A Gene Encoding a DUF247 Domain Protein Cosegregates with the S Self-Incompatibility Locus in Perennial Ryegrass

Manzanares, Chloé; Barth, Susanne; Thorogood, Daniel; Byrne, Stephen; Yates, Steven; Czaban, Adrian; Asp, Torben; Yang, Bicheng; Studer, Bruno

Published in:
Molecular Biology and Evolution

DOI:
10.1093/molbev/msv335

Publication date:
2015

Citation for published version (APA):
Manzanares, C., Barth, S., Thorogood, D., Byrne, S., Yates, S., Czaban, A., Asp, T., Yang, B., & Studer, B. (2015). A Gene Encoding a DUF247 Domain Protein Cosegregates with the S Self-Incompatibility Locus in Perennial Ryegrass. *Molecular Biology and Evolution*, *33*(4), 870-884. https://doi.org/10.1093/molbev/msv335

MBE Advance Access published December 10, 2015

Article (discoveries)

A gene encoding a DUF247 domain protein co-segregates with the *S* self-incompatibility locus in perennial ryegrass

Chloé Manzanares[1,2,3], Susanne Barth[2], Daniel Thorogood[3], Stephen L. Byrne[4], Steven Yates[1], Adrian Czaban[4], Torben Asp[4], Bicheng Yang[5] and Bruno Studer[1]*

[1] Forage Crop Genetics, Institute of Agricultural Sciences, ETH Zurich, 8092 Zurich, Switzerland.

[2] Teagasc Crops, Environment and Land Use Programme, Oak Park Research Centre, Carlow, Ireland.

[3] Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Plas Gogerddan, Aberystwyth, Ceredigion, SY23 3EB, UK.

[4] Department of Molecular Biology and Genetics, Research Centre Flakkebjerg, Aarhus University, Forsøgsvej 1, 4200 Slagelse, Denmark.

[5] BGI-Shenzhen, Building 1, Beishan Industrial Zone, Yantian District, Shenzhen, 518083, China.

*Corresponding author: Bruno Studer

E-mail: bruno.studer@usys.ethz.ch

Key words: Domain of unknown function (DUF) 247, Fine-mapping, Perennial ryegrass (*Lolium perenne* L.), Self-incompatibility (SI), RNA sequencing (RNAseq), *S*-locus

Abstract

The grass family (Poaceae), the fourth largest family of flowering plants, encompasses the most economically important cereal, forage and energy crops, and exhibits a unique gametophytic self-incompatibility (SI) mechanism that is controlled by at least two multi-allelic and independent loci, *S* and *Z*. Despite intense research efforts over the last six decades, the genes underlying *S* and *Z* remain uncharacterised.

Here, we report a fine-mapping approach to identify the male component of the *S*-locus in perennial ryegrass (*Lolium perenne* L.) and provide multiple evidence that a *domain of unknown function 247* (*DUF247*) gene is involved in its determination. Using a total of 10,177 individuals from seven different mapping populations segregating for *S*, we narrowed the *S*-locus to a genomic region containing eight genes, the closest recombinant marker mapping at a distance of 0.016 centimorgan. Of the eight genes co-segregating with the *S*-locus, a highly polymorphic gene encoding for a protein containing a DUF247 was fully predictive of known *S*-locus genotypes at the amino acid level in the seven mapping populations. Strikingly, this genes showed a frame-shift mutation in self-compatible darnel (*Lolium temulentum* L.), whereas all of the self-incompatible species of the *Festuca-Lolium* complex were predicted to encode functional proteins.

Our results represent a major step forward towards understanding the gametophytic SI system in one of the most important plant families and will enable the identification of additional components interacting with the *S*-locus.

Introduction

In flowering plants, self-pollination and subsequent inbreeding leads to increased homozygosity by fixation of alleles at gene loci and can reduce the vigour of plants over successive generations through inbreeding depression (Charlesworth and Willis 2009). To maintain diversity and prevent inbreeding, many angiosperms possess one of several systems of self-incompatibility (SI).

Homomorphic SI prevents self-pollination through physiological recognition of self-pollen by the style and is a widespread mechanism occurring in over half of the angiosperms (de Nettancourt 1977) . In many homomorphic SI systems, SI is genetically governed by a single multi-allelic *S*-locus (Yang et al. 2008) but in the grass gametophytic SI system, it involves at least two multi-allelic and independent loci, *S* and *Z* (Hayman 1956; Lundqvist 1954). This *S*-*Z* system has been found to operate in rye (*Secale cereale* L.) (Lundqvist 1954), sunolgrass (*Phalaris coerulescens* Desf.) (Hayman 1956), perennial ryegrass (*Lolium perenne* L.) (Cornish et al. 1979) and many other grass species as reviewed by Li et al. (1997). With a gametophytic two locus SI system, successful pollination only occurs if the allele of the haploid pollen of at least one of the two loci, *S* or *Z*, is different from the *S* and *Z* alleles in the diploid stigma.

To date, neither of the two loci have been unequivocally characterised at the sequence level, but it is known that in rye (Voylokov et al. 1998) and perennial ryegrass (Thorogood et al. 2002), the *S* and *Z* loci are located on chromosomes 1 and 2, respectively. Bian et al. (2004) delimited the *S* and the *Z* locus in sunolgrass to a region of 0.26 centimorgan (cM) on the short arm of chromosome 1 and 0.9 cM on chromosome 2, respectively. However, the genetic location of the *S*-locus is associated with a large physical region, situated in the subcentromeric region where recombination rates are expected to be reduced compared to more distal parts of the chromosome (Pedersen et al. 1995; Sim et al. 2005). Taking a comparative mapping approach by, for example, using rice (*Oryza sativa* L.) marker and sequence data to deduce the position of the ryegrass *S*-locus, can be challenging because synteny between rice and perennial ryegrass is incompletely conserved; linkage group (LG) 1 of perennial ryegrass in this region is largely syntenic to rice chromosome 5 but with a small insertion of a region syntenic to rice chromosome 10 (Pfeifer et al. 2013).

In bulbous barley (*Hordeum bulbosum* L.), several pistil and anther-specific sequences have been isolated using selective amplification of cDNA from anther and pistil tissue to identify *S*

allele-specific expressed genes (Kakeda 2009; Kakeda et al. 2008). Using a fine-mapping approach with a population of 662 individuals and markers designed on these specific amplicons, three linked candidate genes have been identified: two anther-specific genes, *HAS31* and *HAS175* and one pistil-specific gene *HPS10*. The candidate gene *HAS175* shares sequence similarity to two rice genes located around 26 Mb on rice chromosome 5 (Os05g0532300 and Os05g0535200) and described as proteins containing an F-box domain. F-box proteins (SLFs) are known to be involved in S-RNase SI system of the Solanaceae family as multiple male-determinants (Kubo et al. 2015). The pistil-specific *HSP10* is a highly polymorphic candidate gene located around 6 Mb on the same chromosome in rice (Os05g0198050) and is annotated as a hypothetical protein gene.

In perennial ryegrass, a gene expression study has been conducted by Yang et al. (2009) using the construction of several suppression subtractive hybridization libraries. Seven candidate genes for the *S*-locus have been identified and their expression was confirmed by real-time PCR. Among those candidate genes, four genes (*Can3*, *Can4*, *Can94* and *Can136*) have a protein kinase function which could imply that the grass SI response is determined by a kinase-triggered pathway such as in the *Brassica* sporophytic SI system (*S*-locus receptor kinase (SRK) on the stigma) (Nasrallah 2002) or the downstream SI response of the poppy gametophytic SI system (involvement of mitogen-activated protein kinase, MAPK) (Rudd et al. 2003). One *S*-locus linked bacterial artificial chromosome (BAC) clone *Lp*Bm2 was selected by Shinozuka et al. (2010) in perennial ryegrass. Sequencing of the BAC revealed two candidate genes homologous to Os05g0168800 (a spectrin repeat containing protein) and Os05g0169000 (a thioredoxin-like protein) on rice chromosome 5. A similar approach applied to the *Z* locus identified, on the basis of complementary expression analysis and nucleotide diversity measures, a *ubiquitin-specific protease* and a *DUF247* gene as the most plausible candidates among the currently known *Z* genes (Shinozuka, et al. 2010).

