



OPEN ACCESS

EDITED BY

Liang-Zi Zhou,
University of Regensburg, Germany

REVIEWED BY

Odd Arne Rognli,
Norwegian University of Life Sciences,
Norway
Petr Smykal,
Palacký University in Olomouc,
Czechia

*CORRESPONDENCE

Rowan Herridge
rowan.herridge@otago.ac.nz

SPECIALTY SECTION

This article was submitted to
Plant Development and EvoDevo,
a section of the journal
Frontiers in Plant Science

RECEIVED 04 August 2022

ACCEPTED 27 September 2022

PUBLISHED 18 October 2022

CITATION

Herridge R, McCourt T, Jacobs JME,
Mace P, Brownfield L and Macknight R
(2022) Identification of the genes at S
and Z reveals the molecular basis and
evolution of grass self-incompatibility.
Front. Plant Sci. 13:1011299.
doi: 10.3389/fpls.2022.1011299

COPYRIGHT

© 2022 Herridge, McCourt, Jacobs,
Mace, Brownfield and Macknight. This is
an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use,
distribution or reproduction is
permitted which does not comply with
these terms.

Identification of the genes at S and Z reveals the molecular basis and evolution of grass self-incompatibility

Rowan Herridge^{1*}, Tyler McCourt¹, Jeanne M. E. Jacobs²,
Peter Mace¹, Lynette Brownfield¹ and Richard Macknight¹

¹Department of Biochemistry, University of Otago, Dunedin, New Zealand, ²Forage Science, AgResearch, Christchurch, New Zealand

Self-incompatibility (SI) is a feature of many flowering plants, whereby self-pollen is recognized and rejected by the stigma. In grasses (Poaceae), the genes controlling this phenomenon have not been fully elucidated. Grasses have a unique two-locus system, in which two independent genetic loci (S and Z) control self-recognition. S and Z are thought to have arisen from an ancient duplication, common to all grasses. With new chromosome-scale genome data, we examined the genes present at S- and Z-loci, firstly in ryegrass (*Lolium perenne*), and subsequently in ~20 other grass species. We found that two DUF247 genes and a short unstructured protein (SP/ZP) were present at both S- and Z- in all SI species, while in self-compatible species these genes were often lost or mutated. Expression data suggested that DUF247 genes acted as the male components and SP/ZP were the female components. Consistent with their role in distinguishing self- from non-self, all genes were hypervariable, although key secondary structure features were conserved, including the predicted N-terminal cleavage site of SP/ZP. The evolutionary history of these genes was probed, revealing that specificity groups at the Z-locus arose before the advent of various grass subfamilies/species, while specificity groups at the S-locus arose after the split of Panicoideae, Chloridoideae, Oryzoideae and Pooideae. Finally, we propose a model explaining how the proteins encoded at the S and Z loci might function to specify self-incompatibility.

KEYWORDS

grass, self-incompatibility, DUF247, pollen, reproduction, Poaceae, stigma

Introduction

Self-incompatibility (SI) is a mechanism by which plants are able to distinguish, and reject fertilization of self pollen. SI is mediated by hypervariable genes, expressed in male tissue (anther/tapetum/pollen) and female tissue (pistil/stigma) which interact at the pollen/stigma interface. The genes underlying SI have been described in Solanaceae, Brassicaceae and Papaveraceae, revealing different mechanisms in these families (Foote et al., 1994; Schopfer et al., 1999; Kachroo et al., 2001; Kao and Tsukamoto, 2004). Despite the agronomic importance of the grass family (Poaceae), the genes underlying SI have not been elucidated. Self-fertility is an important domestication trait, and the ability to inbreed plants is highly desirable for F1 hybrid breeding (Glemin and Bataillon, 2009; Herridge et al., 2019).

Grasses have a unique SI system, which involves two genetic loci (S and Z), both of which must be matched in pollen/stigma in order to prevent self-fertilization; however, the underlying molecular mechanism remains unresolved. This system acts gametophytically, which means that the genotype of the haploid pollen contributes to SI response, and SI determinants are produced by the pollen grain itself. This is in contrast to the *Brassica* system, which acts sporophytically, and the SI determinants are deposited on the surface of the pollen grain by the diploid tapetum, preventing fertilization based on the genotype of the paternal (sporophytic) parent, regardless of which allele the pollen grain inherits (Hiscock and McInnis, 2003). The grass system also acts by a self-recognition system, whereby interaction of self pollen on the stigmatic exudate quickly leads to pollen tube arrest (Shivanna et al., 1982). This is in contrast to a non-self-recognition system, as is seen in the Solanaceae family, whereby non-self SI determinants (stigmatic S-RNase in the case of Solanaceae) are recognized by pollen-expressed F-box proteins leading to their degradation, while self S-RNase is not recognized and results in pollen tube arrest (Fujii et al., 2016). Evidence for this non-self mechanism comes from tetraploid grasses, in which SI is maintained, and requires only that one S- and one Z-locus is matched in the stigma (by contrast, tetraploid Solanaceae are self-compatible; Stout and Chandler, 1941; Fearon et al., 1984; Baumann, 2000). Therefore, a positive S/Z interaction is essential to grass SI, and cannot be complimented by a dual S (S/S) or Z (Z/Z) interaction. This suggests that the underlying genes have distinct functions in the downstream signaling pathway that must be integrated to result in pollen tube arrest. The two locus system is thought to have arisen from a duplication occurring in the common ancestor of grasses, although how a one-locus system would operate given the necessity of S/Z integration in modern grasses is an open question (Lundqvist, 1962).

Mapping of the S- and Z-loci in grasses has been performed in a number of species (Yang et al., 2008). In ryegrass (*Lolium perenne*), a DUF247 protein co-segregated with the S-locus, and

had variable sequence between different specificity groups (plants which are capable of crossing with one another, but not themselves; Manzanares et al., 2016). Importantly, this gene was predicted to be non-functional in a self-compatible relative of ryegrass, *L. temulentum* (Manzanares et al., 2016). In addition, no other hypervariable genes were found in the predicted region based on the draft genome sequence developed by Byrne et al. (2015). Similarly, a DUF247 gene was located at the Z-locus in *L. perenne*, although variability of this gene was not probed (Shinozuka et al., 2010; Thorogood et al., 2017). In a SI rice species, *Oryza longistaminata*, two DUF247 genes, and a small pistil-expressed protein (named SP) was present at the S-locus (Lian et al., 2021). With the advent of a high-quality ryegrass genome, as well as abundant high-quality grass genomes now publicly available, we aimed to probe these regions in detail using prior knowledge from the above studies (Frei et al., 2021; Nagy et al., 2022).

Our work revealed that three genes are likely key to the function of both the S- and Z- loci; two DUF247 genes and a gene encoding a short unstructured protein (SP and ZP, respectively). We propose a model where a pollen extracellular receptor, comprising two S- and two Z- DUF247 proteins, specifically recognizes its own stigma excreted SP/ZP signal proteins, activating a pathway that prevents fertilization by its own pollen. SC grasses likely evolved through mutation or loss of these key SI components.

Results

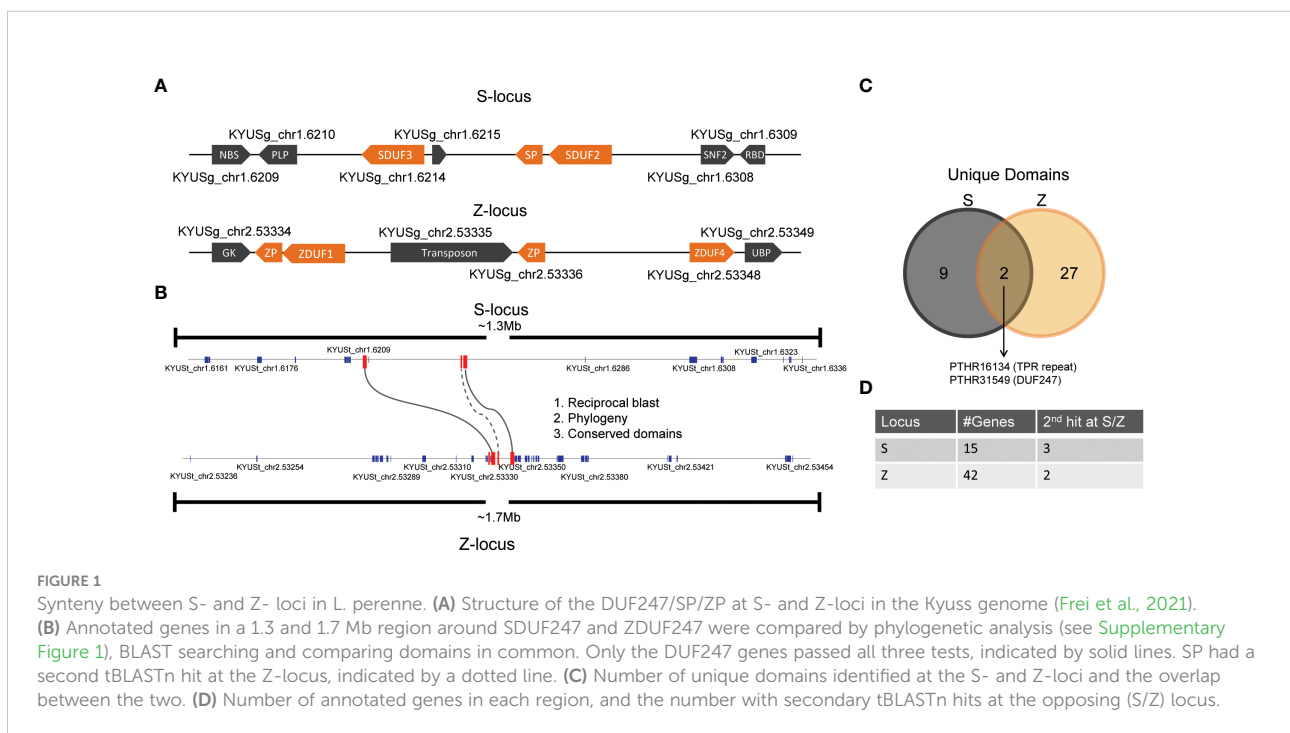
Synteny between the S- and Z- loci indicate that DUF247 and SP/ZP genes are responsible for ryegrass self-incompatibility

