PSIA: A Comprehensive Knowledgebase of Plant Self-incompatibility

Chen Wang (王晨) [1,2,3,#], **Hong Zhao (赵洪)** [4,#], **Hongkui Zhang (张洪魁)** [1,2,3], **Sijie Sun (孙思杰)** [1,2,3], **Yongbiao Xue (薛勇彪)** [1,2,3,4,*]

[1]National Genomics Data Center, China National Center for Bioinformation, Beijing 100101, China
[2]Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
[4]Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China
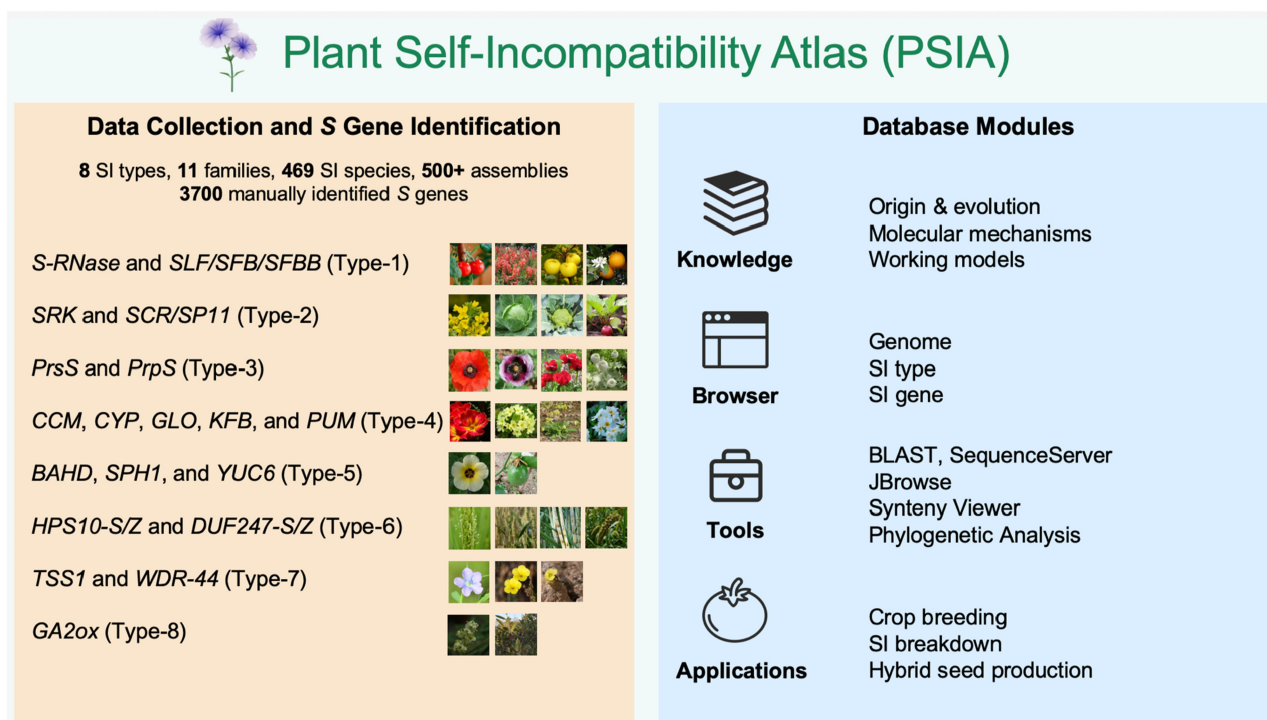*Corresponding author: ybxue@genetics.ac.cn (Xue Y).
#Equal contribution.
Handling Editor: Peng Cui

Abstract

Self-incompatibility (SI) is an important genetic mechanism in angiosperms that prevents inbreeding and promotes outcrossing, with significant implications for crop breeding, including genetic diversity, hybrid seed production, and yield optimization. In eudicots, SI is typically governed by a single *S*-locus containing tightly linked pistil and pollen *S*-determinant genes. Despite major advances in SI research, a centralized, comprehensive resource for SI-related genomic data remains lacking. To address this gap, we developed the Plant Self-Incompatibility Atlas (PSIA), a systematically curated knowledgebase providing an extensive compilation of plant SI, including genomic resources for SI species, *S* gene annotations, molecular mechanisms, phylogenetic relationships, and comparative genomic analyses. The current release of PSIA includes over 500 genome assemblies from 469 SI species. Using known *S* genes as queries, we manually identified and rigorously curated 3700 *S* genes. PSIA provides detailed *S*-locus information from assembled genomes of SI species and offers an interactive platform for browsing, BLAST searches, *S* gene analysis, and data retrieval. Additionally, PSIA serves as a unique platform for comparative genomic studies of *S*-loci, facilitating exploration of the dynamic processes underlying the origin, loss, and regain of SI. As a comprehensive and user-friendly resource, PSIA will greatly advance our understanding of angiosperm SI and serve as a valuable tool for crop breeding and hybrid seed production. PSIA is freely available at http://www.plantsi.cn.

Key words: Plant Self-Incompatibility Atlas; Genome; SI type; *S* gene; Evolution.