The genes putatively involved in SI systems have been identified using different approaches: fine-mapping (Stein et al. 1991; Wang et al. 2003), transcriptome analysis from male and female tissues (Boyes and Nasrallah 1995; Schopfer et al. 1999; Suzuki et al. 1999), proteomic studies (Anderson et al. 1986; Foote 1994; Hinata and Nishio 1978) and DNA sequence analyses (Wheeler et al. 2009). In grasses, Kakeda et al. (2008) used a transcriptomic approach to isolate tissue-specific genes, followed by genetic linkage mapping to identify recombination between the candidate genes and the *S*-locus. In perennial ryegrass,

an opposite strategy was used where the *S*-locus region was first mapped before studying the tissue-specific gene expression pattern of selected candidate genes (Yang, et al. 2008).

A classical linkage mapping approach, based on biparental mapping populations that segregate for a particular SI locus, creates segregation distortion (SD) in the progeny as only one of the two possible *S*-allelic classes is revealed (Thorogood et al. 2005; Thorogood, et al. 2002). These mapping families are usually pseudo-F2 populations resulting from a specific cross of two F1 siblings that show a half-compatible pollination with only one of the *S*-alleles being transmitted through the pollen gametes. This creates SD at *S* and closely linked marker loci and offers the opportunity to locate *S* by determination of allele frequencies at these markers in the offspring. Map-based cloning of SI genes based on SD is very effective, especially in combination with a high throughput genotyping method such as high resolution melting curve analysis (HRM). Due to the specificity and flexibility of HRM to target any type of DNA sequence polymorphism in a specific genome region (Studer et al. 2009), HRM is particularly useful for large-scale mapping projects such as *S*-locus fine-mapping.

The overall objective of this study was to isolate the *S* SI locus in perennial ryegrass. Specifically, we aimed at (1) fine-mapping the *S*-locus region in perennial ryegrass, using bi-parental mapping populations segregating for the *S*-locus, (2) isolating and sequencing BAC clones covering the *S*-locus fine-mapping region in order to identify the genes contained within the *S*-locus region, (3) studying the gene expression of the *S* region by sequencing pollen and stigma RNA from plants representing a variety of *S* genotypes in order to corroborate *S*-locus candidate genes with pollen and stigma expression levels and (4) studying the allelic diversity of the nucleotide and predicted protein sequences.

Results

Fine-mapping the *S*-locus in perennial ryegrass

A total of 10,177 plants from seven perennial ryegrass populations were screened to fine-map the male component of the *S*-locus. These populations were developed from two different genetic backgrounds (DTS and VrnA), and were designed to be heterozygous for the *S*-locus as shown in Figure 1.

The *S*-locus mapping in perennial ryegrass was initiated using published molecular markers in close proximity to the *S*-locus region (Jensen et al. 2005b; Jones et al. 2002; Odonoughue et al. 1992; Studer et al. 2008; Yang, et al. 2009). Twenty markers were tested on a subset of 242 plants from one of the DTS mapping populations (P235/59) in order to identify and estimate the number of homozygous recombinants. Two markers, As_CDO1173 and Hv_BCD921, applied in previous genetic mapping studies (Bian, et al. 2004; Sim, et al. 2005), were used to identify the syntenic region between rice and perennial ryegrass. The genomic region between 1.5 to 5 Mbp on rice chromosome 5 was identified as the target region syntenic to the *S*-locus in perennial ryegrass.

The rice genome sequence syntenic to the perennial ryegrass *S*-locus region was used as a reference to align perennial ryegrass transcripts described in Studer et al. (2012) in order to design ryegrass-specific genic DNA markers. These markers were designed for HRM analysis and tested on the recombinants identified in the 242 plants of the DTS population P235/59. Out of 40 markers, 18 were polymorphic. According to the number and distribution of recombinants, two new flanking markers (05_02720 and 05_02915) were selected to screen an additional 1151 plants of the DTS mapping population P235/59. With these markers, the *S*-locus was delimited to a region of 3.45 cM. From the P235/59 population (in total consisting of 1,393 plants), 48 recombinants were isolated using these two flanking markers. The recombinants were genotyped for six markers designed to map between 05_02720 and 05_02915 in rice: two markers co-segregated with the *S*-locus (05_02827 and 05_02833) and the other four markers identified one, two or three recombinants (Table 1).

Six additional populations, similarly designed to the DTS population P235/59, were used for further fine-mapping of the *S*-locus. These populations consisted of 8,784 plants in total and were screened to identify recombinants between the two markers broadly flanking the *S*-locus

region (As_CDO1173 and 05_03433). These recombinants were genotyped with eight polymorphic markers within the *S*-locus region (Table 1).

The perennial ryegrass *S*-locus region was aligned to the physical genome sequence of rice and Brachypodium (*Brachypodium distachyon*) to identify the gene content of the syntenic region. We identified eight genes in rice and eleven genes in Brachypodium (Figure 2). Markers specifically designed to target the rice and Brachypodium orthologs within the *S*-locus were used to genotype the recombinants identified through fine-mapping. Four markers were polymorphic: 05_RB0150300_1, Bd2g35780_1, Bd2g35707_intr3 and Bd2g35690_intr1. The marker 05_RB0150300_1, designed from the conserved region between the rice gene Os05g0150300 and the Brachypodium gene Bradi2g35740, showed complete linkage to the *S*-locus in all 10,177 plants. The marker Bd2g35707_intr3, based on the same gene as the marker 05_02889, was polymorphic for all the populations and identified one recombinant out of 10,177 plants (<0.01 cM). The marker Bd2g35690_intr1 was located two genes away in Brachypodium and showed two recombinants (<0.02 cM). However, the marker 05_B35780_1, designed from the adjacent Brachypodium gene Bradi2g35780 region, showed a higher recombination frequency than the flanking marker 05_02889 with 48 recombinants. This indicates small-scale rearrangements of homologous genes in the genomic region syntenic to the *S*-locus in these species (Figure 2, thin dash lines).

Establishing the genomic sequence at the *S*-locus in perennial ryegrass

In order to isolate the *S*-locus in perennial ryegrass, the genomic region was sequenced from BAC clones covering the region identified through fine-mapping. Nine BAC clones from two different BAC libraries, LTS18 and NV#20F1-30 (hereafter referred to as F1-30) (Farrar et al. 2007), were selected using the flanking markers showing the least number of recombinants as well as non-recombinant markers. The BAC clones were sequenced using 454-sequencing and the sequences of each BAC clone were individually *de novo* assembled, resulting in one to 40 contigs for each of the nine BAC clones. The order, orientation and overlaps of these BAC sequence contigs were determined using comparative sequence analysis and verified with both the markers used for fine-mapping (Figure 3, marker names given in black) and novel markers developed from the BAC end sequences (Figure 3, marker names given in blue). To further increase contiguity and ordering, we took advantage of the draft genome sequence of perennial ryegrass (Byrne et al. 2015) and isolated five genomic scaffolds

aligning to the BAC contigs, covering 304 Kb of the *S*-locus region (Figure 3). The physical sequence co-segregating with the *S*-locus was estimated to be a minimum size of 247 Kb. The five genome scaffolds complemented with the BAC sequence contigs provided a rich source of sequence information to infer the gene content at the *S*-locus. A sequence gap of unknown length between the scaffolds 5728 and 16486 remains unresolved.

The annotation of the five genome scaffolds using perennial ryegrass expressed sequence tag (EST) databases as well as a BLAST search to the NCBI database revealed a number of conserved genes between perennial ryegrass, rice and Brachypodium (Figure 3 and Table 2). Both empirical and *ab initio* annotation of the *S*-locus region identified at least eleven genes with many repetitive elements such as retrotransposons. Two genes, syntenic to rice and Brachypodium, encoding a pyridoxal 5-phosphate (PLP)-dependent enzyme and a putative nucleotide-binding site leucine-rich repeat (NBS-LRR) protein, were duplicated within the *S* co-segregating region.

Expression profiling of genes at the *S*-locus

The expression pattern of annotated genes at the *S*-locus in perennial ryegrass was assessed in pollen and stigma tissues, as well as in pollinated stigmas expressing incompatible (self-pollination) and compatible (cross-pollination) pollen-stigma interactions, by RNA sequencing (RNAseq). RNAseq was conducted in the parental genotypes of the mapping populations and included three biological replicates. The variation in gene expression was estimated based on ten housekeeping genes used as a reference (Figure 4, Figure S1). The heatmap shows that the expression of the housekeeping genes was not conserved between the pollen and stigma tissue samples (Figure 4). Four housekeeping genes, *β-tubulin*, *EF1-α*, *CPB20* and *eIF4A-1*, have their expression level conserved between pollen and stigma samples, unlike the other six housekeeping genes which are differentially expressed in either pollen or stigma tissue ($p<0.001$). In contrast, the expression of all housekeeping genes was constant between the incompatible and compatible pollen-stigma interactions in perennial ryegrass (Figure S1).