With the recent publication of a high-quality ryegrass genome, we set out to investigate candidate genes that might control self-incompatibility (Frei et al., 2021). DUF247 genes have been identified at both the S- and Z-loci of ryegrass as likely candidates for mediators of SI (Manzanares et al., 2016; Thorogood et al., 2017). Ryegrass contains a large number of DUF247 genes, four DUF247 genes (DUF1-4) form a small clade including DUF2, the aforementioned DUF247 associated with the S-locus (Veckman et al., 2019). Further supporting the role of the DUF247 genes, two DUF247 genes have been identified at the S-locus of the self-incompatible wild perennial rice species *O. longistaminata* (Lian et al., 2021). Alongside the DUF247 gene, a gene encoding a short protein dubbed *S-LOCUS PISTIL* (SP) was also identified as a potential candidate gene within the *O. longistaminata* S-locus, as the SP gene was highly polymorphic and expressed at high levels in the pistil tissue (Lian et al., 2021). As grass SI loci are predicted to have arisen from a duplication, we expected the nature of the genes present at the S- and Z-loci

to be similar (Lundqvist, 1962). We examined the sequence at the S- and Z-loci in the recently published high-quality ryegrass genome, using tBLASTn with DUF247 proteins from *L. perenne* and SP proteins from *O. longistaminata*, revealing two DUF247 genes at each location (Figure 1A; Frei et al., 2021). The flanking genes of SDUF247 in the Kyuss genome were the same as those identified by Manzanares et al. (2016). The DUF247 genes at S- and Z- were the small clade mentioned above, described in Veeckman et al. (2019) and named DUF1-4. We retained this nomenclature, and labeled them with their corresponding locus, yielding SDUF2, SDUF3, ZDUF1 and ZDUF4. We also identified one or two SP-like genes at each location and retained this nomenclature, naming the relevant genes “SP” and “ZP” (Figure 1A). In cases where two ZP-like genes were identified, they were named ZP1 or ZP4 by their proximity to either ZDUF1 or ZDUF4.

The leading hypothesis for the origin of the S- and Z-loci in grasses is that they evolved from a precursor, single locus (Lundqvist, 1962); using this prediction, we aimed to rule out the possibility of other genes in the S- and Z-loci controlling SI specificity. To determine if DUF247/SP/ZP were the only genes in common between S- and Z- we took a region >1Mb centered around the S- and Z-DUF247 genes (on Chromosome 1 and 2; Figure 1B), and extracted all predicted proteins from the *L. perenne* (Kyuss) genome (Frei et al., 2021). Using Interpro to predict protein domains we compared proteins from S- and Z- to identify overlapping protein types. This revealed only DUF247 genes and an FBOX/TPR-repeat gene are present at both loci (Figure 1C; Supplementary Dataset S1). Aligning the two FBOX/

TPR-repeat proteins did not show any regions of substantial alignment, and these genes were not linked in the other two tests (described below). As the SP/ZP proteins do not have predicted domains, they were not included in this analysis. To attempt to address this, and detect other proteins without annotated domains, we produced a phylogenetic tree from the entire list of proteins, with the anticipation that highly similar proteins from S- and Z- may cluster together, but with the caveat that highly polymorphic S/Z-determinants may not cluster or that the high divergence would confound the tree. Nevertheless, the phylogenetic tree produced a cluster with 4 DUF247 proteins showing the highest similarity of any cluster containing proteins from Chromosomes 1 and 2 (Supplementary Figure S1). SP/ZP did not cluster together in this analysis, likely owing to their short size and high degree of polymorphism. Finally, in the case that a protein may only be annotated at one locus, and not the other (as annotations are often produced bioinformatically, unusual genes are occasionally omitted), we performed tBLASTn using protein sequences from S- and Z- with the expectation that the top blast hit would be the native position of each protein, and the second hit would be at the most recent duplication – in the case of S- and Z-determinants this would likely be at the Z- or S- locus, respectively. As expected, the first hit of all predicted proteins was at their native location, and only in the case of the four DUF247 proteins and the SP protein was the second hit at the S- or Z-locus (Figure 1D; Supplemental Dataset S2). The fact that no other genes were in common between S- and Z-loci provides evidence that the four DUF247 genes and the SP/ZP genes were responsible for SI determination.



The expression of the DUF247 and SP/ZP genes in reproductive tissues is consistent with the role in determining self-incompatibility

To mediate specificity, the DUF247 and SP/ZP genes must be expressed in the pollen or stigmatic tissues. Previous experiments from ryegrass indicated that SDUF2 is present in pollen, although with some stigmatic expression (Manzanares et al., 2016). Analysis of SDUF2 (named *OISS2*) and SDUF3 (*OISS1*) and SP from *O. longistaminata* suggested SDUF3 was pollen-specific, SDUF2 expressed in both pollen and pistil, and SP strongly pistil-specific (Lian et al., 2021). One limitation of the prior RNA-seq data analysis in ryegrass was the absence of an appropriate reference genome for identifying transcript abundance from highly variable genes. “SP” for example, was omitted from the analysis performed by Manzanares et al. (2016) likely owing to the inability to map it to the reference genome, despite its presence in the draft genome and RNA-seq data from this study (Byrne et al., 2015). We performed RNA-seq on developing anthers and stigma from a heterozygous *L. perenne* individual from the ONE50 cultivar with a draft genome assembly (referred to as “ONE50a” henceforth; J.M.E. Jacobs unpublished data). Firstly, we aimed to identify well-assembled S- and Z- alleles using BLAST with DUF1-4 as queries. Two S- and Z- alleles were found, as expected of a heterozygous individual, however, only one of each was well-assembled and selected for further analysis (on scaffold_1841 and scaffold_1554). We also screened other ONE50 plants in our collection for the presence of S- and Z-loci that matched this genome using locus-specific primers to the DUF/SP/ZP genes, identifying an individual matching the Z-locus on scaffold_1554.

Consistent with prior results, SDUF2 was primarily anther-expressed, but with some stigmatic expression, while SDUF3 was more anther-biased (Figure 2A). SP was exceptionally highly expressed in stigma (Figure 2A). At the Z-locus ZDUF1 was anther-specific, while ZDUF4 was expressed in both tissues (Figure 2B). ZP was strongly expressed in stigma (Figure 2B). A second putative “ZP” locus was not expressed, suggesting that only one ZP is required for SI, and the presence of a second ZP may be a relic of the single-locus system. As mentioned above, only DUF/SP/ZP sequences that matched the reference ONE50 genome were able to be aligned, evidenced by the second sequenced plant (labeled “ONE50b”) only showing reads aligning to the Z-locus on scaffold_1554, and not the S-locus on scaffold_1841 (Figure 2A and B). To assay the abundance of unmapped DUF/SP/ZP transcripts in both individuals, we assembled transcripts *de novo* from this RNA seq data. In both plants (ONE50a and ONE50b), four SP/ZP and eight S/ZDUF transcripts were assembled; however, in two cases, the ZDUF transcript, and associated ZP transcript were assembled into a single contig

(Supplementary Figure 2). In these instances, ZP expression was likely the predominant factor influencing the expression value (reflected in high stigma expression for these transcripts (Supplementary Figures 2B, D)). Nevertheless, the expression pattern for these transcripts matched the genome-guided approach above, with DUF247 genes expressed in anthers, and SP/ZP highly expressed in stigma (Supplementary Figure 2). Combined, transcripts from SP and ZP made up a large proportion of the transcriptome in our datasets (~26,000 TPM and ~10,000 TPM in each stigma dataset).

To determine if the DUF247 gene is expressed in the male and SP/ZP genes are expressed in the female reproductive tissues in other grasses, we examined expression results using available data from other grass species. Anther and stigma/pistil transcriptomic data was accessible for *Brachypodium distachyon*, *O. sativa*, *Triticum aestivum* and *Sorghum bicolor* (Davidson et al., 2012; Ma et al., 2021). We identified SP/ZP and DUF247 genes from these species using tBLASTn querying the genomic sequence of each species. In all species, S/ZDUF247 and SP/ZP were identified in close proximity, similar to *L. perenne*, indicating that we had identified the correct loci. Some of the genes were annotated (Supplementary Table 1), allowing us to access the available transcriptomic data for these species. Expression patterns of these genes in all species were broadly consistent with those seen in *L. perenne* – modest anther-specific expression of DUF247 and high stigma/pistil expression of SP/ZP (Figures 2C and D). In summary, the expression data from ryegrass and a range of other grasses is consistent with the DUF247 genes function as the anther component and the SP/ZP genes function as the stigma component of the SI mechanism.

Disruption of the DUF247 and SP/ZP genes in self-compatibility highlights their importance for SI

Following the identification of DUF247/SP/ZP in the above species, we next searched available high-quality genomes for the syntenic regions from a large variety of SI and SC species. Should these genes be responsible for SI, they may be disrupted in SC species, either as the cause of SC, or alternatively through genetic drift upon loss of SI at loci outside S and Z. The genetic arrangement of two DUF247 and one or two SP/ZP genes was present in most species examined, allowing us to easily identify S- and Z-loci in these genomes (Figures 3, 4). The flanking genes at the Z-locus were conserved in all species (A glycerol kinase (GK) and ubiquitin-conjugating enzyme (UBC); Figure 1A, Figure 4). The S-locus, however, appears to have different flanking genes in the Pooideae, Oryzoideae, Chloridoideae and Panicoideae, suggesting that it may have translocated in these clades, or alternatively, arisen independently (Figure 3).