Graphical abstract

Introduction

Flowering plants (angiosperms) comprise approximately 90% of all terrestrial plant species, forming the foundation of most ecosystems and playing a pivotal role in human livelihoods [1]. Self-incompatibility (SI), present in over 40% of angiosperms [2], is a reproductive mechanism that prevents self-fertilization in fertile plants. As a strict intraspecific barrier, SI prevents inbreeding and promotes outcrossing, which is essential for maintaining genetic diversity and the adaptive survival of populations [3]. While SI maintains genetic diversity in natural populations, it also presents challenges for breeding programs aiming for trait uniformity. Thus, understanding SI mechanisms is crucial for hybrid seed production and crop yield improvement, particularly in economically important genera such as *Solanum*, *Prunus*, *Citrus*, and *Brassica* [4].

In eudicots, SI is typically controlled by a highly polymorphic, multiallelic *S*-locus [5], whereas in grasses, it is governed by two independently inherited, polymorphic, multiallelic *S* and *Z* loci [6]. These loci encode distinct determinants that mediate pistil–pollen recognition and are inherited as a single segregating unit, with their variants termed *S* haplotypes [5]. In eudicots, self-pollen incompatibility (SPI), leading to self-pollen rejection, is triggered by the self-recognition of pistil and pollen *S* determinants encoded by the same *S* haplotype. In contrast, in grasses, SPI occurs only when both the *S* and *Z* haplotypes match. SI systems can be categorized into homomorphic (gametophytic and sporophytic) and heteromorphic (heterostyly) types based on their association with floral morphology [3,5,7–9]. Gametophytic self-incompatibility (GSI), which rejects self-pollen when its *S* haplotype matches either of the two *S* haplotypes in the pistil, has been identified in 17–25 plant families. In contrast, sporophytic self-incompatibility (SSI) inhibits self-pollen when the *S* haplotype of the sporophytic pollen parent matches either *S* haplotype in the diploid pistil. SSI is present in several plant families, including Brassicaceae, Asteraceae, Convolvulaceae, and Betulaceae [10–14]. Heterostyly, a floral polymorphism found in 28 families, including Primulaceae, Turneraceae, Linaceae, and Oleaceae [15–18], combines morphological and physiological incompatibility to prevent within-morph fertilization.

Given the substantial variation in the molecular mechanisms underlying SI systems across different plant families, these systems can be categorized into eight distinct types. The most widespread mechanism is gametophytic type-1 SI, governed by the pistil *S-RNase* and pollen *SLF*. This mechanism has been identified in Plantaginaceae [19–22], Solanaceae [23–25], Rosaceae [26–28], and Rutaceae [29]. Sporophytic type-2 SI, characteristic of Brassicaceae, is determined by *SRK* and *SP11/SCR* [10,11]. Gametophytic type-3 SI, observed in *Papaver rhoeas*, is controlled by *PrsS* and *PrpS* [30,31]. Heterostyly involves a hemizygous *S*-locus that regulates both floral morphology and incompatibility responses. This mechanism includes type-4 SI in Primulaceae, governed by *CYP*, *GLO2*, *KFB*, *CCM*, and *PUM* [15]; type-5 SI in Turneraceae, defined by *SPH1*, *YUC6*, and *BAHD* [16]; type-7 SI in Linaceae, determined by *TSS1* and *WDR-44* [17]; and type-8 SI in Oleaceae, governed by *GA2ox* [18]. Gametophytic type-6 SI in Poaceae is controlled by the *S* and *Z* loci, both of which encode HPS10 and DUF247I/II [6,32,33].

Advances in third-generation sequencing technologies have significantly improved plant genome assembly, leading to the development of dedicated genomic databases. For example, t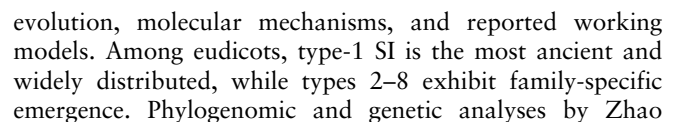he Sol Genomics Network [34] serves as a comprehensive platform for Solanaceae species, integrating genomic, genetic, and breeding data, while the Genome Database for Rosaceae [35] provides a centralized repository for Rosaceae genomic and breeding resources. Despite these advancements, SI information remains largely overlooked in plant genomic studies and family-level databases, resulting in the underutilization of *S*-locus genomic resources. Notably, species within Solanaceae, Plantaginaceae, Rosaceae, and Rutaceae share the same SI type; however, no existing database integrates *S*-locus data across these families for systematic analysis. Given the increasing availability of sequenced plant genomes, a comprehensive SI knowledgebase is urgently needed to facilitate research on SI mechanisms, evolution, and applications in crop breeding.

To address this gap, we developed the Plant Self-Incompatibility Atlas (PSIA), a user-friendly and systematically curated database that integrates over 500 genome assemblies from 469 SI species across 11 plant families, covering all 8 SI types. PSIA consolidates extensive SI information, including its origin, evolution, molecular mechanisms, and working models. Additionally, we systematically identified and annotated 3700 *S* genes, with manual curation ensuring accuracy. PSIA provides a suite of analytical tools, including genome-wide BLAST, *S* gene-specific BLAST, genome browsing, and SI data analysis. To our knowledge, PSIA is the first knowledgebase to systematically compile and analyze SI-related genomic data across all plant species with well-characterized molecular mechanisms. Moreover, it is the first resource to integrate information on SI origin, evolution, and molecular mechanisms at the angiosperm-wide level. Consequently, PSIA represents the most comprehensive SI knowledgebase to date, serving as an essential resource for researchers investigating plant reproductive biology, SI mechanisms, and their implications for plant breeding.