The expression was assessed for each sample using the biological replicates from F1-30 and each annotated gene at the *S*-locus, with a threshold of 50 reads per gene in order to remove any lowly expressed genes. Out of the eleven annotated genes described in Table 2, the *NBS-LRR protein coding* gene 2 on scaffold 5728 was eliminated from the expression study as it did not pass the expression threshold limit. Similarly, the partial *ATP-dependent DNA*

helicase gene on scaffold 5714 was discarded for the comparison of incompatible and compatible pollinations due to expression levels below the threshold limit. None of the genes annotated at the *S*-locus were differentially expressed between incompatible and compatible pollinations at the significance level of $p<0.05$ (Figure S1). A comparison of the gene expression profile between the pollen and the stigma tissue revealed that four genes did not show any significant differences: the *PLP-dependent enzyme* gene (both copies), the *NBS-LRR protein coding* gene and the *SNF2-like* gene. Out of the remaining six annotated genes, three were relatively up-regulated in stigma tissue, namely the Ca^{2+} *binding protein* gene, the *TIR1-like* gene (*LpTIR1)* and the *ds-RNA binding protein* gene (*LpdsRNA)*. The *DUF247* domain containing gene (*LpSDUF247),* the *DNA helicase RecQ* gene and a partial duplication of that gene were relatively up-regulated in pollen ($p<0.005$).

Considering only the genes co-segregating with the *S*-locus, only three genes were differentially expressed between pollen and stigma: the *LpTIR1* and the *LpdsRNA* gene were relatively up-regulated in stigma tissue and the *LpSDUF247* gene was relatively up-regulated in pollen ($p<0.005$).

Determination of the allelic diversity for genes co-segregating with the *S*-locus

In order to further evaluate the genes co-segregating with the *S*-locus, the allelic diversity of each gene was assessed, using RNAseq data from the parental genotypes of the mapping populations. RNAseq raw reads were aligned to the predicted coding sequence (CDS) of the genome scaffolds (hereafter referred to as the reference) for single nucleotide polymorphisms (SNPs) discovery.

According to the design of the mapping populations, the parental genotypes used for RNAseq have known *S* genotypes: P235/58/3 is homozygous S_1S_1 (*S* allele nomenclature is relative), P235/59/3 and P253/59/21 are heterozygous S_1S_2, and F1-30 is heterozygous S_3S_4. On the basis of the homozygous genotype P235/58/3 (S_1S_1), it was possible to infer the allelic sequence of S_2 from the RNA sequences of the genotypes P235/59/3 and P253/59/21. The alleles S_3 and S_4 from the genotype F1-30 were inferred using overlapping reads that covered more than one SNP, enabling manual haplotyping. A fifth allele S_5, isolated from the BAC clones of the BAC library LTS18, was similar to the reference allele on both the DNA and amino acid (aa) sequence level. It is very likely that the genotype LTS18 used for BAC

library construction and the genotype used for the *Lolium* genome sequencing project carry the same S allele, as they share a direct progenitor.

Based on the alleles S_1 to S_5, gene phylogenies were constructed using nucleotide back-translated protein alignments for each annotated gene co-segregating with the *S*-locus in perennial ryegrass (Figure 5). Five out of eight co-segregating genes showed either sequence conservation or the allelic diversity was not sufficiently high to differentiate the five expected *S*-locus genotypes segregating in the seven mapping populations. For the other three genes, the allelic diversity was very low for *LpdsRNA* and *LpTIR1* while an extreme level of sequence divergence at *LpSDUF247* indicated strong diversifying selection on this gene (Table S1). The positional distribution of the allelic diversity along the genome scaffolds identified a unique peak at *LpSDUF247*, flanked by conserved genome regions.

A detailed sequence analysis of the protein-translated *LpTIR1* alleles revealed only three aa polymorphisms between the five alleles (Table S1). Hence, *LpTIR1* failed to differentiate the five expected *S*-locus genotypes as the aa sequence of the alleles S_1, S_2 and S_5 were 100% similar. Similarly, the *LpdsRNA* gene showed only very few aa polymorphisms separating the different alleles (Table S1). In contrast, *LpSDUF247* revealed a very high allelic diversity, also on protein level with 79.9% to 81.4% protein sequence conservation between alleles S_1 to S_5 (Figure 6, Table S1).

Structural and comparative sequence analysis of the *LpSDUF247* gene

Identifying *LpSDUF247* as a promising candidate gene, the *LpSDUF247* orthologs in the self-compatible darnel (*Lolium temulentum* L.) as well as in the self-incompatible annual ryegrass (*Lolium multiflorum* Lam.), westerwolds ryegrass (*Lolium multiflorum* Lam. var. *westerwoldicum* Wittm.) and meadow fescue (*Festuca pratensis* Huds.) were sequenced, translated into protein sequences and aligned with the perennial ryegrass *LpSDUF247* protein sequences (Figure S2). Whereas all of the self-incompatible species of the *Festuca-Lolium* complex were predicted to encode functional proteins, a frame-shift mutation changing the last 24 aa of the C-terminus was found in the self-compatible darnel. These aa are part of a predicted transmembrane domain of the *LpSDUF247* protein, being lost in the self-compatible darnel. Beyond the *Festuca-Lolium* complex, a frame-shift mutation leading to a premature stop codon was found in the *LpSDUF247* ortholog of diploid and self-compatible oat (*Avena atlantica* B. R. Baum & G. Fedak), a close relative to ryegrasses. Other self-compatible

species of the Poaceae family showed either large deletions (as the case for *Oryza brachyantha*) or premature stop codons (as found in *Aegilops tauschii*) in that gene (Figure S2).

Discussion

A combination of fine-mapping, genome sequencing, transcriptome analysis and detailed comparative sequence analysis provides evidence that the *LpSDUF247* gene determines the *S*-locus male component of the gametophytic SI system in perennial ryegrass. Although it proved difficult to isolate the *S*-locus genes using classical fine-mapping, this step was essential to define the genome region containing the *S* locus. Through physical sequence establishment including annotation, gene expression analysis and assessment of the allelic variability, it was possible to establish a list of candidate genes co-segregating with the *S* locus. The reconstruction of alleles at these candidate genes using RNAseq reads from the parental genotypes of the mapping populations enabled further refinement of candidate genes, as the *S* allele composition is known in these parental plants. In addition to allele comparisons, this allowed us to distinguish between polymorphisms not directly associated with SI and candidate polymorphisms matching the expected allele compositions in the mapping families. Most importantly, comparative sequence analyses across species allowed us to functionally associate our findings in self-compatible and incompatible species of the *Festuca-Lolium* complex and other closely related Poaceae species.

Isolation of self-incompatibility genes by classical genetics

Fine-scale comparative linkage mapping was initially considered as a promising strategy to identify and isolate Poaceae SI loci (Shinozuka, et al. 2010) but failed to pinpoint one single locus determining the *S* genes. Although conducted in seven different mapping populations containing a total of 10,177 individual plants, the co-segregating region still represents a large physical region of over 247 kb. This is for two reasons; firstly, on a genome-wide scale, the recombination frequency in subcentromeric regions, where the *S*-locus is located, is expected to be reduced compared to more distal parts of the chromosome (Studer, et al. 2012). An illustration of this reduction in the recombination frequency at *S* is the average genetic distance/physical distance ratio (Mb/cM ratio), which is significantly higher at the *S*-locus (2.47 Mb/cM minimum) when compared to the genome-wide estimation of 1.7- 2.0 Mb/cM by Shinozuka et al. (2010). Interestingly, the recombination frequency varied markedly around the *S*-locus between the different mapping populations: the DTS population P235/59 displayed a higher recombination frequency when compared to the other six populations. The recombination frequency even varied between populations derived from genetically similar parents as was the case for the VrnA-S populations. Secondly, on a local scale, there is a

sound biological reason to block recombination at SI loci, as separation of the pollen and stigma SI determinants by genetic recombination would result in the loss of functionality of the SI system. This genetic and physical linkage might be maintained by distinct chromatin modifications associated with genic and repetitive DNA sequence which is known to influence the pattern of genetic recombination in plant genomes (Henderson 2012). In addition, the genomic constitution around the *S*-locus is important. Dooner and He (2008) found that heterozygous plants differing by local insertions of a retrotransposon cluster have a twofold lower local recombination frequency than homozygotes without the retrotransposon cluster. Indeed, frequency and distribution of retrotransposons can differ between populations and therefore affect their neighbouring genes' recombination rates with up to three-fold variation in the genetic distance between maize (*Zea mays* L.) populations (Williams et al. 1995). In our *S*-locus region, the presence of five elements similar to the perennial ryegrass retrotransposon Camilla (Asp et al. 2011) could affect recombination and prevent crossing over events between the pollen and stigma component, thereby maintaining the functionality of the SI system.