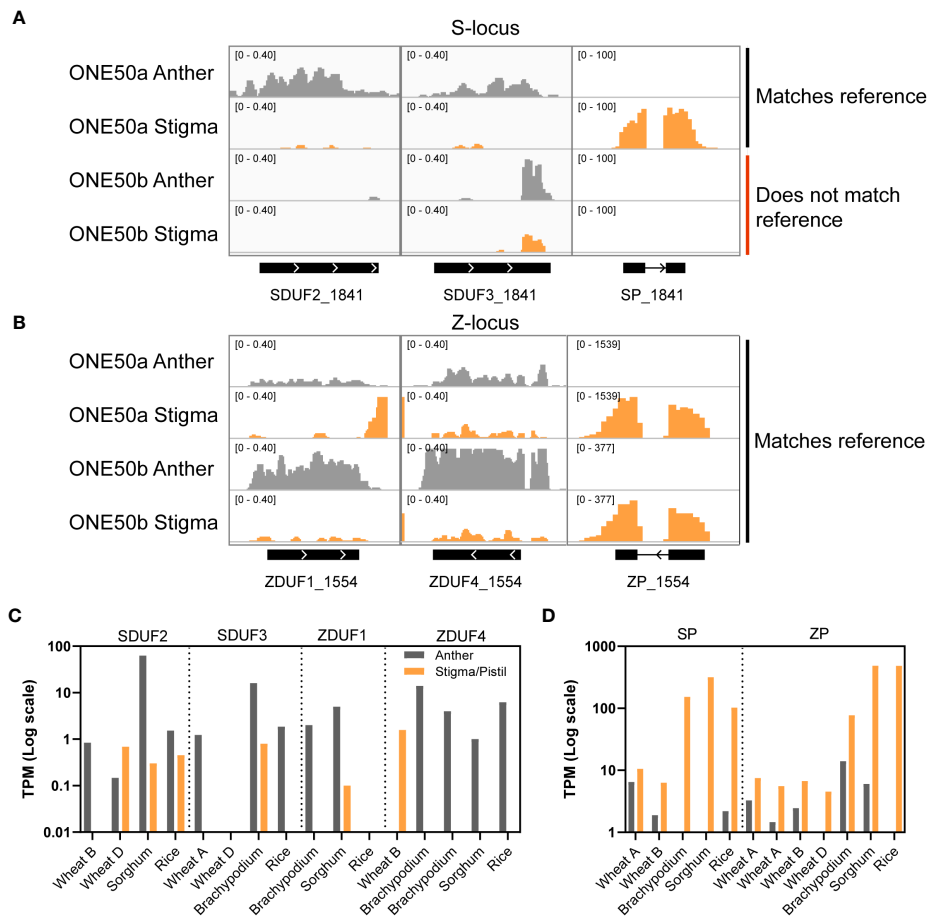


FIGURE 2 Expression of DUF247 and SP/ZP in reproductive tissues from grasses. **(A, B)** RNAseq coverage at the S- **(A)** and Z-locus **(B)** in anther and stigma tissue from two ONE50 individuals. One plant (ONE50a) was a clone of the individual used to create the reference sequence, and had the corresponding genotype at S- and Z-, while the other (ONE50b) only had the corresponding Z-locus genotype. Numbers in the top left of each sub-panel indicate the y-axis scale in units of coverage in reads per million over 25bp windows. **(C, D)** Expression of DUF247 **(C)** and SP/ZP **(D)** genes from rice, wheat, sorghum and brachypodium in anther and stigma (or pistil) from expression databases.

Rearrangements were more common at the Z-locus than the S-locus, with the direction of DUF247 genes changing in relation to flanking genes. The size of the region also varied greatly, often correlating with genome size, from 16 kb in the *Oropetium thomaeum* Z-locus to >3 Mb at the *Panicum virgatum* S-locus (Figures 3, 4). Importantly, all SI species examined contained functional DUF247 and SP/ZP at each locus, while all SC species had frame-shift mutations, or were missing elements (Table 1). One exception was the Weining *Secale cereale* genome, which contained functional versions of all genes – it is possible that a locus outside S- and Z- is responsible for SC in this genotype (Voylov et al., 1998). In summary, disruption of the DUF247 and SP/ZP genes is a common mechanism for grasses to acquire self-compatibility and further supports the role of these genes in specifying SI.

Protein structure predictions suggest that SI involves DUF247 proteins on the pollen tube membrane interacting with SP/ZP proteins secreted by the stigma.

We predicted that these genes would be hypervariable if they were specificity determinants, and that at least one component would be extracellular. Previous studies have shown variability in SDUF2, SDUF3 and SP (Manzanares et al., 2016; Lian et al., 2021). Indeed, alignments of DUF247 from available genomes (reconstructed in cases of frame shifting) showed hypervariability in two segments of each DUF247 member with limited variation elsewhere (Supplementary Figure 3). We used AlphaFold (Deepmind) software to produce a structure of *L. perenne* SDUF2, and subsequently mapped the variability of *L.*

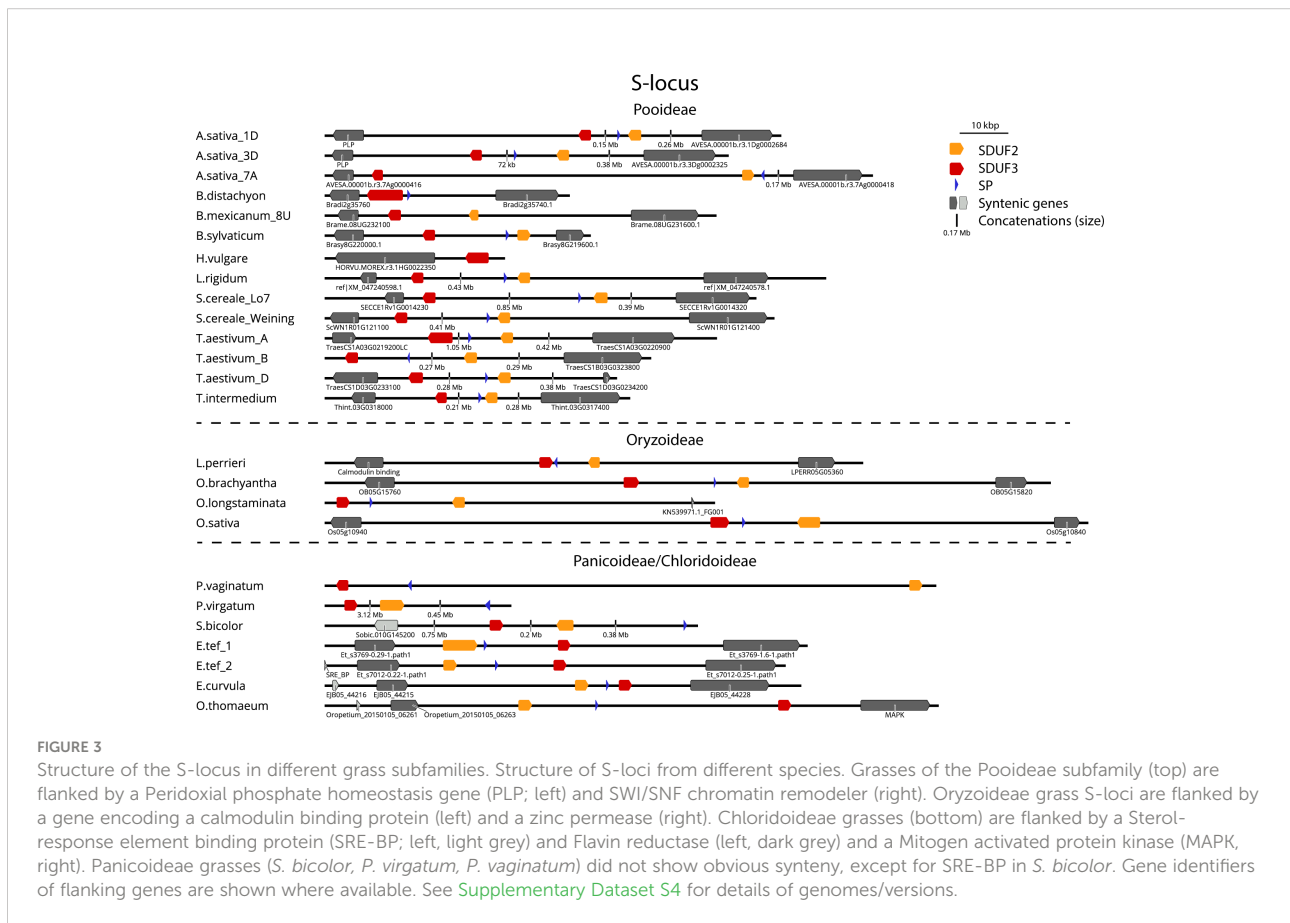
perenne SDUF2 sequence onto these structures (Veeckman et al., 2019; Mirdita et al., 2022). The two hypervariable regions of SDUF2 formed relatively unstructured loops (Supplementary Figure 4). These hypervariable unstructured loops may be involved in SP recognition, and/or oligomerization depending on the presence of other components of the SI system (e.g., SP or other DUF247 proteins).

SP and ZP were almost entirely hypervariable except for a few key residues including a C-terminal Cysteine, and a second cysteine 4-10AA away from the C-terminal (Supplementary Figure 5). Proline was common towards the C-terminal, as well as a conserved “EEK” motif in ZP proteins (Supplementary Figure 5). Small cysteine-rich peptides are common in SI signaling in Papaver and Brassica, likely due to their diffusible nature and ability to be transmitted easily through cell walls (Walker et al., 1996; Schopfer et al., 1999). While the primary sequences of these proteins were highly variable, some key features of secondary structure and functionality were conserved in both DUF247 and SP/ZP proteins, respectively. In the case of DUF247, the transmembrane domain at the C-terminal was well-conserved, suggesting that membrane localization is an essential component of its function, as previously proposed (Manzanares et al., 2016). In the case of SP/ZP, secondary structure prediction suggested that the N-termini typically contain elements of secondary structure (usually an alpha

helix, although occasionally beta-sheet for a subset of ZP proteins; Supplementary Figures 6A, B). Signal peptide prediction software indicated that all SP and ZP were cleaved at the N-terminus, resulting in a short peptide (~65AA) which was expected to be extracellular (Supplementary Figures 5C–F; Supplementary Dataset S3). The only exception was ZP proteins from the PACMAD clade, where the predicted cleavage site did not pass statistical significance (Supplementary Dataset S3). While the cleavage site for ZP proteins occurred slightly closer to the N terminal, the resulting cleaved peptide was of similar length to SP cleaved peptides (Supplementary Figures 6D, E). No GPI anchor was predicted in SP/ZP suggesting that these proteins may be free in the stigmatic exudate (Supplementary Dataset S3). Taken together, these results suggest that DUF247 proteins on the extracellular surface of the pollen tube membrane could interact with extracellular SP/ZP secreted by the stigma.