Database construction and content

Integration of genome resources

Recent studies have identified 8 distinct molecular types of SI, spanning 11 families that include crops, fruits, ornamental plants, and model species (**Figure 1**A and B). To construct a comprehensive knowledgebase of SI, all publicly available genome assemblies for these families were retrieved from multiple databases, including National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/), Genome Warehouse (GWH, https://ngdc.cncb.ac.cn/gwh) [36], Published Plant Genomes (https://www.plabipd.de/plant_genomes_pa.ep), Phytozome (v13, https://phytozome-next.jgi.doe.gov) [37], Sol Genomics Network (https://solgenomics.net) [34], Genome Database for Rosaceae (GDR, https://www.rosaceae.org) [35], Genome Database for Strawberry (GDS, http://eplant.njau.edu.cn/strawberry/) [38], Citrus Genome Database (https://www.citrusgenomedb.org), Brassicaceae Database (BRAD, http://www.brassicadb.cn) [39], and other online resources such as Figshare and Google Drive (Figure 1C). The current release of PSIA includes over 500 genome assemblies from 469 SI species. To enable comprehensive retrieval of SI data, we implemented a fuzzy search function that accepts both scientific and common names (*e.g.*, "*Solanum lycopersicum*" and "tomato", respectively). Figure 1D and Figure S1 provide an overview of the SI data available in PSIA, while Figure S2 displays the species tree of the eight SI types included in the database.

Figure 1 Overview of PSIA

A. Representative species illustrating the eight known SI types. **B**. Phylogenetic relationships of all currently included plant families in PSIA. **C**. Available genome resources for SI species. **D**. Summary of SI-related data in the latest PSIA release. **E**. Evolutionary origins and diversification of the eight SI types. **F**. Schematic representation of the molecular mechanism underlying type-1 SI, alongside a comparative overview of all eight SI types in PSIA. **G**. Experimentally derived working model of SI in *Petunia hybrida*. **H**. Computational pipeline for *S* gene identification. **I**. Key analytical tools integrated into PSIA. **J**. Functional capabilities of the BLAST and JBrowse tools in PSIA. **K**. Example output of synteny analysis comparing *Solanum S*-loci. **L**. Phylogenetic reconstruction of S-RNases across representative families, including Plantaginaceae, Solanaceae, Rosaceae, and Rutaceae. PSIA, Plant Self-Incompatibility Atlas; SI, self-incompatibility.

Knowledge of SI origin, evolution, molecular mechanisms, and working models

To facilitate systematic SI research, PSIA incorporates comprehensive information on plant SI, including its origin, evolution, molecular mechanisms, and reported working models. Among eudicots, type-1 SI is the most ancient and widely distributed, while types 2–8 exhibit family-specific emergence. Phylogenomic and genetic analyses by Zhao

et al. [40] suggest that *S-RNase*, *SLF/SFB/SFBB*, and type-1 *S*-loci trace back to the most recent common ancestor (MRCA) of eudicots. Under selective pressure favoring out-crossing, type-1 SI has been preserved in Plantaginaceae, Solanaceae, Rosaceae, and Rutaceae through selective dele-tion or inactivation of duplicate *S*-loci following whole-genome duplication (WGD) [40]. In contrast, Brassicaceae, Papaveraceae, Primulaceae, Turneraceae, Linaceae, and Oleaceae have lost type-1 SI via *S*-locus deletions or the main-tenance of duplications, giving rise to new, family-specific SI systems (types 2–5, 7, and 8) (Figure 1E). Given that the an-cient type-1 *S*-locus, encoding Class I/II T2 RNase and FBA/FBK proteins, originated from the MRCA of angio-sperms, it is plausible that monocots acquired type-6 SI after the loss of type-1 SI (Figure 1E, Figure S3). To enhance un-derstanding, PSIA provides molecular mechanism summaries for all eight SI types through schematic illustrations (Figure 1F, Figures S4–S10). Additionally, PSIA includes two experimentally derived working models of *Petunia hybrida* SI (Figure 1G) [41,42], providing the theoretical models for type-1 SI research.