Expression profiling of genes co-segregating with the *S*-locus

The gene expression analysis provided insights into global and tissue-specific gene expression patterns of *S* candidate genes. Out of the seven expressed genes co-segregating with the *S*-locus, four genes were not differentially expressed: the *SNF2-like* gene, the *NBS-LRR protein coding* gene and the two duplicated *PLP-dependent enzyme* genes. These genes, with the exception of the *SNF2-like* gene, had a low level of expression in both tissues. However, the low expression of a gene at pollen and stigma maturity does not exclude it from being an *S* candidate, as gene expression in a gametophytic SI reaction is likely to peak before maturity of the functional proteins. The perennial ryegrass SI reaction occurs during the first ten minutes after pollination and therefore, the proteins involved must be functional when the reproductive tissues are mature. The three remaining genes were over-expressed in either stigma (*LpTIR1* and *LpdsRNA*) or pollen (*LpSDUF247*). But different expression levels between pollen and stigma tissue were also observed for six of the ten housekeeping genes. It is known that the expression of housekeeping genes is not always stable across tissue types, as shown by Martin et al (2008). In contrast, the expression of the ten housekeeping genes in incompatible and compatible pollinations, for which the same tissue types have been used, was constant.

In addition to the highly conserved expression of all the ten housekeeping genes, no significant difference in gene expression between incompatible and compatible pollinations was found for the *S* candidate genes. While this seems plausible for the initial self/non-self recognition component, it indicates that the genes involved in any downstream reaction of the SI response are unlikely to be located at the *S*-locus, as is the case in many other SI systems (Isokawa et al. 2010).

Candidate genes co-segregating with the *S*-locus

From the fine-mapping and the gene expression analysis between pollen and stigma at the *S*-locus, three promising candidate genes were identified: *LpTIR1*, *LpdsRNA* and *LpSDUF247*. The *LpTIR1* gene is annotated in rice and Brachypodium as a transporter inhibitor response-1 (TIR1) like protein. The gene is predicted *in silico* to be involved in the auxin signalling pathway and the ubiquitin conjugation pathway. In Arabidopsis (*Arabidopsis thaliana*), TIR1 is part of the Skp1-cullin-F-box TIR1 (SCF^{TIR1}) complex and directly links reception of auxin to degradation of the Aux/indole-3-acetic acid proteins that act as repressors of auxin response factors (ARF) in the absence of auxin (Santner and Estelle 2009). Many SCF complexes are present in plants but the specificity of SCF^{TIR1} is due to the protein TIR1 containing an F-box domain, interacting with the Skp1 protein and a leucine-rich repeat (LRR) domain specifically regulated by auxin. Using transformed Arabidopsis exhibiting functional SI, it has been demonstrated that the abolition of the auxin signalling pathway through ARF3 enhances SI (Tantikanjana and Nasrallah 2012). The SCF complex is also known to be involved in the S-RNase gametophytic SI system. In this mechanism, the F-box protein SLF, expressed in pollen, targets pistil S-RNase from the same allele for degradation. However, the LpTIR1 protein is dissimilar to SLF as it contains a LRR domain unlike SLF, is highly expressed in stigma tissue and does not have a hyper-variable coding region, corresponding to the *S*-allele specificity of SLF (Lai et al. 2002; Sonneveld et al. 2005). LpTIR1 protein sequences are conserved between the alleles S_1, S_2 and the reference protein and only one or two aa changes were observed for the alleles S_3 and S_4, respectively. The allelic variation of the LpTIR1 protein is too low to be responsible for the *S* allele specificity and the recognition between self/non-self-pollen. Therefore, it is unlikely that this gene is coding for the stigma *S* component.

The gene *LpdsRNA*, annotated in rice and Brachypodium as a protein containing two ds-RNA binding domains (DRBDs), is an interesting *S* candidate gene. Ds-RNA binding proteins interact with small ds-RNA molecules via the DRBDs, without specificity to the ds-RNA

sequence (Fierro-Monti and Mathews 2000). The DRBDs are encoded by short and conserved nucleotide sequence motifs and mutations in the corresponding aa sequence can lead to non-functional DRBDs (Krovat and Jantsch 1996). Members of the DRBD-containing protein family include, among others, DICER and RNAse III (Saunders and Barber 2003), and can have many functions such as pre-mRNA editing, RNA regulated kinase activity or RNA transport or gene-specific silencing (Fierro-Monti and Mathews 2000). In perennial ryegrass, *LpdsRNA* is highly expressed in stigma tissue and could link SI with the RNA pathway, similar to the *S*-RNase-mediated system in *Solanaceae* (Kubo et al. 2010) and the control of allelic dominance of the pollen *S*-locus in the *Brassicaceae* plant family (Durand et al. 2014). However, the DRBDs of *LpdsRNA* were highly conserved at both the nucleotide and the protein sequence level in perennial ryegrass and it remains subject of further research if the rather limited allelic variability between the different *S*-alleles found outside the DRBDs could explain the specificity of the stigma *S* component and how *LpdsRNA* could interact with the pollen *S* component or even the *Z*-locus components.

Evidence for *LpSDUF247* as the pollen determinant of the *S* self-incompatibility locus in perennial ryegrass

From the analysis of the genome and transcriptome sequences, it appears that the *LpSDUF247* gene codes for the pollen component of the *S*-locus system. This gene, mapping in the region co-segregating with the *S*-locus in perennial ryegrass, is syntenic to the Brachypodium gene Bradi2g35750, located in the Brachypodium *S*-locus syntenic region and to the rice gene Os05g0198000 located outside the *S*-locus syntenic region, on rice chromosome 5 at around 6.03Mb. Its rice neighbouring gene, Os05g0198050, was described by Kakeda et al. (2008) as a homologue of a potential *S*-candidate gene in bulbous barley. Indeed, Kakeda et al. identified a marker HSP10, mapping onto rice Os05g0198050, which is specifically up-regulated in pistil (Kakeda, et al. 2008). In our study, *LpSDUF247* is relatively up-regulated in pollen tissue, though not highly expressed. This moderate expression is compatible with the observation that it is an intronless gene and that these often display low but tissue-specific expression patterns that are associated with signal transduction and fast acting regulatory pathways important for growth, proliferation and development (Grzybowska 2012), features matching the SI reaction. Additionally, the *LpSDUF247* protein is predicted to have a C-terminal transmembrane helix and an extracellular domain, indicating it may act as a ligand located on the pollen exine. In the sporophytic SI in *Brassica*, the pollen component which is a *S*-locus cysteine rich (SCR) protein, is acting as a ligand located in the pollen exine

(Schopfer, et al. 1999). Because of the cellular location of the *LpSDUF247* protein, the corresponding nuclear gene is not necessarily expected to show extensive expression at pollen maturity.

Another indication that *LpSDUF247* encodes the pollen *S* component is the very high allelic variability, as expected for polyallelic SI loci subjected to frequency-dependent selection (Schierup and Vekemans 2008). Based on this allelic diversity, *LpSDUF247* unambiguously differentiated and fully matched the expected segregation pattern of known *S*-locus genotypes in the seven mapping populations. *LpSDUF247* showed around 80% of sequence conservation at the protein level, comparing the five alleles that were available in this study. However, the high variability was not evenly distributed throughout the sequence and the transmembrane domain towards the C-terminus was more conserved. In the closely related species darnel, the transmembrane domain of the *LpSDUF247* ortholog has been lost through a frame-shift mutation and may give a first explanation of the self-compatibility in this species. Interestingly, all other self-compatible grass species for which sequence information for the *LpSDUF247* ortholog was available showed either frame-shift mutations leading to premature stop codons or large deletions in the predicted protein sequence. It is very intriguing that such loss-of-function mutations occurred in the male determinant candidate, as evolutionary theory predicts that self-compatible mutations in natural populations tend to be found in male components (Tsuchimatsu et al. 2012; Tsuchimatsu et al. 2010; Uyenoyama et al. 2001).

