and *S-like RNases* are linked by red and orange curve lines, respectively. The scale bar represents the number of amino acid substitutions per site.

Most species within the plant kingdom traditionally exhibit bisexual flowers (Matthews and Endress 2004). Yet, the evolution of different mechanisms aimed at curtailing self-fertilization further corroborates the inherent tendency of flowering plants to promote cross-pollination. This promotes increased genetic diversity and the fitness of their offspring. The Cucurbitaceae family, comprising of important vegetables and melons, exhibits unisexuality across most of its member species, indicating a potential link between *S-RNase* genes and the loss of SI (Boualem et al. 2015). It seems likely that Cucurbitaceae species have evolved to prevent self-pollination, exhibiting mechanisms such as monoecy, protandry, and dioecy (Steinbachs and Holsinger 2002). As a result, the *S-RNase*-based GSI mechanism, prevalent in certain plant families, probably carries less relevance (Boualem et al. 2015).

Within this evolutionary context, the well-preserved conservation of genomic context (synteny) within the Cucurbitaceae species offers compelling evidence that the Class III-A genes, inclusive of *S-RNase* genes, and Class I T2 RNase genes are duplicates originated from the eudicot gamma triplication. The study thereby underscores the presence of unisexual flowers and the absence of an *S-RNase*-based GSI mechanism, or similar reproductive systems, as likely contributing factors to such genomic conservation. These findings contribute valuable insights to our understanding of gene evolution mechanisms and plant reproductive system adaptations.

Impact of TEs to *S-RNase* evolution

DSD mode contributes most to Class III genes, which indicate their frequent transpositions. Moreover, we identified lineage-specific clusters of Class III-A genes in families like Rosaceae, Solanaceae, and Rutaceae, hinting at notable lineage-dependent transposition events of *S-RNase* and *S-like RNase* genes. The Maleae-specific gene cluster seemingly aligns with the Maleae-specific WGD event in Rosaceae (Velasco et al. 2010; Xiang et al. 2017).

TEs have often been implicated in inducing dispersed genes and transpositions, thus influencing gene composition, function, and as a result, the evolution of plant genomes (Lisch 2013). The insertion of a MITE transposon near the *S-RNase* promoter resulted in the loss of SI traits in Rutaceae lineage (Hu et al. 2024). The proximity of LTR transposons to the *S-RNase* genes, particularly in the Rosaceae lineage, emphasizes the potential of these TEs to facilitate gene rearrangements and the transposition of *S-RNase* genes, either by introducing them to new genomic landscapes or by displacing them from their initial positions, thus