Evidence for the co-inheritance and evolution of self-incompatibility components

Next, we aimed to examine the evolutionary history of these proteins. Two valid hypotheses that we aimed to probe were that



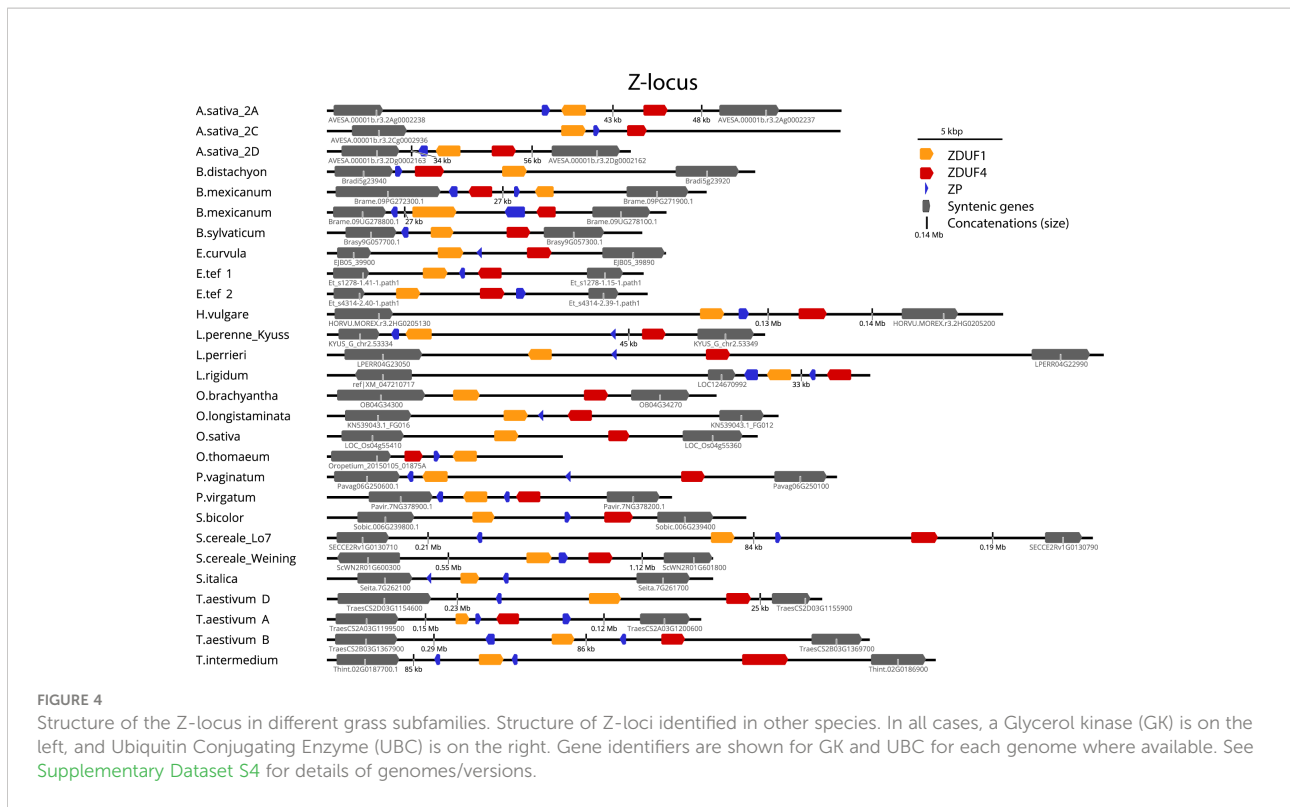


FIGURE 4

Structure of the Z-locus in different grass subfamilies. Structure of Z-loci identified in other species. In all cases, a Glycerol kinase (Gk) is on the left, and Ubiquitin Conjugating Enzyme (UBC) is on the right. Gene identifiers are shown for GK and UBC for each genome where available. See [Supplementary Dataset S4](#) for details of genomes/versions.

the SI system and specificity groups originated before the divergence of a species, or, alternatively, that SI groups formed after the divergence of a species, meaning that each species would have a unique complement of DUF247 and SP/ZP alleles. As a control, we first constructed phylogenetic trees of the Glycerol Kinase genes adjacent to the Z-locus, as we predicted that these would resemble the known phylogeny of each species. This phylogenetic tree formed the expected clades of each family, including Panicoideae, Oryzoideae, Chloridoideae and Pooideae (Figure 5). In addition, closely related species within these families grouped together, such as the three wheat genomes (*T. aestivum*) with rye (*S. cereale*) and the three oat genomes (*Avena sativa*). If specificity groups arose before speciation, we expected different DUF/SP/ZP genes from various species to cluster together, regardless of the relatedness of the species. In contrast, if speciation occurred first, and specificity groups arose later, DUF247/SP/ZP from each species would clade together (in a manner similar to the GK phylogeny).

Firstly, we produced a large phylogenetic tree including all S/ Z-DUF247 proteins, illustrating that the four key proteins are distinct between each grass subfamily (Figure 6). Given that all DUF247 proteins fell within one of the four (DUF1-4) clades, this indicates that the S- and Z-loci were inherited from a common ancestor, and that each locus (S- or Z-) has only originated once (Figure 6). In the case of the S-locus, however, specificity groups appear to have arisen after the divergence of Pooideae, Oryzoideae and the PACMAD clade (Chloridoideae

and Panicoideae), as these clades are present in the SDUF phylogeny (Figure 6). However, specificity groups within the Pooideae (and likely within other clades) arose in a common ancestor and were subsequently inherited into all pooid grasses (versions of *T. aestivum* SDUF3 from the A, B and D genomes, for example, do not cluster together; Figure 6; and see [Supplementary Figures 7, 8](#)). These features were particularly evident when comparing multiple ryegrass SDUF2 proteins identified by [Veckman et al. \(2019\)](#), with our collection of SDUF2 proteins from other species ([Supplementary Figure 7](#)). Despite the wide range of SDUF2 alleles from ryegrass, clades from Oryzoideae, Chloridoideae and Panicoideae were present; however, consistent with the idea that specificity groups arose in a common pooid ancestor, most pooid SDUF2 proteins had orthologs in the ryegrass panel ([Supplementary Figure 7](#)). Strengthening this argument, is the lack of synteny between Oryzoideae, Panicoideae and Pooideae S-loci (Figure 3). This is also reflected in the phylogeny of SP, where distinct clades for each subfamily are present (Figure 7).

At the Z-locus, however, speciation has occurred after specificity arose, evidenced by the mixing of different species throughout the tree (*S. bicolor*, from Panicoideae, for example, clusters with members of the pooid grasses, while *P. virgatum*, another Panicoideae, clusters elsewhere; Figure 6; and see [Supplementary Figure 9](#)). This was also reflected in the phylogeny of ZP, where mixtures of subfamilies cluster together (Figure 7). This is strengthened by the fact that all Z-

TABLE 1 Presence of SI determinants in other grass species.

	SDUF2	SDUF3	SP	ZDUF1	ZDUF4	ZP ^a	Subfamily
<i>Eragrostis curvula</i>	P	P	P	P	P	P	Chloridoideae
<i>Hordeum bulbosum</i> ^b	P	P	P	P	P	P	Pooideae
<i>Lolium perenne</i> (Byrne) ^c	P	P	P	P	P	P (x2)	Pooideae
<i>Lolium perenne</i> (Kyuss)	P	P	P	P	P	P (x2)	Pooideae
<i>Lolium perenne</i> (Velmurugan) ^d	P	P	P	P	P	A	Pooideae
<i>Lolium rigidum</i>	P	P	P	P	P	P (x2)	Pooideae
<i>Oryza longistaminata</i>	P	P	P	P	P	P	Oryzoideae
<i>Panicum virgatum</i> ^d	P	P	P	P	P	P (x2)	Panicoideae
<i>Paspalum vaginatum</i>	P	P	P	P	P	P (x2)	Panicoideae
<i>Thinopyrum intermedium</i>	P	P	P	P	P	P (x2)	Pooideae
<i>Avena sativa</i> (A)	P	P	P	P	P	P	Pooideae
<i>Avena sativa</i> (C)	P	P	P	P	P	P	Pooideae
<i>Avena sativa</i> (D)	P	P	P	P	P	P	Pooideae
<i>Brachypodium distachyon</i>	A	P	P	P	P	P	Pooideae
<i>Brachypodium mexicanum</i> (P)	NH	NH	NH	P	P	P (x2)	Pooideae
<i>Brachypodium mexicanum</i> (U)	P	P	A	P	P	P	Pooideae
<i>Brachypodium sylvaticum</i>	P	P	P	P	P	P	Pooideae
<i>Eragrostis tef</i> ^e	P*	P	P	P*	P	P*	Chloridoideae
<i>Hordeum vulgare</i>	A	P	A	P	P	P	Pooideae
<i>Leersia perrieri</i>	P	P	P	P	P	P	Oryzoideae
<i>Oropetium thomaeum</i>	P	P	P	P	P	P	Chloridoideae
<i>Oryza brachyantha</i>	P	P	P	P	P	P	Oryzoideae
<i>Oryza sativa</i>	P	P	P	P	P	P	Oryzoideae
<i>Secale cereale</i> (Lo7)	P	P	P	P	P	P (x2)	Pooideae
<i>Secale cereale</i> (Weining) ^f	P	P	P	P	P	P (x2)	Pooideae
<i>Setaria viridis</i>	NH	NH	NH	P	A	P (x2)	Panicoideae
<i>Sorghum bicolor</i>	P	P	P	P	P	P	Panicoideae
<i>Triticum aestivum</i> (A)	P	P	P	P	P	P (x2)	Pooideae
<i>Triticum aestivum</i> (B)	P	P	P	P	P	P (x2)	Pooideae
<i>Triticum aestivum</i> (D)	P	P	P	P	P	P	Pooideae

P, present; Red shading indicates disruption of ORF, green shading indicates no disruption. A, absent. NH, no hits (entire locus was not identified).