The nature of the stigma component of the *S*-locus system remains to be resolved. The low allelic diversity of any *S* gene except *LpSDUF247* either indicates that this gene may determine both the male and female specificity, or suggests the presence of the female specificity candidate at the sequence gab in close vicinity to *LpSDUF247*. Interestingly, in some other grasses, the orthologs of *LpSDUF247* have been found to be duplicated, as for example in rice, wild rice (*Oryza barthii*) and sorghum (*Sorghum bicolor* L.) (Figure S4). Further sequence analyses will resolve the presence of another copy of *LpSDUF247* or other stigma specificity candidates such as the *HSP10* homolog expressed at the barley stigma (Kakeda, et al. 2008) in the gap in the genomic sequence. A duplication of *LpSDUF247* could lead to the functional unit of the pollen and stigma component, by the creation of dimers. Recently, a similar gene encoding for a protein containing a DUF247 domain has been identified as a plausible candidate for the *Z* SI locus in perennial ryegrass (Shinozuka, et al. 2010). This supports the hypothesis of a duplicative origin of the two-locus SI system based

on pollen-stigma physiology and allelic interactions, proposed more than fifty years ago (Lundqvist 1962). However, the interaction between the *LpSDUF247* genes identified at the *S* and the *Z* locus remains to be characterized.

In conclusion, the present study provides multiple evidence that the *LpSDUF247* gene is the male determinant of the *S*-locus in the gametophytic SI system of perennial ryegrass. This is a major step forward towards understanding the SI response and constitutes the basis to isolate additional SI components. Moreover, it provides opportunities to study signatures of selection in the *S*-locus region in a wide set of self-incompatible and self-compatible ryegrass species. As gametophytic SI has been reported in both diploid and polyploid species within the tribes Triticeae, Poeae, and Paniceae, and seems to be monophyletic in the Poaceae family (Yang, et al. 2008), our work will be of major significance for evolutionary studies of speciation in some of the most important crop species. In the long run, we envisage, on the basis of our findings, modelling the SI cascade in Poaceae species. For self-incompatible Poaceae crop species, further understanding and identifying of target genes to efficiently regulate the SI mechanisms could simplify and accelerate breeding procedures in one of the largest, economically and ecologically most important plant families.

Material and Methods

Biparental mapping populations

Seven mapping populations of two different genetic backgrounds were used for the fine-mapping of the *S*-locus in perennial ryegrass, each designed to be heterozygous for the *S*-locus. The initial mapping population, the DTS population P235/59 as well as the two related fine-mapping populations P235/63 and P235/64, were derived from the perennial ryegrass ILGI mapping family (Jones, et al. 2002). All three mapping populations, called DTS populations, derived from crosses between siblings resulting from two crosses between ILGI population (P150/112) plants. A cross between the plant P150/112/129 ($S_{12}Z_{13}$) as female and plant P150/112/132 ($S_{12}Z_{12}$) as the pollinator resulted in an F1 population, referred to as P232/88. By the same method, a cross between the plant P150/112/41 ($S_{12}Z_{13}$) as female and plant P150/112/38 ($S_{13}Z_{13}$) as pollinator plant produced the F1 population P235/40. By crossing a plant from each of these two populations P232/88 and P235/40, the three populations P235/59, P235/63 and P235/64, were obtained.

The other four populations used for fine-mapping of the *S*-locus, the VrnA-S populations, are not related to the ILGI family and are derived from the VrnA mapping population originating from a cross between a Danish ecotype Falster and a genotype from the Italian variety Veyo (Jensen et al. 2005a). The two F1 plants of the VrnA population, F1-30 and NV#20F1-39 (hereafter referred to as F1-39), were selected and crossed in order to produce the F2 VrnA population. Finally, selected individuals of the F2 VrnA population were back-crossed with F1-39 in order to create the VrnA-S populations. In total, four different VrnA-S populations have been used for fine-mapping: VrnA-S216, VrnA-S267, VrnA-S324 and VrnA-S404.

DNA extraction methods

The DNA of 1393 plants from the population P235/59 was extracted using the DNeasy 96 Plant Kit (Qiagen, Hilden, Germany). The DNA of the other populations used for fine-mapping was extracted using an in-house protocol designed for high-throughput DNA extraction. Young leaf samples (approximately 15cm long) were collected in 96 well collection plates before being dried overnight in an oven at 60°C. Once dried, samples were ground with a Geno/Grinder 2000 (Spex SamplePrep, Metuchen, NJ) for 35 seconds at 1500 strokes/min. The plates were spun for 30 seconds at 3700rpm in order to collect the plant powder at the bottom and to avoid contamination. Using a Rainin Liquidator 96 Pipetting

System (Anachem Ltd., Luton, UK), 380µl of extraction buffer was added to each sample (0.2M NaCl, 0.05M Tris-HCl pH 8.0, 0.01M Na-EDTA, 0.01M DTT, 1M SDS). Samples were vortexed and incubated in a water-bath at 60°C for 10 minutes. Fifty µl of 5M K-acetate was added to each sample; the samples were vortexed again and placed on ice to incubate for 20 minutes. The plates were centrifuged for 5 minutes at 3700rpm and 125µl of the supernatant was pipetted into a new 96-well plate. One hundred µl of isopropanol was added to each sample and the plate was mixed by inversion before being incubated for 30 minutes at -20°C. The plate was then centrifuged for 15 minutes at 3000rpm and the isopropanol (supernatant) was discarded by pipetting. The samples were dried at 60°C in a vacuum-centrifuge (Eppendorf, Hamburg, Germany) for approximately 20 minutes (to allow the isopropanol to evaporate). Finally, the DNA was dissolved in 20µl of sterile distilled water and left for at least 4 hours on ice before being used.

Marker development

Publicly available DNA markers (Jensen, et al. 2005b; Jones, et al. 2002; Odonoughue, et al. 1992; Studer, et al. 2008; Yang, et al. 2009) were used to target the *S*-locus region in perennial ryegrass. Out of twenty markers tested on a subset of 242 plants from the DTS mapping populations P235/59, As_CDO1173 and Hv_BCD921 were used as anchors to establish the synteny between the *S*-locus region and the rice genome sequence. Alignment of the marker sequences and the rice genome sequence (RAP Build 3 of *Oriza sativa japonica*, NCBI) using BLASTN analysis identified the region between 1.5 to 5 Mbp on rice chromosome 5 as the *S* target region in rice.

In order to maximize the amplification and polymorphism rate of DNA markers in the mapping populations, perennial ryegrass-specific EST sequences were used for the development of markers suitable for HRM analysis. Using the CLC Genomics Workbench software (CLC bio, Aarhus, Denmark), an alignment between the rice sequence syntenic to the *S*-locus region and approximately 802,156 high quality cDNA reads (generated by Roche 454 sequencing) from various tissues of the perennial ryegrass genotypes F1-30 and F1-39 (Studer, et al. 2012) was conducted. Primers were designed in regions being conserved between rice and perennial ryegrass to amplify sequence polymorphisms between the cDNA reads of the two perennial ryegrass genotypes. Using the Primer3 software (Untergasser et al. 2012), a PCR product size of 80 to 150bp was targeted. Out of 40 markers developed and tested for polymorphism, 14 markers were screened in the mapping populations (Table 1 and Table S2).

Marker genotyping

The genotyping of the segregating populations was done using HRM. DNA was amplified in a 7µl PCR reaction using 1X LightScanner master mix (BioFire Diagnostics Inc., UT, USA), 0.3µM of each forward and reverse primer and 1µl of DNA (approximately 20ng/µl). An additional 14µl of mineral oil was added before the PCR to avoid sample evaporation during the melting curve analysis. PCR amplifications were conducted with a first step at 95°C for 2 minutes, followed by 40 cycles of 30 seconds at 94°C, 30 seconds at 63°C and 30 seconds at 72°C, with a final extension cycle for 2 minutes at 72°C. For optimal melting analysis, the PCR was finalized with a denaturation step at 94°C for 30 seconds followed by 30 seconds at 25°C. The melting of the PCR amplicons was done between 60 and 98°C using the 96-well LightScanner (BioFire Diagnostics Inc., UT, USA). For genotyping, the LightScanner® and Call-IT® software package (BioFire Diagnostics Inc., UT, USA) was used.