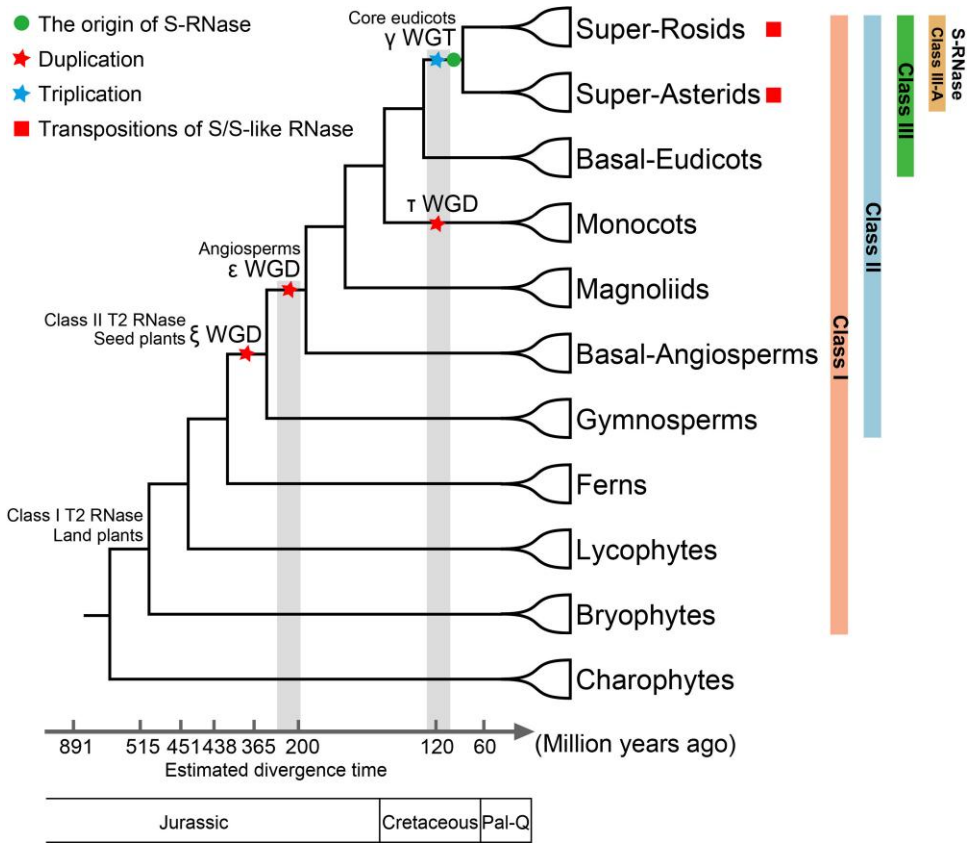

Figure 6. Evolution trajectory of *S-RNase* genes in land plants. The figure outlines the hypothesized origin and diversification of *S-RNase* genes from RNase T2 gene family over land plant evolution. The green circle represents the proposed origination of *S-RNase* genes, while the red squares denote instances of gene transpositions. Key WGD events are indicated with red pentagrams, while whole-genome triplication (WGT) events are marked with blue pentagrams, positioned along the branches of the phylogenetic tree. The timeline at the bottom illustrates estimated divergence times, calibrated based on previous studies (Morris et al. 2018; Wu et al. 2020). The phylogenetic placement of Magnoliids corresponds to the APG IV 11 classification (Chase et al. 2016). Pal-Q, Paleogene-Quaternary.

contributing to the formation of 2 lineage-specific genomic contexts in Maleae and Prunoideae.

Interestingly, the diversity in operational mechanisms within the GSI system, characterized by differential pollen recognition in 2 major branches of Rosaceae (Maleae and Prunoideae), appears to be driven by divergence in their syntenic genomic contexts (Akagi et al. 2016; Ashkani and Rees 2016; Fujii et al. 2016; Matsumoto and Tao 2016). The significant genomic disparities introduced by TE insertion within *S-RNase* genes likely contribute to the mechanistic diversity observed within pollen recognition processes.

Conclusion

Previous research has shown that Class I T2 RNase genes are present across all major land plant lineages, originating in bryophytes, while Class II T2 RNase genes emerged later, restricted to seed plants (Ramanauskas and Igić 2017). Building on this foundation, our study integrates a phylogenomic approach with synteny-based analysis to propose a detailed model for the origins and evolutionary trajectory of *S-RNase* genes, drawing on the observations discussed above (Fig. 6). We revealed that Class III-A *S-like RNase* genes in Cucurbitaceae and Class I T2 RNase genes are duplicates originating from the core-eudicot gamma triplication event. Furthermore, we highlighted the distinct genomic landscapes shaping *S-RNases* and *S-like RNases* within Class III-A across different angiosperm lineages. This comprehensive understanding of T2 RNase gene evolution enhances our knowledge of plant reproductive biology and evolutionary genetics.

Materials and methods

Plant genome resources and identification of the RNase T2 family genes

In our study, we analyzed 130 fully sequenced and annotated plant reference genomes across various taxa: 66 superrosids, 23 superasterids, 5 basal eudicots, 25 monocots, 6 Magnoliids, and 4 basal angiosperms. Within the basal angiosperms, predating the monocot–eudicot divergence, we incorporated *Amborella trichopoda*, *Euryale ferox*, *Nymphaea colorata*, and *N. thermarum* (Supplementary Fig. S1).

Genome annotations and coding sequences (CDSs) were downloaded from established repositories like NCBI, Ensemble, GigaDB, CoGe, and Phytozome and specific databases such as GDR, Sol Genomics Network, Citrus Genome Database, and others (refer to Supplementary Table S1 for detailed links). We processed annotated protein sequences and genome annotation files to include gene positional data from each genome. Species names were abbreviated into 3- or 4-letter codes and paired with their respective protein IDs. Details of the genomes and their associated information are provided in Supplementary Table S1.

Base on previous studies (Zhu et al. 2020; Lv et al. 2022; Zhao et al. 2022; Zhu et al. 2023), we identified T2 RNase family genes by using the seed alignment file of the T2 RNase domain

(PF00445) from the Pfam database to construct a Hidden Markov Model file. Potential candidates were initially identified using HMMER3.0 (Finn et al. 2011) with the default inclusion threshold (*E*-value < 1*E*−3). To ensure the completeness and accuracy of the results, the presence of the characteristic T2 RNase domain in the protein sequences was further verified using the SMART database (https://smart.embl.de/) and Pfam (http://pfam.xfam.org/). Additionally, protein sequences were analyzed with InterProScan (version 5.60) using the parameters (-appl Pfam, SMART, SuperFamily -dp -f tsv -goterms -iprlookup) (Jones et al. 2014), and only those sequences containing T2 RNase domain were retained. Correlation and statistic tests were performed using the "stats" package in R (v4.2.3). For a comprehensive list of the RNase T2 gene family members across 130 angiosperms genomes, please refer to Supplementary Table S2.