^aOccasionally more than one ZP gene was identified, marked by (x2).

^bThe *Hordeum bulbosum* genome is fragmented, no disruptions to DUF ORFs were found despite this.

^cThe sequenced plant was self-fertile, the location of the self-fertility allele is not published

^dThe *Lolium perenne* (Velmurugan) Z-locus was split across multiple scaffolds with ambiguous stretches contained within – the absence of ZP in this genome may be due to incompleteness

^e*Eragrostis tef* is tetraploid, cases where one allele was interrupted are indicated by a *.

^fWeining rye may contain SC mutations outside the S- and Z-loci.

loci examined were syntenic (occurring between UBC and GK proteins; Figure 4). Interestingly, sub-clusters of ZP including “ZP1” and “ZP4” also appear (so named by their proximity to ZDUF1 and ZDUF4), suggesting that ZP1 and ZP4 may be genuinely different proteins, perhaps interacting with their corresponding ZDUF protein (Figure 7).

From an evolutionary standpoint, the co-inheritance of SP/ZP and each DUF247 gene is essential to maintain SI. We compared SP and ZP phylogenetic trees with their corresponding S/Z-DUF247 trees to investigate whether these proteins were co-inherited, with the caveat that in SC species, these genes may have become unlinked. A strong association between SP proteins and SDUF3 was observed, but not so

between SDUF2 and SP (Supplementary Figure 8). Smaller clades of linkage were observed for various ZP and ZDUF1/4 groupings (Supplementary Figure 9). Of particular note was the strong association between ZDUF4 and a small ZP4 clade, while ZDUF1 appeared to clade more similarly to ZP1 proteins (Supplementary Figure 9). This analysis is confounded by the fact that most species examined have potentially been SC for long periods and recombination may have occurred between, for example, SDUF2 and SDUF3/SP in many species. There is also uncertainty around the original/functional protein sequences in these species which have likely degraded over time, therefore impacting the creation of accurate phylogenetic trees. Nevertheless, the associations observed supports the idea that

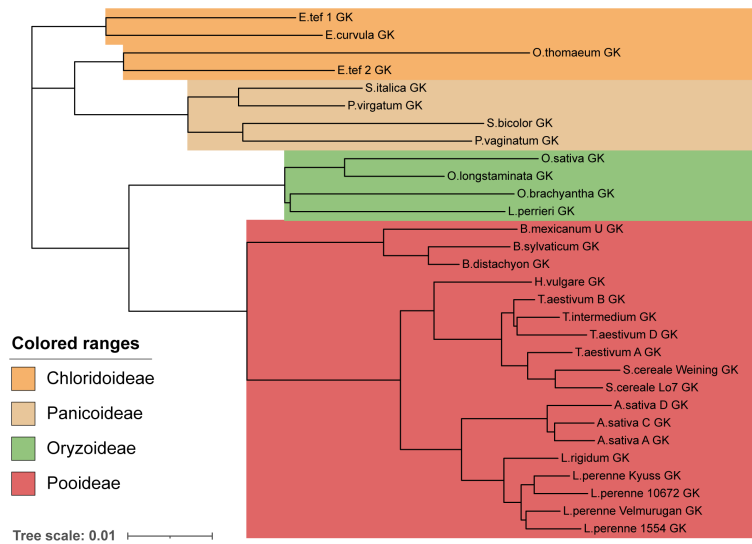


FIGURE 5
Phylogeny of grass species represented by Glycerol Kinase protein similarity. A phylogenetic tree constructed using the protein sequence of a Glycerol Kinase adjacent to all Z-loci. To produce the tree a global alignment with a BLOSUM62 cost matrix was performed, followed by tree building using a Jukes Cantor genetic distance model and nearest neighbor tree-building method. The phylogenetic tree resembles the expected relationship between species. Scale bar indicates substitutions per site.

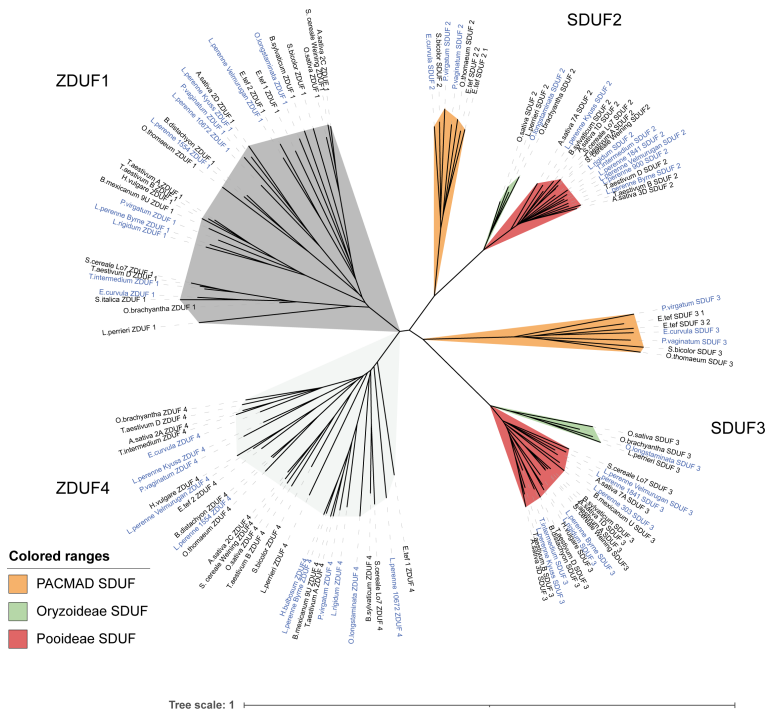


FIGURE 6
Phylogeny of S/Z-DUF247 genes from various grass species. A phylogenetic tree constructed using the sequences of all S/Z-DUF247 protein sequences. To produce the tree a global alignment with a BLOSUM62 cost matrix was performed, followed by tree building using a Jukes Cantor genetic distance model and nearest neighbor tree-building method. Subclades of SDUF2 and SDUF3 based on subfamily are shown. ZDUF1/ZDUF4 did not produce subclades. Self-incompatible species are highlighted with blue text. Scale bar indicates substitutions per site.

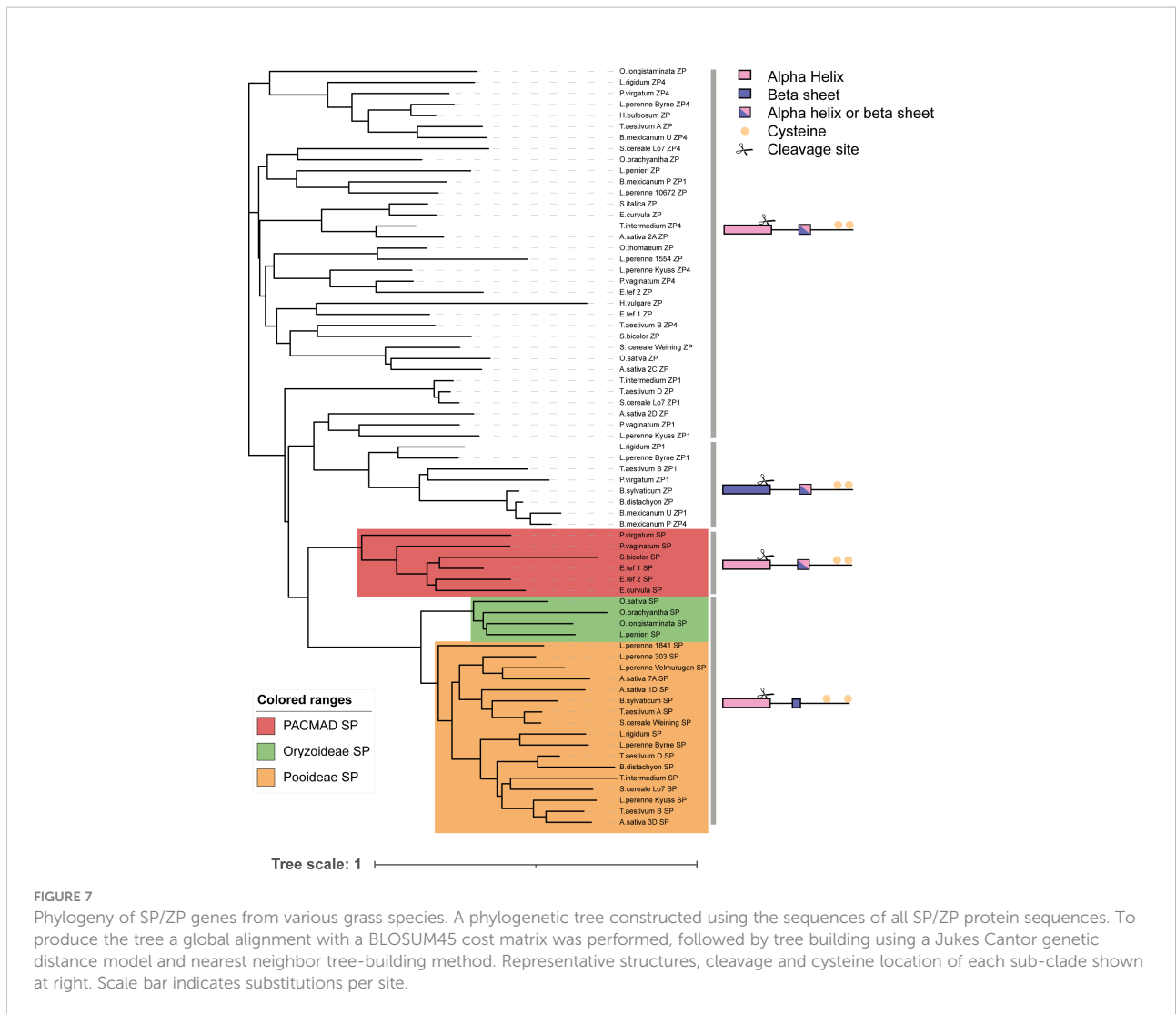


FIGURE 7

Phylogeny of SP/ZP genes from various grass species. A phylogenetic tree constructed using the sequences of all SP/ZP protein sequences. To produce the tree a global alignment with a BLOSUM45 cost matrix was performed, followed by tree building using a Jukes Cantor genetic distance model and nearest neighbor tree-building method. Representative structures, cleavage and cysteine location of each sub-clade shown at right. Scale bar indicates substitutions per site.

these genes, which are nearby, are co-inherited and evolve together.