BAC library screen

Two perennial ryegrass BAC libraries, from the genotypes LTS18 and F1-30, have been used in this project (Farrar, et al. 2007). The average insert size of both libraries is approximately 100Kb and taken together, the two libraries represent almost ten-fold coverage of genome equivalents.

BAC clones covering the *S*-locus were selected by screening both BAC libraries with the *S*-locus flanking markers as well as the co-segregating markers. The screening was done by PCR using an amplification volume of 10µl containing 0.25 unit DreamTaq DNA Polymerase (Fermentas, St. Leon-Rot, Germany), 1X DreamTaq Buffer, 200µM dNTPs (Roche Applied Biosciences, Penzberg, Germany), 0.2µM of each forward and reverse primer and 1µl of the BAC DNA or bacteria clones. The amplification was done at 95°C for 4 min, followed by 30 cycles of 30 s at 94°C, 30 s at a primer-specific temperature and 1 min at 72°C, and finally 10 min at 72°C. The PCR products were then analysed on 1% agarose gels (1% agarose, 0.5X TAE) using 10% GelRed (Biotium, Hayward, CA, USA). Isolated BAC clones were used for DNA extraction according to the method described Farrar et al. (2007) and sequenced from both ends (M13 primers) using Sanger sequencing. Primers specific to each BAC end were developed and tested for specificity by PCR (as described above). The presence and absence of PCR amplicons on each BAC clone was used to order and orientate the isolated BAC clones.

Shotgun sequencing of the BAC clones covering the *S*-locus was done using the 454 GS FLX System (Roche Applied Biosciences, Basel, Switzerland). The nine BAC clones were pooled together into a quarter of a run, with an estimated coverage for each BAC clone of 120 times. The assembly of the reads into contigs was done using the software GS De Novo Assembler version 2.3 (Newbler, Roche, Basel, Switzerland). Using the CLC Genomics Workbench, the sequences were trimmed and filtered from any vector sequences (*Hind*III-digested pIndigoBAC-5 vector), sequencing adapters and bacterial genome sequences.

A draft assembly of the perennial ryegrass genome became available during this study, and we made use of it to identify assembled scaffolds that had hits to the BAC contigs covering the *S*-locus. This enabled us to identify five genomic scaffolds that spanned the region. These scaffolds were annotated using the MAKER 2 annotation pipeline (Holt and Yandell 2011). Briefly, an in house comprehensive transcriptome database enabled the prediction of genes directly from EST evidence. These initial predictions were used to develop a training file for the *de novo* gene predictor SNAP (Korf 2004), and a new round of gene prediction was performed. These new predictions were then used to generate an updated training file for SNAP and the gene prediction was repeated a final time.

Gene expression analysis

For the gene expression analysis, selected parental genotypes of the fine-mapping populations were used. These genotypes are known for their *S* allele composition and represent a diverse set of *S*-locus alleles. The plant F1-30, being heterozygous S_3S_4, was complemented with three other plants in order to increase the *S* allelic diversity; the plants P235/59/3 and P235/59/21 (heterozygous S_1S_2) from the population DTS population P235/59 and the plant P235/58/3 (homozygous S_1S_1) from the population P235/58 (reciprocal cross as P235/59).

Three biological replicates of pollen and stigma samples were collected from each of the different genotypes. The pollen was harvested into clear bags and transferred into microcentrifuge tubes; the pollen volume was approximately 0.1 ml per sample. The stigmas were collected from unopened florets with non-dehiscent anthers. Non-pollinated pistils at different stages of maturity were sampled and the ovary was removed using a razor blade. Approximately 50 pistils were sampled to create one stigma sample.

In addition to pollen and stigma tissue-specific samples, self-incompatible and self-compatible pollen-stigma interactions were analysed. Using *in-vivo* pollinations, three biological replicates from both self-pollinations (self-incompatible) and cross-pollinations

(self-compatible) were collected. The semi *in-vivo* pollinations were conducted as described by Thorogood et al. (2002). The genotype F1-30 was used for self-pollination, while F1-30 was pollinated with pollen from an unrelated genotype of the cultivar Foxtrot (DLF-Trifolium, Store Heddinge, Denmark) for the cross-pollination. In both cases, semi *in-vivo* pollinated pistils were collected approximately one hour after pollination, the ovaries were removed and the stigma samples were placed in liquid nitrogen.

Total RNA was extracted from the pollen, the self-incompatible and the self-compatible samples using the RNeasy™ Plant Mini Kit following the manufactures instructions (Qiagen, Valencia, CA). For the stigma samples which contained much less material, the RNA extraction was conducted using the PicoPure RNA Isolation kit (Arcturus Bioscience, Inc., CA, USA). RNA integrity was measured with an RNA 6000 Nano Labchip™ on the Agilent 2100 Bioanalyzer™ (Agilent Technologies, Santa Clara, CA).

The sequencing libraries were generated using the TruSeq Stranded mRNA Sample Preparation kit (Illumina, San Diego, USA) according to the manufactures instructions. The cDNA libraries were sequenced using Illumina HiSeq2000 sequencing and 76 bp paired-end reads were generated for each of them. All libraries were pooled into two groups and each group was distributed over two lanes, with an expectancy minimum of 30M reads/sample (AROS Applied Biotechnology AS, Aarhus, Denmark). The sequence data was uploaded to Sequence Read Archive (SRA) at NCBI under project (xxxxx, filled upon acceptance).

The alignment of the transcripts to the *S*-locus genomic sequence (scaffolds of the *Lolium* genome) was done using Bowtie 2, version 2.1.0 (Langmead and Salzberg 2012). The expression level between pollen and stigma tissues and between self-pollination and cross-pollination was done using R and Rsamtools as part of Bioconductor software package (Gentleman et al. 2004) to extract the read counts. EdgeR (Robinson et al. 2010) was used to calculate the biological coefficient of variation (BCV) in order to assess the variation between replicates and to calculate the differential expression between samples for each annotation. For each gene annotation, the logFC ($\log_2$ Fold Change; difference between tissues expression), the logCPM ($\log_2$ Count Per Million; normalised average between all samples) and P Value (exact test for the negative binomial distribution) were obtained. A normalization of the gene expression was also calculated using the RPKM (Reads per Kb of exon model) (Mortazavi et al. 2008).

Allelic diversity study

Allele reconstruction at genes contained in the *S*-locus region was done using the CLC Genomics Workbench. For expressed genes, the raw transcriptome reads from the different genotypes available were aligned to the predicted CDS of the genome scaffolds (hereafter referred to as the reference) for SNP detection (minimum read depth > 4). Using the raw transcriptome reads from the genotype P235/58/3 (homozygous S_1S_1), the allele S_1 was predicted. The allele S_2 was then predicted using the genotypes P235/59/3 and P235/59/21 (heterozygous S_1S_2). As for the alleles S_3 and S_4, the prediction was done using the genotype F1-30, separating the two alleles using reads long enough to cover two adjacent SNPs. Finally, an additional allele was predicted, S_5, using the BAC sequence covering the gene of interest.

Gene phylogenies and detailed allelic diversity measures were calculated based on CDS of the alleles S_1 to S_5, extracted for each gene co-segregating with the *S*-locus. Codon alignments were made using back translations via TranslatorX, version 1.1 (Abascal et al. 2010) where protein sequences were aligned using MAFFT, version 7.147b (Katoh and Standley 2013). Conserved regions in the alignments were selected by Gblocks, version, 0.91b (Castresana 2000) with a minimum block length of five. Parsing of Gblocks output from protein alignments to nucleotides alignments was done using TranslatorX. Phylogentic trees were constructed with PhyML, version 3.1 (Guindon et al. 2010) using nearest neighbour interchange and approximate likelihood ratio test returning Chi-squared-based parametric branch supports. Trees were rendered using the APE package in R statistical environment, version 3.2 (Paradis et al. 2004). For DNA polymorphism analyses, the software DNA Sequence Polymorphism DNASP, version 4.901, was used (Rozas 2009).

Acknowledgment

The authors would like to acknowledge Stephan Hentrup at Aarhus University for excellent technical support, Dr. Zeljko Micic from Deutsche Saatveredelung AG for helpful advises in the lab, Dr Susan Girdwood at IBERS for BAC clone sequencing and Dr Matthew Hegarty at IBERS for Illumina library preparation. We sincerely thank Dr. Ian Armstead for his very helpful comments on the manuscript and Prof. Dr. Achim Walter for hosting the Forage Crop Genetics group at ETH Zurich.