Phylogenetic reconstruction

We obtained the amino acid sequences of reported *S-RNase* and *S-like RNase* genes from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and listed them in Supplementary Table S3. For the large-scale phylogenetic gene tree of the RNase T2 gene family across 130 angiosperm species, full-length amino acid sequences were aligned with MAFFT (v7.475), with the parameters (-localpair –maxiterate 1000 –thread 20 –reorder) and the L-INS-I strategy (Katoh and Standley 2013). The alignment was subsequently manually curated in MEGA (v11) to remove gaps and discard sequences lacking conserved motifs (Tamura et al. 2011). Optimal amino acid substitution models were determined using the ModelFinder algorithm in IQ-TREE (-m MF -T AUTO). The "WAG +R8" model was identified as the best-fit model for the RNase T2 gene family across the 130 species (Nguyen et al. 2015; Kalyaanamoorthy et al. 2017). The ML phylogenetic tree was computed with IQ-TREE (v2.0.3) with the following parameters: -m WAG+R8 -alrt 1000 -bb 1000 -nt 10 (Lin et al. 2013). For the small-scale phylogenetic tree of 11 Rosaceae species, the analysis was performed with IQ-TREE (v2.0.3) using the parameters: -m VT +R6 -alrt 1000 -bb 1000 -nt 10. The resulting ML gene trees were visualized and annotated using the online tools iTOL v5 (https://itol.embl.de/) and Figtree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) (Letunic and Bork 2021).

Determination of gene duplication modes

The origins of T2 RNase genes in each species were categorized into 5 different modes of gene duplication: WGD, TD, proximal duplication (PD), TRD, and DSD. This classification was performed using DupGen_finder (https://github.com/qiao-xin/DupGen_finder) with the default parameters (Qiao et al. 2019) (Wang et al. 2012).

Specifically, tandem duplicates (TDs) refer to homologous gene copies located adjacent to each other on the same chromosome. Proximal gene pairs (PD) are nontandem pairs separated by 10 or fewer genes on the same chromosome. TRD, often labeled as "transposed," entails distally located pairs where one gene is syntenic and the other is nonsyntenic. This arrangement creates a gene pair comprising an ancestral and a novel locus. DSD events generate 2 gene copies, which are neither adjacent nor syntenic.

Any remaining duplicates that did not fall under WGD, TD, PD, or TRD were considered as dispersed duplicates. The precedence of duplicated genes was established as follows: WGD > TD > PD > TRD > DSD. Singleton genes, which lack homologous counterparts in the target species, were not classified as a type of duplication. The results of this categorization are provided in Supplementary Tables S4 and S5.

Synteny network analysis and phylogenomic profiling of synteny clusters

The SynNet construction pipeline (https://github.com/zhaotao1987/SynNet-Pipeline), developed by Zhao and Schranz (2017), was employed to build synteny networks, which encompass intra and intergenome synteny comparisons and synteny relationships among all the genes from the studied angiosperm genomes. The first step involved an all-against-all reciprocal comparison of protein sequences from each of the 130 angiosperms species under study using Diamond v3.3.2 (Buchfink et al. 2015). Next, we used MCScanX to calculate genomic collinearity (i.e. conserved gene order and content across multiple species genomes) between all pairwise genome combinations under default parameters (minimum match size for a collinear block = 5 genes, max gaps allowed = 25 genes) (Wang et al. 2012).

Following that, with the gene list of all candidate T2 RNase genes, a subnetwork specifically for the T2 RNase genes was extracted from the entire synteny network database (Supplementary Table S6). Each node in this subnetwork represents a gene, edges signify syntenic connections between genes, and edge lengths are meaningless (unweighted). To cluster the network, we implemented the Infomap algorithm (map equation framework) (Rosvall and Bergstrom 2008). The R package "*igraph*" was used to perform network statistical analysis. The resulting T2 RNase gene family synteny network was visualized and analyzed using Cytoscape v.3.8.0 (Shannon et al. 2003) and Gephi (Bastian et al. 2009).

We used the iTOL v5 web server (https://itol.embl.de/) to map syntenic relations onto the constructed phylogenetic tree. We assessed the copy number of syntenic genes across species within each community. These profiles, illustrating the count of syntenic T2 RNase genes in each genome for each synteny cluster, were organized and visualized using "Jaccard" distance and "ward.D" clustering. Detailed information about the synteny network analysis and resulting profiles are provided in Supplementary Tables S7 to S10.

We displayed notable microsynteny contexts across species using MCscan (Python version) implemented in the JCVI package (https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version)) (Tang et al. 2015). Further details can be found in Supplementary Table S11.