Discussion

Here we describe the genes underlying the S- and Z-locus in Poaceae. The nature of each locus is similar, comprising two DUF247 genes, and a small gene encoding a cleaved, extracellular peptide. From this we propose a model which incorporates all six genes to instigate a self-incompatibility response in pollen (Figure 8). As both S- and Z-loci must match, we propose that a tetramer of DUF247 proteins forms upon recognition of self SP/ZP by DUF247 dimers made up of SDUF2/SDUF3 and ZDUF1/ZDUF4. This satisfies the requirement for the integration of S- and Z-signals leading to pollen-tube arrest. In *Brassica*, recognition of SCR/SP11 (a small cysteine-rich peptide) by SRK induces dimerization of SRK,

forming a 2:2 SCR/SRK heterotetramer, providing a basis for our model (Ma et al., 2016). In addition, it is known that the grass SI response occurs very quickly after pollen tube germination. It is not required that the pollen tube reaches the stigmatic surface, or further down the style, providing support for SP/ZP being free-floating in the stigmatic exudate (Shivanna et al., 1982). Downstream integration of the S- and Z- signals may be occurring through known loci, for example the “SF locus” on chromosome 5 present in ryegrass and rye (Voytkov et al., 1998; Do Canto et al., 2018).

Evolutionary analysis of the DUF247/SP/ZP proteins suggested that the Z-locus was the original locus and S has resulted from a subsequent duplication (Figure 6). While the Z-locus appears to have retained a stable position, the S-locus has translocated and diversified between grass subfamilies (Figure 3). The presence of two ZP-like genes at some Z-loci suggests that a single locus system may comprise a “Z-locus” with two DUF247 genes and two ZP genes. Such a locus would

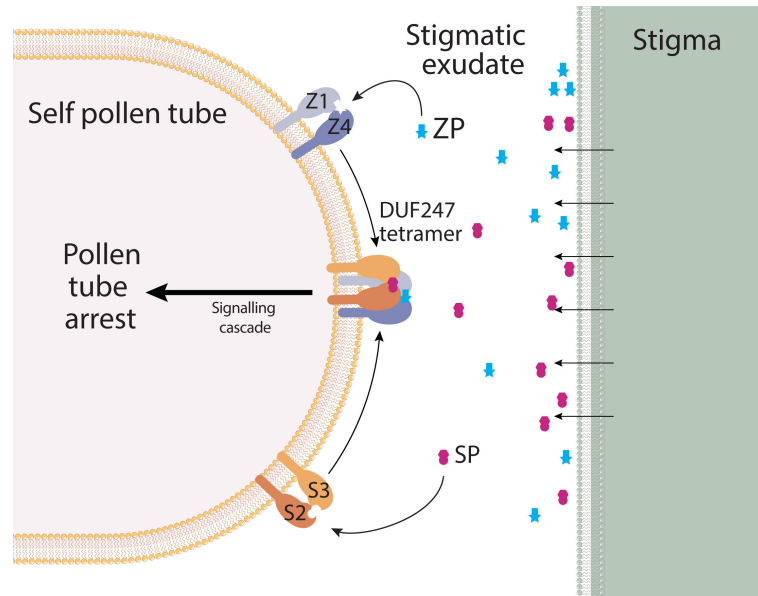


FIGURE 8

Model for SI in grasses. DUF247 proteins from S- and Z-loci form dimers at the pollen tube surface, and are anchored to the membrane by a conserved transmembrane domain. SP/ZP peptides are free-floating in the stigmatic exudate, allowing pollen tube arrest to occur immediately upon pollen germination. Signals from S- and Z- must be incorporated, therefore we propose a tetramerization of DUF247 proteins, resulting in a downstream signaling cascade that results in pollen tube arrest. Absence of any component, or non-self components would result in a lack of signaling, thus allowing pollen tube growth.

have to have been capable of inducing pollen-tube arrest by itself, perhaps by forming a tetramer similar to our proposed S/Z-tetramer (Figure 8). Upon the advent of the S-locus, the second “ZP” gene has likely become redundant. Evidence for this appears in our data as a second ZP-like gene was present on scaffold_1554; however, it was not expressed in stigma or pollen samples. As the S-locus appears to have duplicated only once in the ancestral grass species, but diversified much later, it suggests that it may have had some function during this period. Modifiers are common in other SI species such as *SLG* in *Brassica* or *Sli* in *Solanum chacoense*, and these genes are often related or linked to the specificity determinants (Boyes et al., 1991; Eggers et al., 2021). One hypothesis is that the S-locus initially acted as an enhancer of SI response, until it later diversified, becoming a second determinant of specificity. Such an evolutionary pathway would satisfy the ability of the Z-locus to function independently prior to the duplication that resulted in the S-locus. Over time, presence of the S-locus could have allowed the Z-locus to lose its ability to signal pollen-tube arrest independently. One idea on the evolution of SI, is that this pathway shares a common ancestor with plant defence systems (Hiscock et al., 1996; Nasrallah, 2005). Intriguingly, overexpression of a DUF247 gene in *Arabidopsis* leads to SAR-like defense response, supporting this hypothesis for grass SI (Kondou et al., 2013).

Surprisingly, there is scant evidence for a two-locus system operating in the PACMAD clade. To our knowledge only

Miscanthus sinensis has been analyzed for this purpose, showing a two locus system operates in this species (Jiang et al., 2017). Confounding this analysis is the extensive genome duplication in *M. sinensis*, and the apparent presence of two genomic loci that are similar to the Z-locus from other grasses on chromosome 11 and 12, with a putative SDUF gene on chromosome 18. None of the DUF genes in *M. sinensis* had full ORFs of the same size as other species – it may be that splicing is occurring in these genes, or alternatively, shortened DUF proteins may still be functional in *M. sinensis*. Until the S/Z-loci from *M. sinensis* have been mapped, this remains speculative. The large size of the “S-locus” in *S. bicolor* and particularly *P. virgatum* is also intriguing (Figure 3). In *P. virgatum* the SDUF3 gene is over 3Mb away from SDUF2, but SP is found within 500kbp of SDUF2, for example. Repression of recombination across large distances (>6 Mb) is also possible in plants, as evidenced by sex-loci in various species, which may also be occurring in the Panicoideae (Rifkin et al., 2021). The S-locus was much smaller in the three Chloridoideae species examined here (7kb in *Eragrostis curvula*, ~15kb in *E. tef* and 35kb in *O. thomaeum*). The functional *P. virgatum* Z-locus was present on Chromosome 7N, interestingly the version on Chromosome 7K contained many mutations, suggesting that upon becoming polyploid the second Z-locus has become non-functional. Similarly, *Paspalum vaginatum* had a second version of the S-locus on Chromosome 9 which contained many non-

functional mutations, while the version on Chromosome 7 was fully intact (Table 1).

Our model, shown in Figure 8, proposes that SI involves all four DUF247 and at least one SP/ZP, with the hypervariability of each gene/protein determining the specificity of the SI mechanism. Further work is required to prove this model, such as showing that the DUF247 proteins interact physically to form S/Z tetramer and this complex recognizes self SP/ZP protein. We are currently undertaking these studies as well as using gene editing to determine if each of the S- and Z-components are essential for SI.

Materials and methods

Plant growth conditions

Ryegrass plants of ONE50 cultivar were grown at 22°C in 16h/8h light/dark cycles. To induce flowering, plants were placed at 4°C for 6 weeks with 8h/16h light/dark cycles. Plant ONE50a was a clonal copy of the individual plant used for generating all sequence data to generate the draft genome assembly of ONE50.

Tissue collection and RNA preparation

Developing anthers were dissected from young flowers prior to dehiscence and placed on dry ice. Stigma were removed from flowers upon emergence and separated from ovaries and placed on dry ice before storing at -80°C. RNA was prepared using ~50mg of tissue. Frozen tissue was disrupted by bead bashing in a 1.5 mL Eppendorf tube with a stainless steel bead for 1 minute at 30 rps. Trizol (600 µL) was added to disrupted tissue, making sure to mix thoroughly as tissue thawed. The mixture was then centrifuged at 13,000 rpm for 1 minute, prior to further homogenization with a plastic micropestle. The fully-homogenized sample was centrifuged for a further minute, and the supernatant was transferred to a new tube with an equal volume of ethanol. RNA was extracted using a Zymo Quickzol miniprep kit according to manufacturer's instructions, including on column DNase treatment. RNA was eluted in 25 µL RNase-free H₂O.

RNA sequencing

Sequencing and cDNA library preparation was performed by the Otago Genomics Facility. RNA was quality assessed using a bioanalyzer having a RNA integrity number >7. ~800 ng RNA was used to make cDNA libraries using the Illumina Truseq stranded mRNA kit. Libraries were sequenced using a NextSeq 2000 with 2x100bp paired-end read length.