S gene identification and curation

The *S*-loci from assembled genomes of SI species were com-prehensively analyzed in PSIA (Figure 1H) based on well-characterized types of molecular mechanisms. First, we re-trieved all publicly available *S* gene sequences from the NCBI Nucleotide and Protein databases, compiling a dataset of 1275 nucleotide sequences and 1130 protein sequences. Using BLAST+ [43] with default parameters, we identified *S* gene candidates across the collected genome assemblies. Putative *S* gene coding regions were manually curated to en-sure accurate splice site annotations. To refine candidate se-lection, we applied multiple filtering criteria, including the presence of conserved domains and genomic linkage distan-ces. Specifically, type-1 *SLF/SFB/SFBB* genes were identified based on the presence of F-box and FBA/FBK domains, type-1 *S-RNase* genes were characterized by the Ribonuclease T2 family domain, and type-6 *DUF247I/II-S/Z* genes were dis-tinguished by the DUF247 domain. Additionally, genomic linkage distances were incorporated as an additional filtering criterion, as *S* genes within each SI family exhibit characteris-tic genomic distributions, with linked male and female deter-minant *S* genes typically falling within a similar genomic span. To enhance candidate selection accuracy, we compared these distances with those reported in previous SI studies. For example, the *S*-locus spans approximately 17 Mb in *S. lyco-persicum* (Solanaceae) [44], 1.2 Mb in *Antirrhinum* species (Plantaginaceae) [45], 1.36–2.64 Mb in *Pyrus* and *Malus* spe-cies (Rosaceae) [46], and 198–378 kb in *Citrus* species (Rutaceae) [29]. Based on the aforementioned strategies, we identified and manually curated a total of 3700 *S* genes in PSIA.

Useful tools

PSIA integrates user-friendly tools, including BLAST, SequenceServer, JBrowse, Synteny Viewer, and Phylogenetic Analysis (Figures 1I–L). These tools can be accessed through the navigation bar on the database interface (Figure S11). Sequence similarity analysis is a fundamental bioinformatics method for detecting sequence homology and inferring po-tential functional relationships among genes. While the NCBI BLAST database remains a standard tool, its utility in SI

research is constrained. To overcome these limitations, PSIA integrates both ViroBLAST [47] and SequenceServer [48] platforms, providing specialized analytical capabilities for SI studies including genome-wide and *S* gene-specific BLAST. Local genome-wide BLAST databases have been established using several genome assemblies representing different SI types. The BLAST procedure is straightforward, and optimal alignment results can be obtained using default parameters (Figure S12). PSIA hosts the most comprehensive collection of *S* gene sequences related to plant SI. To facilitate the utili-zation of these *S* gene resources, local databases of SequenceServer have been developed using thousands of *S* genes identified and curated by PSIA. SequenceServer ena-bles faster and more accurate identification of *S* genes, with a user-friendly workflow (Figure S13), which is expected to sig-nificantly enhance the efficiency of SI research.

Genome browsers implemented on servers enable users to visualize genomic data online. For genome assemblies with accessible GFF files, a JBrowse function [49] has been pro-vided, enabling researchers to browse the entire genome, *S*-lo-cus regions, and *S* gene annotations (Figure S14). The SynVisio component (https://github.com/kiranbandi/synvisio) facilitates comparative genomic studies of different genome assemblies and *S*-loci. Synteny analysis of the *S*-loci will con-tribute to a deeper understanding of the dynamic evolution of SI (Figures S15 and S16). Additionally, phylogenetic trees have been constructed using IQ-TREE [50] based on the *S* genes identified in this study, which will aid in investigating their origins and relationships (Figure S17).

Implementation

PSIA runs on a Linux-based Apache (https://httpd.apache.org) web server using PHP (a popular general-purpose script-ing language, especially suited to web development; https://www.php.net) as the backend. The web interface is built with Bootstrap 5 (a powerful, extensible, and feature-packed fron-tend toolkit; https://getbootstrap.com), HTML5, CSS, and JavaScript. MySQL (http://www.mysql.org) has been adopted as the relational database system for information storage. All code is developed using Visual Studio (https://visualstudio.microsoft.com), a powerful and versatile integrated development environment (IDE) for developers and teams. PSIA is hosted on the Elastic Cloud Server (ECS) of HUAWEI CLOUD (https://activity.huaweicloud.com) to provide stable web services. The website has been tested to ensure function-ality across various operating systems and web browsers, in-cluding Google Chrome, Firefox, and Microsoft Edge.