Ortholog identification and phylogenomic profiling of orthologous clusters

The Rosaceae dataset in this study consisted of 11 genomes, representing 3 subfamilies: Amygdaloideae (*M. domestica*, *Py. communis*, *E. japonica*, *G. trifoliata*, *Prunus persica*, *P. mume*, and *P. salicina*), Rosoideae (*Rubus occidentalis*, *Rosa chinensis*, and *Fragaria vesca*), and Dryadoideae (*Dryas drummondii*). We used OrthoFinder (version 2.5.2) with the parameters (-S diamond -M msa -A mafft -I 1.5) to identify orthologous gene groups (orthogroups) from the protein sequences of these 11 Rosaceae species (Emms and Kelly 2019).

From this analysis, we obtained a phylogenomic profile matrix comprising 28,069 nonredundant multigene clusters (orthogroups). Each orthogroup and each synteny cluster were annotated with the number of represented species to determine the dissimilarity among all the clusters. We then computed a dissimilarity index and executed hierarchical clustering using the "ward.D" method. The resulting cluster heatmap was visualized using "pheatmap," and detailed information about the identified orthogroups can be found in Supplementary Table S12.

Synonymous substitution rates (*Ks*) calculation and identification of transposons

For the Cucurbitaceae genomes, *Ks* dot plots were generated using the WGDI v.0.62 toolkit (Sun et al. 2022). Density distribution curves for *Ks* values were created using Kspeaks (-kp), and multi-peak fitting was performed with PeaksFit (-pf). The estimated mean peak values from these curves were used to date WGD events. For ACEK mapping, the "-icl" parameter was applied to identify collinear genes between ACEK and specific species, and the "-km" parameter was used to map ACEK. Finally, WGDI with the "-d" parameter and the ancestor_left (*C. sativus*) was used to produce the homologous dot plot.

We utilized the extensive de novo TE annotator (EDTA v1.9.4), with the parameter: –sensitive 1 –anno 1, to identify TEs (Su et al. 2021). Only intact TEs were considered for our study. Detailed information about the *Ks* values in the compared Cucurbitaceae genomes and the identified TEs can be found in Supplementary Table S13.

Gene expression analysis

The RNA-Seq data for this study were obtained from the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra). We performed quality control on the raw reads using Fastp v.0.12.4, applying parameters (-f 12 -F 12 -l 50) to ensure data cleanliness and high quality (Chen et al. 2018). The cleaned, high-quality reads derived from the samples were then mapped to the CDSs of the annotated genes in the reference genomes using Kallisto v.0.46.2 (Bray et al. 2016). The transcripts per million values were used to quantify the expression levels of genes within tissues. Comprehensive details regarding the RNA-seq data used can be found in Supplementary Table S14.

Annotations of the *S-locus* in various species

Annotations of the S-locus were based on previous studies (Lv et al. 2022; Zhao et al. 2022). The GSI *S-locus* comprises of *S-RNases* and closely linked *SLFs*. Detailed information regarding the identified *S-locus* is provided in Supplementary Table S15.

Accession numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers listed in Supplementary Table S3.

Acknowledgments

We would like to thank High-Performance Computing (HPC) of Northwest A&F University (NWAFU) for providing computing resources.

Author contributions

T.Z., F.M., C.G., and Y.S. designed the study. Y.L., Y.Z., S.H., B.G., J.L., Z.Y., and F.Y. conducted the analyses. T.Z., Y.L., J.X., Z.L., and M.E.S. analyzed the data. Y.L. and T.Z. wrote the paper. All co-authors read and edited the manuscript.

Supplementary data

The following materials are available in the online version of this article.

Supplementary Figure S1. Phylogenetic relationships of the 130 angiosperm and 11 Rosaceae genomes analyzed in this study.

Supplementary Figure S2. Comparative analysis of T2 RNase gene counts across studies.

Supplementary Figure S3. Phylogenetic relationship and gene counts of the 3 major classes of the RNase T2 gene family in 130 angiosperm species.

Supplementary Figure S4. Gene counts of different classes across different clades and correlation analysis.

Supplementary Figure S5. Phylogenetic relationships of the RNase T2 gene family in 130 angiosperm species.

Supplementary Figure S6. The phylogenetic tree and motif structures of the Class III T2 RNase genes in 130 angiosperm species.

Supplementary Figure S7. A detailed phylogenetic tree of Class III-A RNase genes.

Supplementary Figure S8. Gene expression levels of the *S-RNase* genes in pistil from 16 eudicot exemplar species with GSI systems.

Supplementary Figure S9. Genomic location of *S-locus*, linked *S-RNases* (Class III) and *SLFs* in 21 eudicot exemplar species from 5 plant families with the GSI system.

Supplementary Figure S10. Modes of gene duplications for identified angiosperm T2 RNase genes.