Mapping and analysis of RNA seq data

Fastq files were mapped to a draft ONE50 genome (J.M.E. Jacobs unpublished data) using STAR aligner v2.1.10a (Dobin et al., 2013). Alignment files (.bam) were processed in the Integrated Genomics Viewer (IGV) to create normalized count files with a window size of 25bp. Regions containing S- and Z-genes were identified by tBLASTn with DUF1-4 protein sequences as queries, followed by manual annotation of the SP/ZP genes (Veeckman et al., 2019). For *de novo* transcript assembly and quantification, Trinity software (v2.12.0) was used on combined anther and stigma datasets from each individual ("ONE50a" and "ONE50b") under default settings with adapter trimming (Grabherr et al., 2011). Transcripts were then quantified using trimmed sequences from either anther or stigma using the salmon alignment-free estimation method (Patro et al., 2017).

Identification of S- and Z-loci in other species

tBLASTn using DUF1-4 protein sequences was used to probe genome assemblies from various grass species (see Supplementary Dataset S4 for details). The combination of DUF1/4 flanked by UBC and GK genes was used to classify a region as a Z-locus, while the combination of DUF2/3 flanked by SWI/SNF and PLP genes (Pooideae), calmodulin binding protein and zinc permease (Oryzoideae), or MAPK and Flavin reductase (Chloridoideae). Except where noted, genomes were accessed using Phytozome (<https://phytozome-next.jgi.doe.gov/>), GrainGenes (<https://wheat.pw.usda.gov/>), or Ensembl Plants (<http://plants.ensembl.org/index.html>). Ryegrass genomes were downloaded from online repositories and analyzed in the same manner with local copies (Byrne et al., 2015; Velmurugan et al., 2016; Frei et al., 2021). The following genome assemblies have been described in the primary literature *Eragrostis curvula* (Carballo et al., 2019), *E. tef* (Cannarozzi et al., 2014), *Hordeum vulgare* (Mascher et al., 2021), *Leersia perrieri*, *O. longistaminata*, *O. brachyantha* (Stein et al., 2018), *O. sativa* (Kawahara et al., 2013), *Oropetium thomaeum* (VanBuren et al., 2015), *Panicum virgatum* (Lovell et al., 2021), *Sorghum bicolor* (McCormick et al., 2017), *Secale cereale* (Lo7) (Rabanus-Wallace et al., 2021), *Secale cereale* (Weining) (Li et al., 2021), *Setaria italica* (Bennetzen et al., 2012), *Triticum aestivum* (Zhu et al., 2021).

Secondary structure and feature prediction of SP/ZP proteins

Sequences of SP/ZP were aligned using ClustalW and secondary structure was analyzed using Ali2d software (Gabler et al., 2020). Signal peptide prediction was performed using SignalP

6.0 (Teufel et al., 2022). GPI anchor prediction was performed using NetGPI and subcellular localization prediction was performed using DeepLoc (Gislason et al., 2021; Thumuluri et al., 2022).

Protein prediction and phylogeny

Genes encoding DUF247 and SP/ZP were identified using tBLASTn on relevant genomes (Supplementary Dataset S4). In cases where DUF247 ORFs were disrupted, predicted sequences were reconstituted by aligning translated nucleotide sequences to the appropriate (functional) DUF247 protein, and subsequently concatenating amino acid sequences from each frame that aligned well. SP/ZP proteins were annotated by alignment to closely related SP/ZP sequences, detecting likely splice sites, and producing an in-frame SP/ZP protein, followed by recursively testing the proteins by alignment to other SP/ZP predicted proteins. Phylogenetic trees were created in Geneious (Biomatters) using parameters described in the legend of relevant figures. Trees were visualized in the interactive tree of life (iTOL; Letunic and Bork, 2021). Structural models were generated using the ColabFold implementation of AlphaFold (Jumper et al., 2021; Mirdita et al., 2022) and models were compared using ChimeraX (Pettersen et al., 2021).

Data availability statement

RNAseq datasets generated in this study have been deposited to the NCBI Sequence Read Archive (SRA) BioProject ID PRJNA864892. Access to the ONE50 genome was provided by J.M.E. Jacobs (AgResearch, Lincoln) and may be available upon request.

Author contributions

RH, RM, PM, and LB designed experiments and supervised the project. RH and TM carried out experiments. JJ provided the ONE50 draft genome assembly and a clonal copy of the plant used data to generate the draft genome assembly of ONE50. RH wrote the manuscript and produced figures. RM, PM, and JJ edited the manuscript. All authors contributed to the article and approved the submitted version.

References

- Baumann, U. (2000). Self-incompatibility in the grasses. *Ann. Bot.* 85, 203–209. doi: 10.1006/anbo.1999.1056
- Bennetzen, J. L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A. C., et al. (2012). Reference genome sequence of the model plant setaria. *Nat. Biotechnol.* 30 (6), 555–561. doi: 10.1038/nbt.2196
- Boyes, D. C., Chen, C. H., Tantikanjana, T., Esch, J. J., and Nasrallah, J. B. (1991). Isolation of a second *s*-locus-related cDNA from brassica oleracea: genetic

Funding

This study received funding from Otago Innovation Ltd via RH. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. RM and LB are Q15 funded by MBIE Smart Idea Project UOOX1911. The ONE50 genome assembly is funded by Genomics Aotearoa project 'High Quality Genomes and Population Genomics' to AgResearch (JMEJ).

Acknowledgments

We would like to thank Charles Hefer (AgResearch) for generating the draft genome assembly of ONE50 used in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1011299/full#supplementary-material>

relationships between the *s* locus and two related loci. *Genetics* 127 (1), 221–228. doi: 10.1093/genetics/127.1.221

Byrne, S. L., Nagy, I., Pfeifer, M., Armstead, I., Swain, S., Studer, B., et al. (2015). A synteny-based draft genome sequence of the forage grass *loium perenne*. *Plant J.* 84 (4), 816–826. doi: 10.1111/tpj.13037

Cannarozzi, G., Plaza-Wuthrich, S., Esfeld, K., Larti, S., Wilson, Y. S., Girma, D., et al. (2014). Genome and transcriptome sequencing identifies breeding targets in