This work was supported by Teagasc Walsh Fellow PhD stipend, the Danish Council for Independent Research, Technology and Production Sciences (FTP, grant no: 09-065762), supported by the Swiss National Science Foundation (SNSF Professorship grant no: PP00P2 138988), BBSRC Institute Strategic Programme Grant (ref. BB/J004405/1) and the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no: GA-2010-267243 - PLANT FELLOWS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Abascal F, Zardoya R, Telford MJ 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Research 38: W7-W13.

Anderson MA, Cornish EC, Mau SL, Williams EG, Hoggart R, Atkinson A, Bonig I, Grego B, Simpson R, Roche PJ, Haley JD, Penschow JD, Niall HD, Tregear GW, Coghlan JP, Crawford RJ, Clarke AE 1986. Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. Nature 321: 38-44.

Asp T, Byrne S, Gundlach H, Bruggmann R, Mayer KFX, Andersen JR, Xu ML, Greve M, Lenk I, Lübberstedt T 2011. Comparative sequence analysis of *VRN1* alleles of *Lolium perenne* with the co-linear regions in barley, wheat, and rice. Molecular Genetics and Genomics 286: 433-447.

Bian XY, Friedrich A, Bai JR, Baumann U, Hayman DL, Barker SJ, Langridge P 2004. High-resolution mapping of the S and Z loci of *Phalaris coerulescens*. Genome 47: 918-930.

Boyes DC, Nasrallah JB 1995. An anther-specific gene encoded by an S locus haplotype of *Brassica* produces complementary and differentially regulated transcripts. Plant Cell 7: 1283-1294.

Byrne SL, Nagy I, Pfeifer M, Armstead I, Swain S, Studer B, Mayer K, Campbell JD, Czaban A, Hentrup S, Panitz F, Bendixen C, Hedegaard J, Caccamo M, Asp T 2015. A synteny-based draft genome sequence of the forage grass *Lolium perenne*. Plant Journal: in press.

Castresana J 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17: 540-552.

Charlesworth D, Willis JH 2009. The genetics of inbreeding depression. Nature Reviews Genetics 10: 783-796.

Cornish MA, Hayward MD, Lawrence MJ 1979. Self-incompatibility in ryegrass. I. Genetic control in diploid *Lolium perenne* L. Heredity 43: 95-106.

de Nettancourt D. 1977. Incompatibility in angiosperms. Springer, New York, NY.

Dooner HK, He LM 2008. Maize genome structure variation: Interplay between retrotransposon polymorphisms and genic recombination. Plant Cell 20: 249-258.

Durand E, Meheust R, Soucaze M, Goubet PM, Gallina S, Poux C, Fobis-Loisy I, Guillon E, Gaude T, Sarazin A, Figeac M, Prat E, Marande W, Berges H, Vekemans X, Billiard S, Castric V 2014. Dominance hierarchy arising from the evolution of a complex small RNA regulatory network. Science 346: 1200-1205.

Farrar K, Asp T, Lübberstedt T, Xu ML, Thomas AM, Christiansen C, Humphreys MO, Donnison IS 2007. Construction of two *Lolium perenne* BAC libraries and identification of BACs containing candidate genes for disease resistance and forage quality. Molecular Breeding 19: 15-23.

Fierro-Monti I, Mathews MB 2000. Proteins binding to duplexed RNA: one motif, multiple functions. Trends in Biochemical Sciences 25: 241-246.

Foote HCC 1994. Cloning and expression of a distinctive class of self-incompatibility (S) gene from P*apaver rhoeas* L. Proc Natl Acad Sci USA 91: 2265-2269.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge YC, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang JH 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biology 5: R80.

Grzybowska EA 2012. Human intronless genes: functional groups, associated diseases, evolution, and mRNA processing in absence of splicing. Biochem Biophys Res Commun 424: 1-6.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology 59: 307-321.

Hayman D 1956. The genetical control of incompatibility in *Phalaris Coerulescens* Desf. Australian Journal of Biological Sciences 9: 321-331.

Henderson IR 2012. Control of meiotic recombination frequency in plant genomes. Current Opinion in Plant Biology 15: 556-561.

Hinata K, Nishio T 1978. *S*-allele specificity of stigma proteins in *Brassica oleracea* and *B. campestris*. Heredity 41: 93-100.

Holt C, Yandell M 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC bioinformatics 12: 491.

Isokawa S, Osaka M, Shirasawa A, Kikuta R, Komatsu S, Horisaki A, Niikura S, Takada Y, Shiba H, Isogai A, Takayama S, Suzuki G, Suwabe K, Watanabe M 2010. Novel self-compatible lines of *Brassica rapa* L. isolated from the Japanese bulk-populations. Genes Genet Syst 85: 87-96.

Jensen LB, Andersen JR, Frei U, Xing Y, Taylor C, Holm PB, Lübberstedt T 2005a. QTL mapping of vernalization response in perennial ryegrass (*Lolium perenne* L.) reveals co-location with an orthologue of wheat *VRN1*. Theoretical and Applied Genetics 110: 527-536.

Jensen LB, Muylle H, Arens P, Andersen CH, Holm PB, Ghesquière M, Julier B, Lübberstedt T, Nielsen KK, Riek JD, Roldán-Ruiz I, Roulund N, Taylor C, Vosman B, Barre P 2005b. Development and mapping of a public reference set of SSR markers in *Lolium perenne* L. Molecular Ecology Notes 5: 951-957.

Jones ES, Mahoney NL, Hayward MD, Armstead IP, Jones JG, Humphreys MO, King IP, Kishida T, Yamada T, Balfourier F, Charmet G, Forster JW 2002. An enhanced molecular marker based genetic map of perennial ryegrass (*Lolium perenne*) reveals comparative relationships with other Poaceae genomes. Genome 45: 282-295.

Kakeda K 2009. S locus-linked F-box genes expressed in anthers of *Hordeum bulbosum*. Plant Cell Reports 28: 1453-1460.

Kakeda K, Ibuki T, Suzuki J, Tadano H, Kurita Y, Hanai Y, Kowyama Y 2008. Molecular and genetic characterization of the S locus in *Hordeum bulbosum* L., a wild self-incompatible species related to cultivated barley. Molecular Genetics and Genomics 280: 509-519.

Katoh K, Standley DM 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. Molecular Biology and Evolution 30: 772-780.

Korf I 2004. Gene finding in novel genomes. BMC bioinformatics 5: 59.

Krovat BC, Jantsch MF 1996. Comparative mutational analysis of the double-stranded RNA binding domains of *Xenopus laevis* RNA-binding protein A. Journal Of Biological Chemistry 271: 28112-28119.

Kubo K-i, Entani T, Takara A, Wang N, Fields AM, Hua Z, Toyoda M, Kawashima S-i, Ando T, Isogai A, Kao T-h, Takayama S 2010. Collaborative non-self recognition system in S-RNase-based self-incompatibility. Science 330: 796-799.

Kubo K-i, Paape T, Hatakeyama M, Entani T, Takara A, Kajihara K, Tsukahara M, Shimizu-Inatsugi R, Shimizu KK, Takayama S 2015. Gene duplication and genetic exchange drive the evolution of S-RNase-based self-incompatibility in *Petunia*. Nature Plants 1: 14005.

Lai Z, Ma WS, Han B, Liang LZ, Zhang YS, Hong GF, Xue YB 2002. An F-box gene linked to the self-incompatibility (*S*) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. Plant Molecular Biology 50: 29-42.

Langmead B, Salzberg SL 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357-359.

Li XM, Paech N, Nield J, Hayman D, Langridge P 1997. Self-incompatibility in the grasses: Evolutionary relationship of the S gene from *Phalaris coerulescens* to homologous sequences in other grasses. Plant Molecular Biology 34: 223-232.

Lundqvist A 1962. The nature of the two-loci incompatibility system in grasses. I. The hypothesis of a duplicative origin. Hereditas 48: 153-168.

Lundqvist A 1954. Studies on self-sterility in rye, *Secale cereale* L. Hereditas 40: 278-294.

Martin RC, Hollenbeck VG, Dombrowski JE 2008. Evaluation of reference genes for quantitative RT-PCR in *Lolium perenne*. Crop Science 48: 1881-1887.

Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5: 621-628.