Case study

PSIA integrates genome assemblies from species with well-characterized SI molecular mechanisms, enabling systematic identification and analysis of *S* genes. To demonstrate its util-ity in exploring *S*-locus information through high-quality ge-nomic assemblies, we present a case study of *S. lycopersicum*, a representative species of the Solanaceae family. The PSIA interface provides access to SI type classifications and *S* gene information for Solanaceae via the "Browse" section. Users can navigate to the "Species" section for a direct link to *S. lycopersicum* or utilize a fuzzy search (*e.g.*, inputting "tomato") to retrieve relevant entries, including wild tomato species (Figure S11A).
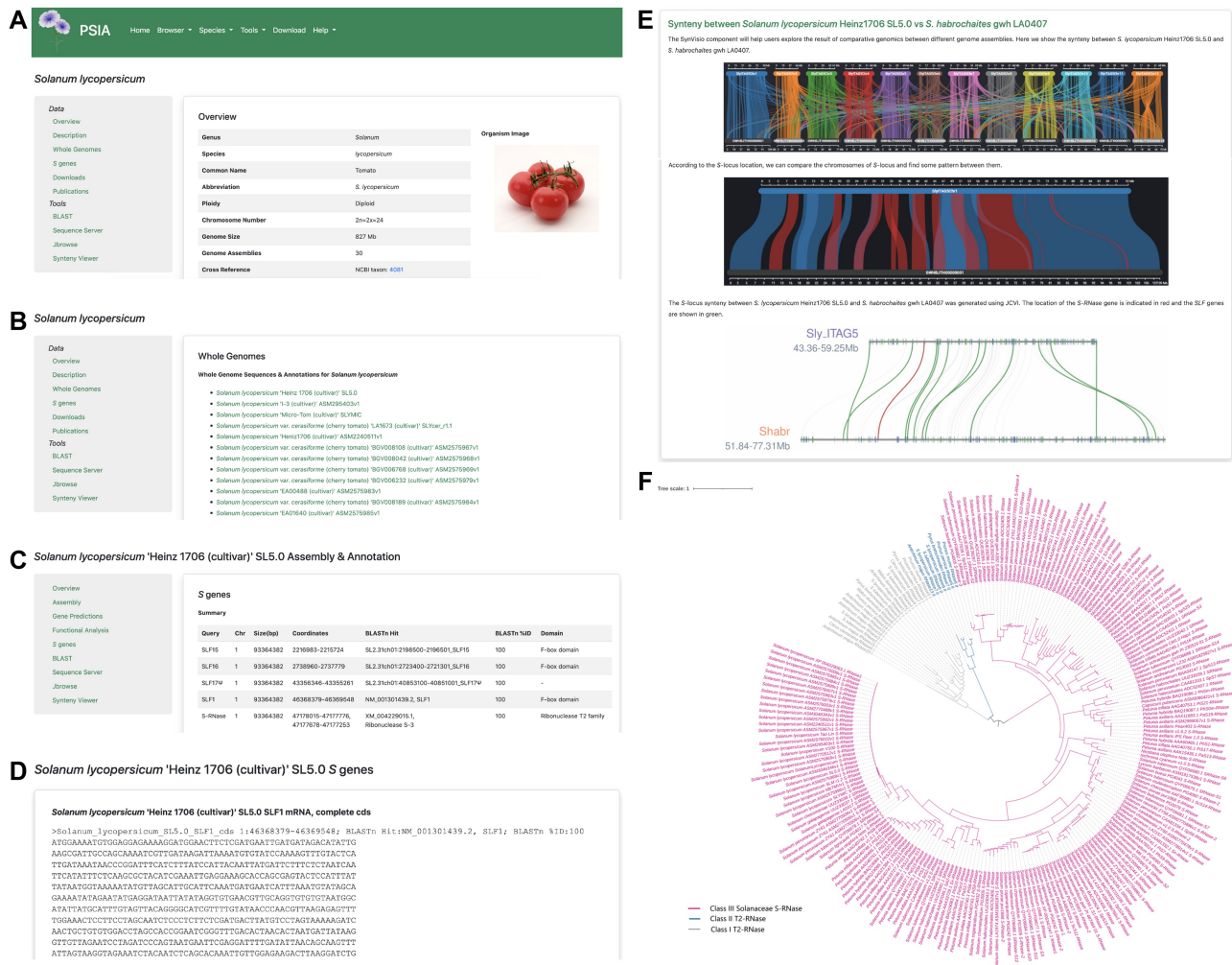
Figure 2 SI information and analysis of *Solanum lycopersicum*
A. Species page of *S. lycopersicum*. **B**. Genome assembly list of *S. lycopersicum*. **C**. Summary of *S* genes in *S. lycopersicum* Heinz 1706 SL5.0 assembly. **D**. Sequence of the *SLF1* gene in the *S. lycopersicum* Heinz 1706 SL5.0 assembly. **E**. Synteny analysis output between *S. lycopersicum* Heinz 1706 SL5.0 and *S. habrochaites* gwh LA0407. **F**. Phylogenetic analysis output of S-RNases in Solanaceae.

The *S. lycopersicum* interface comprises two main sections: "Data" and "Tools". The "Data" section includes: (1) Overview, summarizing species-specific details (Figure 2A); (2) Description, providing taxonomic and biological information; (3) Whole Genomes, listing 30 available genome assemblies (Figure 2B); (4) *S* genes, summarizing the annotation of identified *S* genes across these assemblies; (5) Downloads, offering *S* gene sequences from the 30 genomes; and (6) Publications, compiling relevant genomic literature. For example, selecting "Heinz 1706 (cultivar) SL5.0 assembly" under "Whole Genomes" redirects users to a dedicated page featuring: (1) Overview, summarizing assembly statistics; (2) Assembly, providing genome download links; (3) Gene Predictions, detailing structural annotations; (4) Functional Analysis, presenting gene function annotations; and (5) *S* genes, displaying sequence information specific to the selected assembly (Figure 2C and D).