- the orphan crop tef (*Eragrostis tef*). *BMC Genomics* 15, 581. doi: 10.1186/1471-2164-15-581
- Carballo, J., Santos, B., Zappacosta, D., Garbus, I., Selva, J. P., Gallo, C. A., et al. (2019). A high-quality genome of *eragrostis curvula* grass provides insights into poaceae evolution and supports new strategies to enhance forage quality. *Sci. Rep.* 9 (1), 10250. doi: 10.1038/s41598-019-46610-0
- Davidson, R. M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S. H., et al. (2012). Comparative transcriptomics of three poaceae species reveals patterns of gene expression evolution. *Plant J.* 71 (3), 492–502. doi: 10.1111/j.1365-313X.2012.05005.x
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15–21. doi: 10.1093/bioinformatics/bts635
- Do Canto, J., Studer, B., Frei, U., and Lübberstedt, T. (2018). Fine mapping a self-fertility locus in perennial ryegrass. *Theor. Appl. Genet.* 131 (4), 817–827. doi: 10.1007/s00122-017-3038-6
- Eggers, E. J., van der Burgt, A., van Heusden, S. A. W., de Vries, M. E., Visser, R. G. F., Bachem, C. W. B., et al. (2021). Neofunctionalisation of the *sli* gene leads to self-compatibility and facilitates precision breeding in potato. *Nat. Commun.* 12 (1), 1–9. doi: 10.1038/s41467-021-24267-6
- Fearon, C. H., Hayward, M. D., and Lawrence, M. J. (1984). Self-incompatibility in ryegrass VII. @ the determination of incompatibility genotypes in autotetraploid families of *lolium perenne* l. *Heredity* 53 (2), 403–413. doi: 10.1038/hdy.1984.97
- Foote, H. C., Ride, J. P., Franklin-Tong, V. E., Walker, E. A., Lawrence, M. J., and Franklin, F. C. (1994). Cloning and expression of a distinctive class of self-incompatibility (S) gene from papaver rhoeas l. *Proc. Natl. Acad. Sci. U.S.A.* 91 (6), 2265–2269. doi: 10.1073/pnas.91.6.2265
- Frei, D., Veeckman, E., Grogg, D., Stoffel-Studer, I., Morishima, A., Shimizu-Inatsugi, R., et al. (2021). Ultralong Oxford nanopore reads enable the development of a reference-grade perennial ryegrass genome assembly. *Genome Biol. Evol.* 13 (8), 1–6. doi: 10.1093/gbe/evab159
- Fujii, S., Kubo, K., and Takayama, S. (2016). Non-self- and self-recognition models in plant self-incompatibility. *Nat. Plants* 2 (9), 16130. doi: 10.1038/nplants.2016.130
- Gabler, F., Nam, S. Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., et al. (2020). Protein sequence analysis using the MPI bioinformatics toolkit. *Curr. Protoc. Bioinf.* 72 (1), 1–30. doi: 10.1002/cpbi.108
- Gislason, M. H., Nielsen, H., Almagro Armenteros, J. J., and Johansen, A. R. (2021). Prediction of GPI-anchored proteins with pointer neural networks. *Curr. Res. Biotechnol.* 3, 6–13. doi: 10.1016/j.crbiot.2021.01.001
- Glemin, S., and Bataillon, T. (2009). A comparative view of the evolution of grasses under domestication. *New Phytol.* 183 (2), 273–290. doi: 10.1111/j.1469-8137.2009.02884.x
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29 (7), 644–652. doi: 10.1038/nbt.1883
- Herridge, R. P., Macknight, R. C., and Brownfield, L. R. (2019). Prospects for F1 hybrid production in ryegrass. *New Z. J. Agric. Res.* 1–11, 405–415. doi: 10.1080/00288233.2018.1559867
- Hiscock, S. J., Kues, U., and Dickinson, H. G. (1996). Molecular mechanisms of self-incompatibility in flowering plants and fungi - different means to the same end. *Trends Cell Biol.* 6 (11), 421–428. doi: 10.1016/s0962-8924(96)10037-4
- Hiscock, S. J., and McInnis, S. M. (2003). Pollen recognition and rejection during the sporophytic self-incompatibility response: Brassica and beyond. *Trends Plant Sci.* 8 (12), 606–613. doi: 10.1016/j.tplants.2003.10.007
- Jiang, J. X., Guan, Y. F., McCormick, S., Juvik, J., Lubberstedt, T., and Fei, S. Z. (2017). Gametophytic self-incompatibility is operative in *miscanthus sinensis* (Poaceae) and is affected by pistil age. *Crop Sci.* 57 (4), 1948–1956. doi: 10.2135/cropsci2016.11.0932
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi: 10.1038/s41586-021-03819-2
- Kachroo, A., Schopfer, C. R., Nasrallah, M. E., and Nasrallah, J. B. (2001). Allele-specific receptor-ligand interactions in brassica self-incompatibility. *Science* 293 (5536), 1824–1826. doi: 10.1126/science.1062509
- Kao, T. H., and Tsukamoto, T. (2004). The molecular and genetic bases of s-RNase-based self-incompatibility. *Plant Cell* 16 Suppl, S72–S83. doi: 10.1105/tpc.016154
- Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013). Improvement of the *oryza sativa* nipponbare reference genome using next generation sequence and optical map data. *Rice* 6 (1), 1–10. doi: 10.1186/1939-8433-6-4
- Kondou, Y., Noguchi, K., Kutsuna, S., Kawashima, M., Yoneda, A., Ishibashi, M., et al. (2013). Overexpression of DWARF AND LESION FORMATION 1 (DLE1) causes altered activation of plant defense system in *arabidopsis thaliana*. *Plant Biotechnol.* 30 (4), 385–U177. doi: 10.5511/plantbiotechnology.13.0605a
- Leticia, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49 (W1), W293–W296. doi: 10.1093/nar/gkab301
- Lian, X., Zhang, S., Huang, G., Huang, L., Zhang, J., and Hu, F. (2021). Confirmation of a gametophytic self-incompatibility in *oryza longistaminata*. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.576340
- Li, G., Wang, L., Yang, J., He, H., Jin, H., Li, X., et al. (2021). A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nat. Genet.* 53 (4), 574–584. doi: 10.1038/s41588-021-00808-z
- Lovell, J. T., MacQueen, A. H., Mamidi, S., Bonnette, J., Jenkins, J., Napier, J. D., et al. (2021). Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* 590 (7846), 438–444. doi: 10.1038/s41586-020-03127-1
- Lundqvist, A. (1962). The nature of the two-loci incompatibility system in grasses. *Hereditas* 48 (1-2), 153–168. doi: 10.1111/j.1601-5223.1962.tb01804.x
- Ma, R., Han, Z., Hu, Z., Lin, G., Gong, X., Zhang, H., et al. (2016). Structural basis for specific self-incompatibility response in brassica. *Cell Res.* 26 (12), 1320–1329. doi: 10.1038/cr.2016.129
- Manzanares, C., Barth, S., Thorogood, D., Byrne, S. L., Yates, S., Czaban, A., et al. (2016). A gene encoding a DUF247 domain protein cosegregates with the s self-incompatibility locus in perennial ryegrass. *Mol. Biol. Evol.* 33 (4), 870–884. doi: 10.1093/molbev/msv335
- Mascher, M., Wicker, T., Jenkins, J., Plott, C., Lux, T., Koh, C. S., et al. (2021). Long-read sequence assembly: a technical evaluation in barley. *Plant Cell* 33 (6), 1888–1906. doi: 10.1093/plcell/koab077
- Ma, S., Wang, M., Wu, J., Guo, W., Chen, Y., Li, G., et al. (2021). WheatOmics: A platform combining multiple omics data to accelerate functional genomics studies in wheat. *Mol. Plant* 14 (12), 1965–1968. doi: 10.1016/j.molp.2021.10.006
- McCormick, R. F., Truong, S. K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., et al. (2017). The sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 93 (2), 338–354. doi: 10.1111/tpj.13781
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* 19 (6), 679–682. doi: 10.1038/s41592-022-01488-1
- Nagy, I., Veeckman, E., Liu, C., Bel, M. V., Vandepoele, K., Jensen, C. S., et al. (2022). Chromosome-scale assembly and annotation of the perennial ryegrass genome. *BMC Genomics* 23 (1), 505. doi: 10.1186/s12864-022-08697-0
- Nasrallah, J. B. (2005). Recognition and rejection of self in plant self-incompatibility: comparisons to animal histocompatibility. *Trends Immunol.* 26 (8), 412–418. doi: 10.1016/j.it.2005.06.005
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14 (4), 417–419. doi: 10.1038/nmeth.4197
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., et al. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* 30 (1), 70–82. doi: 10.1002/pro.3943
- Rabanus-Wallace, M. T., Hackauf, B., Mascher, M., Lux, T., Wicker, T., Gundlach, H., et al. (2021). Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. *Nat. Genet.* 53 (4), 564–573. doi: 10.1038/s41588-021-00807-0
- Rifkin, J. L., Beaudry, F. E. G., Humphries, Z., Choudhury, B. I., Barrett, S. C. H., and Wright, S. I. (2021). Widespread recombination suppression facilitates plant sex chromosome evolution. *Mol. Biol. Evol.* 38 (3), 1018–1030. doi: 10.1093/molbev/msaa271
- Schopfer, C. R., Nasrallah, M. E., and Nasrallah, J. B. (1999). The male determinant of self-incompatibility in brassica. *Science* 286 (5445), 1697–1700. doi: 10.1126/science.286.5445.1697
- Shinozuka, H., Cogan, N. O. I., Smith, K. F., Spangenberg, G. C., and Forster, J. W. (2010). Fine-scale comparative genetic and physical mapping supports map-based cloning strategies for the self-incompatibility loci of perennial ryegrass (*Lolium perenne* l.). *Plant Mol. Biol.* 72 (3), 343. doi: 10.1007/s11103-009-9574-y
- Shivanna, K. R., Heslop-Harrison, Y., and Heslop-Harrison, J. (1982). The pollen-stigma interaction in the grasses. 3. features of the self-incompatibility response. *Acta Botanica Neerlandica* 31 (4), 307–319. doi: 10.1111/j.1438-8677.1982.tb01637.x
- Stein, J. C., Yu, Y., Copetti, D., Zwickl, D. J., Zhang, L., Zhang, C., et al. (2018). Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *oryza*. *Nat. Genet.* 50 (2), 285–296. doi: 10.1038/s41588-018-0040-0
- Stout, A. B., and Chandler, C. (1941). Change from self-incompatibility to self-compatibility accompanying change from diploidy to tetraploidy. *Science* 94 (2431), 118. doi: 10.1126/science.94.2431.118

- Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gislason, M. H., Pihl, S. I., Tsigos, K. D., et al. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* 40, 1023–1025. doi: 10.1038/s41587-021-01156-3
- Thorogood, D., Yates, S., Manzanares, C., Skot, L., Hegarty, M., Blackmore, T., et al. (2017). A novel multivariate approach to phenotyping and association mapping of multi-locus gametophytic self-incompatibility reveals s, z, and other loci in a perennial ryegrass (Poaceae) population. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01331
- Thumhuri, V., Almagro Armenteros, J. J., Johansen, A. R., Nielsen, H., and Winther, O. (2022). DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res.* 50 (W1), W228–W234. doi: 10.1093/nar/gkac278
- VanBuren, R., Bryant, D., Edger, P. P., Tang, H., Burgess, D., Challabathula, D., et al. (2015). Single-molecule sequencing of the desiccation-tolerant grass *oropetium thomaeum*. *Nature* 527 (7579), 508–511. doi: 10.1038/nature15714
- Veeckman, E., Van Glabeke, S., Haegeman, A., Muylle, H., van Parijs, F. R. D., Byrne, S. L., et al. (2019). Overcoming challenges in variant calling: exploring sequence diversity in candidate genes for plant development in perennial ryegrass (*Lolium perenne*). *DNA Res.* 26 (1), 1–12. doi: 10.1093/dnares/dsy033
- Velmurugan, J., Mollison, E., Barth, S., Marshall, D., Milne, L., Creevey, C. J., et al. (2016). An ultra-high density genetic linkage map of perennial ryegrass (*Lolium perenne*) using genotyping by sequencing (GBS) based on a reference shotgun genome assembly. *Ann. Bot.* 118 (1), 71–87. doi: 10.1093/aob/mcw081
- Voylokov, A. V., Korzun, V., and Borner, A. (1998). Mapping of three self-fertility mutations in rye (*Secale cereale* L.) using RFLP, isozyme and morphological markers. *Theor. Appl. Genet.* 97 (1-2), 147–153. doi: 10.1007/s001220050879
- Walker, E. A., Ride, J. P., Kurup, S., Franklin-Tong, V. E., Lawrence, M. J., and Franklin, F. C. (1996). Molecular analysis of two functional homologues of the S3 allele of the papaver rhoeas self-incompatibility gene isolated from different populations. *Plant Mol. Biol.* 30 (5), 983–994. doi: 10.1007/BF0020809
- Yang, B., Thorogood, D., Armstead, I., and Barth, S. (2008). How far are we from unravelling self-incompatibility in grasses? *New Phytol.* 178 (4), 740–753. doi: 10.1111/j.1469-8137.2008.02421.x
- Zhu, T., Wang, L., Rimbart, H., Rodriguez, J. C., Deal, K. R., De Oliveira, R., et al. (2021). Optical maps refine the bread wheat *triticum aestivum* cv. Chinese spring genome assembly. *Plant J.* 107 (1), 303–314. doi: 10.1111/tj.15289