Supplementary Figure S11. A network view depicting the syntenic relationships of the entire T2 RNase gene family across 130 angiosperm species.

Supplementary Figure S12. Overview of the network and network statistic for each of the three classes of the RNase T2 gene family.

Supplementary Figure S13. Synteny networks of the Class III T2 RNase genes.

Supplementary Figure S14. Maximum-likelihood gene tree and corresponding syntenic relationships of the Class III T2 RNase genes.

Supplementary Figure S15. Unrooted tree of the major subclades of the Class III T2 RNase tree.

Supplementary Figure S16. Synteny and phylogenetic analysis highlighting Class III-A *S-like RNase* genes from Cucurbitaceae syntenic to Class I *S-like RNase* genes in angiosperms.

Supplementary Figure S17. Phylogenetic tree of the RNase T2 gene family in Rosaceae.

Supplementary Figure S18. Microsynteny network of the RNase T2 gene family in Rosaceae genomes.

Supplementary Figure S19. Phylogenomic profiling of the microsynteny clusters from 11 Rosaceae species.

Supplementary Figure S20. Maximum-likelihood phylogenetic tree of Rosaceae T2 RNase genes and corresponding synteny network clusters.

Supplementary Figure S21. Phylogenomic profiling of the orthologous gene clusters of 11 Rosaceae species.

Supplementary Figure S22. Organization of the *S-locus* genes and transposable elements in the genomes of various eudicot GSI species.

Supplementary Table S1. Comprehensive genome annotation details for the 130 angiosperm species used in this study.

Supplementary Table S2. Gene counts of the RNase T2 gene family members in each subclass and amino acid sequences of the identified T2 RNase genes.

Supplementary Table S3. Amino acid sequences of known *S-RNase* and *S-like RNase* genes.

Supplementary Table S4. Count of genes exhibiting various duplication modes within the RNase T2 gene family.

Supplementary Table S5. Count of genes categorized by various duplication modes within the RNase T2 gene family across different classes.

Supplementary Table S6. The full edge list of the T2 RNase genes synteny network across 130 angiosperm species.

Supplementary Table S7. Phylogenomic profiling of the synteny clusters of the RNase T2 gene family.

Supplementary Table S8. T2 RNase genes identified in 11 Rosaceae species.

Supplementary Table S9. The edge list of the T2 RNase synteny network of 11 Rosaceae species.

Supplementary Table S10. Phylogenomic profiling of the synteny clusters of the T2 RNase genes in 11 Rosaceae species.

Supplementary Table S11. Microsynteny relationships of the S-*RNase* genes in Maleae and *Prunus* species, referenced by the genome of *M. domestica*.

Supplementary Table S12. Phylogenomic profiling of the orthogroups of the RNase T2 gene family in Rosaceae.

Supplementary Table S13. Information of *Ks* distributions among Cucurbitaceae genomes and the identified TEs near S-*RNases*.

Supplementary Table S14. RNA-seq data used in this study.

Supplementary Table S15. Genomic structure information of S-*locus* in diverse eudicots species with GSI systems.

Funding

This work was financially supported by the State Key Research and Development Project-Youth Scientist Program (grant no. 2023YFD1202400), the National Natural Science Foundation of China (32102338), and the Chinese Universities Scientific Fund (2452021133).

Conflict of interest statement. None declared.

Data availability

Datasets used in this study are available at Zenodo (https://doi.org/10.5281/zenodo.14302948). This includes multiple sequence alignments and the phylogenetic trees. All the data supporting the research presented in this study are included in supplementary data.

References

Akagi T, Henry IM, Morimoto T, Tao R. Insights into the prunus-specific S-RNase-based self-incompatibility system from a genome-wide analysis of the evolutionary radiation of S locus-related F-box genes. *Plant Cell Physiol*. 2016:57(6):1281–1294. https://doi.org/10.1093/pcp/pcw077

Ashkani J, Rees DJ. A comprehensive study of molecular evolution at the self-incompatibility locus of Rosaceae. *J Mol Evol*. 2016:82(2–3):128–145. https://doi.org/10.1007/s00239-015-9726-4

Asquini E, Gerdol M, Gasperini D, Igic B, Graziosi G, Pallavicini A. S-RNase-like sequences in styles of coffea (Rubiaceae). evidence for S-RNase based gametophytic self-incompatibility? Trop. *Plant Biol*. 2011:4(3):237–249. https://doi.org/10.1007/s12042-011-9085-2

Azizkhani N, Mirzaei S, Torkzadeh-Mahani M. Genome-wide identification and characterization of legume T2 Ribonuclease gene family and analysis of GmaRNS9, a soybean T2 Ribonuclease gene, function in nodulation. 3 *Biotech*. 2021:11(12):495. https://doi.org/10.1007/s13205-021-03025-x

Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *Proc Int AAAI Conf Weblogs Soc Media*. 2009:8:361–362. https://doi.org/10.13140/2.1.1341.1520

Boualem A, Troadec C, Camps C, Lemhemdi A, Morin H, Sari MA, Fraenkel-Zagouri R, Kovalski I, Dogimont C, Perl-Treves R, *et al*. A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science*. 2015:350(6261):688–691. https://doi.org/10.1126/science.aac8370

Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016:34(5):525–527. https://doi.org/10.1038/nbt.3519

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015:12:59–60. https://doi.org/10.1038/nmeth.3176

Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabberley DJ, Sennikov AN, Soltis PS, Stevens PF. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc*. 2016:181(1):1–20. https://doi.org/10.1111/boj.12385

Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018:34(17):i884–i890. https://doi.org/10.1093/bioinformatics/bty560

De Nettancourt D. *Incompatibility and incongruity in wild and cultivated plants*. Berlin: Springer-Verlag; 2001.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019:20(1):238. https://doi.org/10.1186/s13059-019-1832-y

Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011:39(Web Server issue):W29–W37. https://doi.org/10.1093/nar/gkr367

Foote HC, Ride JP, Franklin-Tong VE, Walker EA, Lawrence MJ, Franklin FC. Cloning and expression of a distinctive class of self-incompatibility (S) gene from *Papaver rhoeas L*. *Proc Natl Acad Sci U S A*. 1994:91(6):2265–2269. https://doi.org/10.1073/pnas.91.6.2265

Franklin-Tong NV, Franklin FC. Gametophytic self-incompatibility inhibits pollen tube growth using different mechanisms. *Trends Plant Sci*. 2003:8(12):598–605. https://doi.org/10.1016/j.tplants.2003.10.008

Fujii S, Kubo K, Takayama S. Non-self- and self-recognition models in plant self-incompatibility. *Nat Plants*. 2016:2(9):16130. https://doi.org/10.1038/nplants.2016.130

Giacomo P, Léveillé-Bourret É, Yousefi N, Choudhury RR, Keller B, Diop SI, Duijsings D, Pirovano W, Lenhard M, Szövényi P, *et al*. Comparative genomics elucidates the origin of a supergene controlling floral heteromorphism. *Mol Biol Evol*. 2022:39(2):msac035. https://doi.org/10.1093/molbev/msac035

Hu J, Liu C, Du Z, Guo F, Song D, Wang N, Wei Z, Jiang J, Cao Z, Shi C, *et al*. Transposable elements cause the loss of self-incompatibility in citrus. *Plant Biotechnol J*. 2024:22(5):1113–1131. https://doi.org/10.1111/pbi.14250

Hua ZH, Fields A, Kao TH. Biochemical models for S-RNase-based self-incompatibility. *Mol Plant*. 2008:1(4):575–585. https://doi.org/10.1093/mp/ssn032

Huu CN, Keller B, Conti E, Kappel C, Lenhard M. Supergene evolution via stepwise duplications and neofunctionalization of a floral-organ identity gene. *Proc Natl Acad Sci U S A*. 2020:117(37):23148–23157. https://doi.org/10.1073/pnas.2006296117

Igic B, Kohn JR. Evolutionary relationships among self-incompatibility RNases. *Proc Natl Acad Sci U S A*. 2001:98(23):13167–13171. https://doi.org/10.1073/pnas.231386798

Irie M. Structure-function relationships of acid ribonucleases: lysosomal, vacuolar, and periplasmic enzymes. *Pharmacol Ther*. 1999:81(2):77–89. https://doi.org/10.1016/s0163-7258(98)00035-7

Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ, *et al*. A genome triplication associated with early diversification of the core eudicots. *Genome Biol*. 2012:13(1):R3. https://doi.org/10.1186/gb-2012-13-1-r3

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014:30(9):1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017:14(6):587–589. https://doi.org/10.1038/nmeth.4285

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013:30(4):772–780. https://doi.org/10.1093/molbev/mst010

Kubo K, Entani T, Takara A, Wang N, Fields AM, Hua Z, Toyoda M, Kawashima S, Ando T, Isogai A, *et al*. Collaborative non-self recognition system in S-RNase-based self-incompatibility. *Science*. 2010:330(6005):796–799. https://doi.org/10.1126/science.1195243

Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021:49(W1):W293–w296. https://doi.org/10.1093/nar/gkab301

Liang M, Cao Z, Zhu A, Liu Y, Tao M, Yang H, Xu Q Jr, Wang S, Liu J, Li Y, *et al*. Evolution of self-compatibility by a mutant S(m)-RNase in citrus. *Nat Plants*. 2020:6(2):131–142. https://doi.org/10.1038/s41477-020-0597-3