PSIA's tools enable in-depth analysis of *S*-loci and *S* genes in tomatoes. The genome-wide BLAST enables homology searches against local databases of *S. lycopersicum* genomes, coding sequences (CDS), and proteins. This eliminates the need for users to manually download genome data and execute BLAST+ commands via Linux. Instead, users can input a query sequence, select a species-specific database, and utilize optimized default parameters for homology searches (Figure S12). The integrated SequenceServer hosts a comprehensive repository of *S. lycopersicum* *SLF* and *S-RNase* genes, significantly enhancing the accuracy and efficiency of *S* gene identification in newly assembled genomes of tomatoes (Figure S13). To further explore *S*-locus architecture, PSIA integrates JBrowse for visualizing *S*-locus genomic features and annotated *S* gene structures. For instance, using the *S* gene annotations identified in *S. lycopersicum* "Heinz 1706 (cultivar)" SL5.0 assembly (Figure 2C), users can visualize the complete gene structure and annotation details of *S-RNase* by entering genomic coordinates 1:47178015-47177253 in JBrowse (Figure S14). Comparative synteny analysis of the *S*-loci between cultivated tomato (*S. lycopersicum*) and wild tomato (*Solanum habrochaites*) revealed conserved collinearity as well as structural variations (Figure 2E, Figure S15). Notably, wild tomatoes retain functional SI systems that promote outcrossing and serve as reservoirs of novel alleles. Comprehensive characterization of *S* genes in wild and cultivated tomato species enables precise manipulation of cross-pollination barriers, thereby enhancing hybrid breeding efficiency. Extended synteny analysis across

Solanum species (Figure S16) further elucidated *S*-locus diversification within the genus. To investigate the evolutionary trajectory of *S* genes within Solanaceae, we performed phylogenetic analyses of *S-RNase* and *SLF* genes across Solanaceae species (Figure 2F, Figure S18) and extended our analysis to four families exhibiting type-1 SI (Figures S19 and S20). These phylogenetic insights contribute to a broader understanding of type-1 SI systems.

Discussion and conclusion

Research on angiosperm SI dates back to the 19th century. While substantial progress has been made, the molecular mechanisms underlying this complex reproductive system remain incompletely understood. SI research holds both fundamental biological significance and practical importance, particularly in crop breeding, quality improvement, and fruit production [4]. The availability of high-quality plant genomes has greatly facilitated the identification and analysis of *S* genes, especially in the most widespread type-1 SI system. Type-1 *S*-loci typically span large genomic regions containing a female determinant (*S-RNase*) and approximately 15–32 male determinants (*SLF/SFB/SFBB* genes) [29,44–46], necessitating chromosome-scale assemblies for comprehensive structural resolution. However, existing plant genomic databases largely overlook *S* gene annotation and analysis. For example, the Sol Genomics Network primarily focuses on limited Solanaceae species, emphasizing gene expression, quantitative trait loci (QTLs), and genetic markers [34]. To date, no database has systematically integrated available genome assemblies to comprehensively annotate and analyze *S* genes across taxa, representing a critical gap that hinders SI research.

The PSIA knowledgebase aims to provide the most comprehensive online platform for plant SI research by integrating genomic data, *S*-locus information, *S* gene sequences, and analytical tools. For researchers unfamiliar with SI systems, PSIA offers a systematic introduction to SI classification, origin, evolution, and experimentally derived working models. For SI specialists, the platform features a locally implemented BLAST service via SequenceServer, enabling rapid and accurate identification of *S* genes in newly assembled genomes using the most extensive collection of *S* gene sequences. Additionally, bioinformatics users can download *S* gene sequences for customized analyses. PSIA's data integration spans angiosperm-wide comparisons, facilitating macroevolutionary investigations of SI dynamics. Notably, the four families exhibiting type-1 SI include numerous crops, fruits, and ornamental species. A comprehensive understanding of the molecular mechanisms and evolutionary trajectory of type-1 SI requires the systematic integration of *S* gene data from these four families. Moreover, the platform's Synteny Viewer supports both interspecies (*e.g.*, cultivated *vs.* wild relatives) and genus-level comparisons, providing insights into the dynamic evolution of SI. PSIA's Phylogenetic Analysis module incorporates the most extensive *S* gene sequence collection to date, significantly advancing evolutionary studies of these loci.

PSIA's comprehensive annotation and analysis of *S* genes can enhance hybrid breeding in tomatoes. Wild tomato (*S. habrochaites*) possesses valuable traits, such as disease resistance, stress tolerance, and improved fruit quality, which breeders aim to transfer into cultivated tomato (*S.*

lycopersicum). Interspecific hybridization between cultivated (self-compatible) and wild (self-incompatible) tomatoes requires a comprehensive understanding of *S* gene function to either maintain SI for breeding diversity or overcome SI for successful hybrid production. Manipulating *S* gene expression offers a viable approach to overcoming reproductive barriers and enabling gene flow between wild and cultivated species. For example, disrupting or silencing *S* genes using CRISPR-Cas9 or RNA interference can restore self-compatibility in hybrids, thereby improving seed production efficiency. Additionally, synteny analysis between cultivated and wild tomatoes enables researchers to track evolutionary changes in reproductive mechanisms, providing insights into how domestication has influenced SI systems in tomatoes.

PSIA has significant potential for further advancements, and we plan to continuously update and expand its functionality. For instance, the database will periodically incorporate genomic and *S*-locus information from representative species across the 11 families with well-characterized molecular mechanisms. Furthermore, as novel molecular mechanisms of SI are discovered in other plant families, PSIA will systematically integrate these new findings, providing a continuously evolving and comprehensive knowledgebase. Given the tissue-specific expression patterns of pistil and pollen *S* genes, future versions will integrate relevant transcriptome data to enable *S* gene expression analysis. Additional functionalities, such as primer-design tools, will be implemented based on user needs, broadening the platform's applicability in functional genomics and breeding research. These planned updates will ensure that PSIA remains a comprehensive and up-to-date resource for SI research.