Liang M, Yang W, Su S, Fu L, Yi H, Chen C, Deng X, Chai L. Genome-wide identification and functional analysis of S-RNase involved in the self-incompatibility of citrus. *Mol Genet Genomics*. 2017:292(2):325–341. https://doi.org/10.1007/s00438-016-1279-8

Lin Y, Hu F, Tang J, Moret BM. Maximum likelihood phylogenetic reconstruction from high-resolution whole-genome data and a tree of 68 eukaryotes. *Pac Symp Biocomput*. 2013:2013:285–296. https://doi.org/10.1142/9789814447973_0028

Lisch D. How important are transposons for plant evolution? *Nat Rev Genet*. 2013:14(1):49–61. https://doi.org/10.1038/nrg3374

Luhtala N, Parker R. T2 Family ribonucleases: ancient enzymes with diverse roles. *Trends Biochem Sci*. 2010:35(5):253–259. https://doi.org/10.1016/j.tibs.2010.02.002

Lv S, Qiao X, Zhang W, Li Q, Wang P, Zhang S, Wu J. The origin and evolution of RNase T2 Family and gametophytic self-incompatibility system in plants. *Genome Biol Evol*. 2022:14(7):evac093. https://doi.org/10.1093/gbe/evac093

Ma L, Wang Q, Zheng Y, Guo J, Yuan S, Fu A, Bai C, Zhao X, Zheng S, Wen C, *et al*. Cucurbitaceae genome evolution, gene function and molecular breeding. *Hortic Res*. 2022:9:uhab057. https://doi.org/10.1093/hr/uhab057

MacIntosh. RNase T2 family: enzymatic properties, functional diversity, and evolution of ancient ribonucleases. In: *Nucleic acids and molecular biology*. 2011. p. 89–114.

MacIntosh GC, Hillwig MS, Meyer A, Flagel L. RNase T2 genes from rice and the evolution of secretory ribonucleases in plants. *Mol Genet Genomics*. 2010:283(4):381–396. https://doi.org/10.1007/s00438-010-0524-9

Matsumoto D, Tao R. Distinct self-recognition in the prunus S-RNase-based gametophytic self-incompatibility system. *Hort. J*. 2016:85:289–305. https://doi.org/10.2503/hortj.MI-IR06

Matthews ML, Endress PK. Comparative floral structure and systematics in Cucurbitales (Corynocarpaceae, Coriariaceae, Tetramelaceae, Datiscaceae, Begoniaceae, Cucurbitaceae, Anisophylleaceae). *Bot J Linn Soc*. 2004:145(2):129–185. https://doi.org/10.1111/j.1095-8339.2003.00281.x

Morimoto T, Akagi T, Tao R. Evolutionary analysis of genes for S-RNase-based self-incompatibility reveals S locus duplications in the ancestral Rosaceae. *Hort. J*. 2015:84:233–242. https://doi.org/10.2503/hortj.MI-060

Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. The timescale of early land plant evolution. *Proc Natl Acad Sci U S A*. 2018:115(10):E2274–E2283. https://doi.org/10.1073/pnas.1719588115

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015:32(1):268–274. https://doi.org/10.1093/molbev/msu300

Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol*. 2019:20(1):38. https://doi.org/10.1186/s13059-019-1650-2

Ramanauskas K, Igić B. The evolutionary history of plant T2/S-type ribonucleases. *PeerJ*. 2017:5:e3790. https://doi.org/10.7717/peerj.3790

Ramanauskas K, Igić B. RNase-based self-incompatibility in cacti. *New Phytol*. 2021:231(5):2039–2049. https://doi.org/10.1111/nph.17541

Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci U S A*. 2008:105(4):1118–1123. https://doi.org/10.1073/pnas.0706851105

Ruelens P, de Maagd RA, Proost S, Theißen G, Geuten K, Kaufmann K. FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes. *Nat Commun*. 2013:4:2280. https://doi.org/10.1038/ncomms3280

Schopfer CR, Nasrallah ME, Nasrallah JB. The male determinant of self-incompatibility in Brassica. *Science*. 1999:286(5445):1697–1700. https://doi.org/10.1126/science.286.5445.1697

Schultz DT, Haddock SHD, Bredeson JV, Green RE, Simakov O, Rokhsar DS. Ancient gene linkages support ctenophores as sister to other animals. *Nature*. 2023:618(7963):110–117. https://doi.org/10.1038/s41586-023-05936-6

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003:13(11):2498–2504. https://doi.org/10.1101/gr.1239303

Steinbachs JE, Holsinger KE. S-RNase–mediated gametophytic self-incompatibility is ancestral in eudicots. *Mol Biol Evol*. 2002:19(6):825–829. https://doi.org/10.1093/oxfordjournals.molbev.a004139

Su W, Ou S, Hufford MB, Peterson T. A tutorial of EDTA: extensive de novo TE annotator. *Methods Mol Biol*. 2021:2250:55–67. https://doi.org/10.1007/978-1-0716-1134-0_4

Sun P, Jiao B, Yang Y, Shan L, Li T, Li X, Xi Z, Wang X, Liu J. WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol Plant*. 2022:15(12):1841–1851. https://doi.org/10.1016/j.molp.2022.10.018

Suzuki G, Kai N, Hirose T, Fukui K, Nishio T, Takayama S, Isogai A, Watanabe M, Hinata K. Genomic organization of the S locus: identification and characterization of genes in SLG/SRK region of S(9) haplotype of *Brassica campestris* (syn. rapa). *Genetics*. 1999:153(1):391–400. https://doi.org/10.1093/genetics/153.1.391

Takayama S, Isogai A. Self-incompatibility in plants. *Annu Rev Plant Biol*. 2005:56:467–489. https://doi.org/10.1093/10.1146/annurev.arplant.56.032604.144249

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011:28(10):2731–2739. https://doi.org/10.1093/molbev/msr121

Tang H, Krishnakumar V, Li J, Zhang X. jcvi: JCVI utility libraries. Zenodo; 2015. [accessed 2024 Jan 23]. https://doi.org/10.5281/ZENODO.31631

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, *et al*. The genome of the domesticated apple (*Malus* × *domestica Borkh*.). *Nat Genet*. 2010:42:833–839. https://doi.org/10.1038/ng.654

Vieira J, Fonseca NA, Vieira CP. An S-RNase-based gametophytic self-incompatibility system evolved only once in eudicots. *J Mol Evol*. 2008:67(2):179–190. https://doi.org/10.1007/s00239-008-9137-x

Vieira J, Pimenta J, Gomes A, Laia J, Rocha S, Heitzler P, Vieira CP. The identification of the Rosa S-locus and implications on the evolution of the Rosaceae gametophytic self-incompatibility systems. *Sci Rep*. 2021:11(1):3710. https://doi.org/10.1038/s41598-021-83243-8

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, *et al*. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012:40(7):e49. https://doi.org/10.1093/nar/gkr1293

Wheeler MJ, de Graaf BH, Hadjiosif N, Perry RM, Poulter NS, OsmanK, Vatovec S, Harper A, Franklin FC, Franklin-Tong VE. Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. *Nature*. 2009:459(7249):992–995. https://doi.org/10.1038/nature08027

Wu S, Han B, Jiao Y. Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol Plant*. 2020:13(1):59–71. https://doi.org/10.1016/j.molp.2019.10.012

Xiang Y, Huang CH, Hu Y, Wen J, Li S, Yi T, Chen H, Xiang J, Ma H. Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol Biol Evol*. 2017:34(2):262–281. https://doi.org/10.1093/molbev/msw242

Zhao H, Zhang Y, Zhang H, Song Y, Zhao F, Zhang Y, Zhu S, Zhang H, Zhou Z, Guo H, *et al*. Origin, loss, and regain of self-incompatibility in angiosperms. *Plant Cell*. 2022:34(1):579–596. https://doi.org/10.1093/plcell/koab266

Zhao T, Holmer R, de Bruijn S, Angenent GC, van den Burg HA, Schranz ME. Phylogenomic synteny network analysis of MADS-box transcription factor genes reveals lineage-specific transpositions, ancient tandem duplications, and deep positional conservation. *Plant Cell*. 2017:29(6):1278–1292. https://doi.org/10.1105/tpc.17.00312

Zhao T, Schranz ME. Network approaches for plant phylogenomic synteny analysis. *Curr Opin Plant Biol*. 2017:36:129–134. https://doi.org/10.1016/j.pbi.2017.03.001

Zhao T, Schranz ME. Network-based microsynteny analysis identifies major differences and genomic outliers in mammalian and angiosperm genomes. *Proc Natl Acad Sci U S A*. 2019:116(6):2165–2174. https://doi.org/10.1073/pnas.1801757116

Zhu S, Zhang Y, Copsy L, Han Q, Zheng D, Coen E, Xue Y. The snapdragon genomes reveal the evolutionary dynamics of the S-Locus supergene. *Mol Biol Evol*. 2023:40(4):msad080. https://doi.org/10.1093/molbev/msad080

Zhu X, Li Q, Tang C, Qiao X, Qi K, Wang P, Zhang S, Wu J. Comprehensive genomic analysis of the RNase T2 gene family in Rosaceae and expression analysis in *Pyrus bretschneideri*. *Plant Syst Evol*. 2020:306(4):71. https://doi.org/10.1007/s00606-020-01700-9