In conclusion, PSIA represents a critical advancement in plant SI research, leveraging high-quality genome assemblies to systematically explore *S*-locus information and addressing the current limitations in SI-related genomic databases. Looking ahead, regular updates incorporating newly sequenced genomes and the latest insights into SI molecular mechanisms will enhance its utility as a continuously evolving resource. By providing a comprehensive, user-friendly platform, PSIA will facilitate novel discoveries in plant reproductive biology, improve our understanding of *S*-locus evolution, and support applied research in crop breeding and quality improvement.

Data availability

PSIA is freely available at http://www.plantsi.cn. PSIA has been submitted to Database Commons [51] at the National Genomics Data Center (NGDC), China National Center for Bioinformation (CNCB), which is publicly accessible at https://ngdc.cncb.ac.cn/databasecommons/database/id/8169.

CRediT author statement

Writing – review & editing. All authors have read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Supplementary material

Acknowledgments

ORCID

0000-0003-4430-5378 (Chen Wang)
0000-0003-1137-8028 (Hong Zhao)
0000-0002-4238-5530 (Hongkui Zhang)
0000-0002-4786-6819 (Sijie Sun)
0000-0002-6895-8472 (Yongbiao Xue)

References

[1] Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, et al. Phylogenomics and the rise of the angiosperms. Nature 2024;629:843–50.

[2] Igic B, Lande R, Kohn JR. Loss of self-incompatibility and its evolutionary consequences. Int J Plant Sci 2008;169:93–104.

[3] Fujii S, Kubo K, Takayama S. Non-self- and self-recognition models in plant self-incompatibility. Nat Plants 2016;2:16130.

[4] Muñoz-Sanz JV, Zuriaga E, Cruz-García F, McClure B, Romero C. Self-(in)compatibility systems: target traits for crop-production, plant breeding, and biotechnology. Front Plant Sci 2020;11:195.

[5] Takayama S, Isogai A. Self-incompatibility in plants. Annu Rev Plant Biol 2005;56:467–89.

[6] Yang B, Thorogood D, Armstead I, Barth S. How far are we from unravelling self-incompatibility in grasses? New Phytol 2008; 178:740–53.

[7] De Nettancourt D. Incompatibility and incongruity in wild and cultivated plants. Berlin Heidelberg: Springer; 2001.

[8] Franklin-Tong VE. Self-incompatibility in flowering plants: evolution, diversity, and mechanisms. Berlin Heidelberg: Springer; 2008.

[9] Zhang Y, Zhao Z, Xue Y. Roles of proteolysis in plant self-incompatibility. Annu Rev Plant Biol 2009;60:21–42.

[10] Schopfer CR, Nasrallah ME, Nasrallah JB. The male determinant of self-incompatibility in *Brassica*. Science 1999;286:1697–700.

[11] Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K. The *S* receptor kinase determines self-incompatibility in *Brassica* stigma. Nature 2000;403:913–6.

[12] Bala M, Rehana S, Singh MP. Self-incompatibility: a targeted, unexplored pre-fertilization barrier in flower crops of Asteraceae. J Plant Res 2023;136:587–612.

[13] Kowyama Y, Tsuchiya T, Kakeda K. Sporophytic self-incompatibility in *Ipomoea trifida*, a close relative of sweet potato. Ann Bot 2000;85:191–6.

[14] Hill RJ, Baldassi C, Snelling JW, Vining KJ, Mehlenbacher SA. Fine mapping of the locus controlling self-incompatibility in European hazelnut. Tree Genet Genomes 2021;17:1–14.

[15] Li J, Cocker JM, Wright J, Webster MA, McMullan M, Dyer S, et al. Genetic architecture and evolution of the *S* locus supergene in *Primula vulgaris*. Nat Plants 2016;2:16188.

[16] Shore JS, Hamam HJ, Chafe PDJ, Labonne JDJ, Henning PM, McCubbin AG. The long and short of the *S*-locus in *Turnera* (Passifloraceae). New Phytol 2019;224:1316–29.

[17] Gutiérrez-Valencia J, Fracassetti M, Berdan EL, Bunikis I, Soler L, Dainat J, et al. Genomic analyses of the *Linum* distyly supergene reveal convergent evolution at the molecular level. Curr Biol 2022;32:4360–71.

[18] Castric V, Batista RA, Carré A, Mousavi S, Mazoyer C, Godé C, et al. The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene. Curr Biol 2024;34:1967–76.

[19] Xue Y, Carpenter R, Dickinson HG, Coen ES. Origin of allelic diversity in antirrhinum *S* locus RNases. Plant Cell 1996;8:805–14.

[20] Lai Z, Ma W, Han B, Liang L, Zhang Y, Hong G, et al. An F-box gene linked to the self-incompatibility (*S*) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. Plant Mol Biol 2002;50:29–42.

[21] Qiao H, Wang F, Zhao L, Zhou J, Lai Z, Zhang Y, et al. The F-box protein AhSLF-S_2 controls the pollen function of S-RNase-based self-incompatibility. Plant Cell 2004;16:2307–22.

[22] Li M, Zhang D, Gao Q, Luo Y, Zhang H, Ma B, et al. Genome structure and evolution of *Antirrhinum majus* L. Nat Plants 2019;5:174–83.

[23] Anderson MA, Cornish E, Mau SL, Williams E, Hoggart R, Atkinson A, et al. Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. Nature 1986;321:38–44.

[24] McClure BA, Haring V, Ebert PR, Anderson MA, Simpson RJ, Sakiyama F, et al. Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. Nature 1989;342:955–7.

[25] Sijacic P, Wang X, Skirpan AL, Wang Y, Dowd PE, McCubbin AG, et al. Identification of the pollen determinant of S-RNase-mediated self-incompatibility. Nature 2004;429:302–5.

[26] Sassa H, Nishio T, Kowyama Y, Hirano H, Koba T, Ikehashi H. Self-incompatibility (*S*) alleles of the Rosaceae encode members of a distinct class of the T2/S ribonuclease superfamily. Mol Gen Genet 1996;250:547–57.

[27] Ushijima K, Sassa H, Dandekar AM, Gradziel TM, Tao R, Hirano H. Structural and transcriptional analysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. Plant Cell 2003;15:771–81.

[28] Sassa H, Kakui H, Miyamoto M, Suzuki Y, Hanada T, Ushijima K, 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 2007;175:1869–81.

[29] Liang M, Cao Z, Zhu A, Liu Y, Tao M, Yang H, et al. Evolution of self-compatibility by a mutant S_m-*RNase* in citrus. Nat Plants 2020;6:131–42.

[30] 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;91:2265–9.

[31] Wheeler MJ, de Graaf BH, Hadjiosif N, Perry RM, Poulter NS, Osman K, et al. Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. Nature 2009;459:992–5.

[32] Rohner M, Manzanares C, Yates S, Thorogood D, Copetti D, Lübberstedt T, et al. Fine-mapping and comparative genomic analysis reveal the gene composition at the *S* and *Z* self-incompatibility loci in grasses. Mol Biol Evol 2023;40:msac259.

[33] Wang Y, Zhao H, Zhu S, Zhang H, Chen Y, Tao X, et al. Control of gametophytic self-incompatibility in the African wild rice. Research Square 2022; https://doi.org/10.21203/rs.3.rs-2121145/v1.

[34] Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, Strickler SR, et al. The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. Nucleic Acids Res 2015;43:D1036–41.

[35] Jung S, Lee T, Cheng CH, Buble K, Zheng P, Yu J, et al. 15 years of GDR: new data and functionality in the genome database for Rosaceae. Nucleic Acids Res 2019;47:D1137–45.

[36] Ma Y, Zhao X, Jia Y, Han Z, Yu C, Fan Z, et al. The updated Genome Warehouse: enhancing data value, security, and usability to address data expansion. Genomics Proteomics Bioinformatics 2025;23:qzaf010.

[37] Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 2012;40:D1178–86.

[38] Zhou Y, Qiao Y, Ni Z, Du J, Xiong J, Cheng Z, et al. GDS: a genomic database for strawberries (*Fragaria* spp.). Horticulturae 2022;8:41.

[39] Chen H, Wang T, He X, Cai X, Lin R, Liang J, et al. BRAD V3.0: an upgraded Brassicaceae database. Nucleic Acids Res 2022; 50:D1432–41.

[40] Zhao H, Zhang Y, Zhang H, Song Y, Zhao F, Zhang Y, et al. Origin, loss, and regain of self-incompatibility in angiosperms. Plant Cell 2022;34:579–96.

[41] Zhao H, Song Y, Li J, Zhang Y, Huang H, Li Q, et al. Primary restriction of S-RNase cytotoxicity by a stepwise ubiquitination and degradation pathway in *Petunia hybrida*. New Phytol 2021; 231:1249–64.

[42] Tian H, Zhang H, Huang H, Zhang Y, Xue Y. Phase separation of S-RNase promotes self-incompatibility in *Petunia hybrida*. J Integr Plant Biol 2024;66:986–1006.

[43] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10.

[44] Li W, Chetelat RT. Unilateral incompatibility gene *ui1.1* encodes an *S*-locus F-box protein expressed in pollen of *Solanum* species. Proc Natl Acad Sci U S A 2015;112:4417–22.

[45] Zhu S, Zhang Y, Copsy L, Han Q, Zheng D, Coen E, et al. The snapdragon genomes reveal the evolutionary dynamics of the *S*-locus supergene. Mol Biol Evol 2023;40:msad080.

[46] Gu C, Xu Y, Wu L, Wang X, Qi K, Qiao X, et al. Long-read genome sequencing reveals the sequence characteristics of pear self-incompatibility locus. Mol Hortic 2025;5:13.

[47] Deng W, Nickle DC, Learn GH, Maust B, Mullins JI. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. Bioinformatics 2007;23:2334–6.

[48] Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, et al. Sequenceserver: a modern graphical user interface for custom BLAST databases. Mol Biol Evol 2019;36:2922–4.

[49] Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, et al. JBrowse 2: a modular genome browser with views of synteny and structural variation. Genome Biol 2023;24:74.

[50] Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 2020;37:1530–4.

[51] Ma L, Zou D, Liu L, Shireen H, Abbasi AA, Bateman A, et al. Database Commons: a catalog of worldwide biological databases. Genomics Proteomics Bioinformatics 2023;21:1054–8.