Nasrallah JB 2002. Recognition and rejection of self in plant reproduction. Developmental Biology 247: 523-523.

Odonoughue LS, Wang Z, Roder M, Kneen B, Leggett M, Sorrells ME, Tanksley SD 1992. An RFLP-based linkage map of oats based on a cross between 2 diploid taxa (*Avena atlantica* X *A. hirtula*). Genome 35: 765-771.

Paradis E, Claude J, Strimmer K 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20: 289-290.

Pedersen C, Giese H, Linde-Laursen I 1995. Towards an integration of the physical and the genetic chromosome maps of barley by in situ hybridization. Hereditas 123: 77-88.

Pfeifer M, Martis M, Asp T, Mayer KFX, Lübberstedt T, Byrne S, Frei U, Studer B 2013. The perennial ryegrass GenomeZipper: targeted use of genome resources for comparative grass genomics. Plant Physiology 161: 571-582.

R Development Core Team 2008. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Robinson MD, McCarthy DJ, Smyth GK 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140.

Rozas J 2009. DNA sequence polymorphism analysis using DnaSP. Bioinformatics for DNA Sequence Analysis 537: 337-350.

Rudd JJ, Osman K, Franklin FCH, Franklin-Tong VE 2003. Activation of a putative MAP kinase in pollen is stimulated by the self-incompatibility (SI) response. FEBS Letters 547: 223-227.

Santner A, Estelle M 2009. Recent advances and emerging trends in plant hormone signalling. Nature 459: 1071-1078.

Saunders LR, Barber GN 2003. The dsRNA binding protein family: critical roles, diverse cellular functions. FASEB J 17: 961-983.

Schierup MH, Vekemans X 2008. Genomic consequences of selection on self-incompatibility genes. Current Opinion in Plant Biology 11: 116-122.

Schopfer CR, Nasrallah ME, Nasrallah JB 1999. The male determinant of self-incompatibility in *Brassica*. Science 286: 1697-1700.

Shinozuka H, Cogan NOI, Smith KF, Spangenberg GC, Forster JW 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 Molecular Biology 72: 343-355.

Sim S, Chang T, Curley J, Warnke SE, Barker RE, Jung G 2005. Chromosomal rearrangements differentiating the ryegrass genome from the Triticeae, oat, and rice genomes using common heterologous RFLP probes. Theoretical and Applied Genetics 110: 1011-1019.

Sonneveld T, Tobutt KR, Vaughan SP, Robbins TP 2005. Loss of pollen-S function in two self-compatible selections of *Prunus avium* is associated with deletion/mutation of an S haplotype-specific F-box gene. Plant Cell 17: 37-51.

Stein JC, Howlett B, Boyes DC, Nasrallah ME, Nasrallah JB 1991. Molecular-cloning of a putative receptor protein-kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. Proc Natl Acad Sci USA 88: 8816-8820.

Studer B, Asp T, Frei U, Hentrup S, Meally H, Guillard A, Barth S, Muylle H, Roldán-Ruiz I, Barre P, Koning-Boucoiran C, Uenk-Stunnenberg G, Dolstra O, Skøt L, Skøt K, Turner L, Humphreys M, Kölliker R, Roulund N, Nielsen KK, Lübberstedt T 2008. Expressed sequence tag-derived microsatellite markers of perennial ryegrass (*Lolium perenne* L.). Molecular Breeding 21: 533-548.

Studer B, Byrne S, Nielsen RO, Panitz F, Bendixen C, Islam MS, Pfeifer M, Lübberstedt T, Asp T 2012. A transcriptome map of perennial ryegrass (*Lolium perenne* L.). BMC Genomics 13: 140.

Studer B, Jensen LB, Fiil A, Asp T 2009. “Blind” mapping of genic DNA sequence polymorphisms in *Lolium perenne* L. by high resolution melting curve analysis. Molecular Breeding 24: 191-199.

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

Tantikanjana T, Nasrallah JB 2012. Non-cell-autonomous regulation of crucifer self-incompatibility by Auxin Response Factor ARF3. Proc Natl Acad Sci USA 109: 19468-19473.

Thorogood D, Armstead IP, Turner LB, Humphreys MO, Hayward MD 2005. Identification and mode of action of self-compatibility loci in *Lolium perenne* L. Heredity 94: 356-363.

Thorogood D, Kaiser WJ, Jones JG, Armstead I 2002. Self-incompatibility in ryegrass 12. Genotyping and mapping the *S* and *Z* loci of *Lolium perenne* L. Heredity 88: 385-390.

Tsuchimatsu T, Kaiser P, Yew CL, Bachelier JB, Shimizu KK 2012. Recent loss of self-incompatibility by degradation of the male component in allotetraploid *Arabidopsis kamchatica*. PLoS Genetics 8.

Tsuchimatsu T, Suwabe K, Shimizu-Inatsugi R, Isokawa S, Pavlidis P, Stadler T, Suzuki G, Takayama S, Watanabe M, Shimizu KK 2010. Evolution of self-compatibility in *Arabidopsis* by a mutation in the male specificity gene. Nature 464: 1342-1346.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG 2012. Primer3-new capabilities and interfaces. Nucleic Acids Research 40: e115.

Uyenoyama MK, Zhang Y, Newbigin E 2001. On the origin of self-incompatibility haplotypes: Transition through self-compatible intermediates. Genetics 157: 1805-1817.

Voylokov AV, Korzun V, Borner A 1998. Mapping of three self-fertility mutations in rye (*Secale cereale* L.) using RFLP, isozyme and morphological markers. Theoretical and Applied Genetics 97: 147-153.

Wang Y, Wang X, Skirpan AL, Kao TH 2003. S-RNase-mediated self-incompatibility. Journal of Experimental Botany 54: 115-122.

Wheeler MJ, de Graaf BHJ, Hadjiosif N, Perry RM, Poulter NS, Osman K, Vatovec S, Harper A, Franklin FCH, Franklin-Tong VE 2009. Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. Nature 459: 992-995.

Williams CG, Goodman MM, Stuber CW 1995. Comparative recombination distances among *Zea mays* L inbreds, wide crosses and interspecific hybrids. Genetics 141: 1573-1581.

Yang B, Thorogood D, Armstead I, Barth S 2008. How far are we from unravelling self-incompatibility in grasses? New Phytologist 178: 740-753.

Yang B, Thorogood D, Armstead I, Franklin F, Barth S 2009. Identification of genes expressed during the self-incompatibility response in perennial ryegrass (*Lolium perenne* L.). Plant Molecular Biology 70: 709-723.

Figures

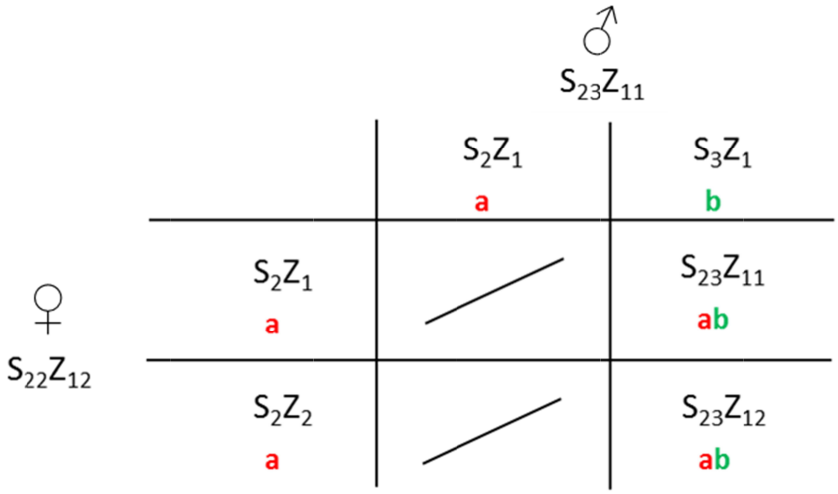

Figure 1. Design of the biparental mapping populations segregating for the *S*-locus in perennial ryegrass (*Lolium perenne* L.)

The parental genotypes, a plant with the *S* allele composition S_2S_3 as the male and S_2S_2 as the female parent, are sharing one *S* allele (S_2). Two *S* pollen gamete genotypes are produced yet only the S_3 gamete is transmitted to produce heterozygous S_2S_3 progeny as the S_2 pollen gamete is incompatible with the female parent. With this population design, any marker closely linked to the *S*-locus will also be heterozygous (ab) in the mapping population while recombinants and selfs will be homozygous (aa).

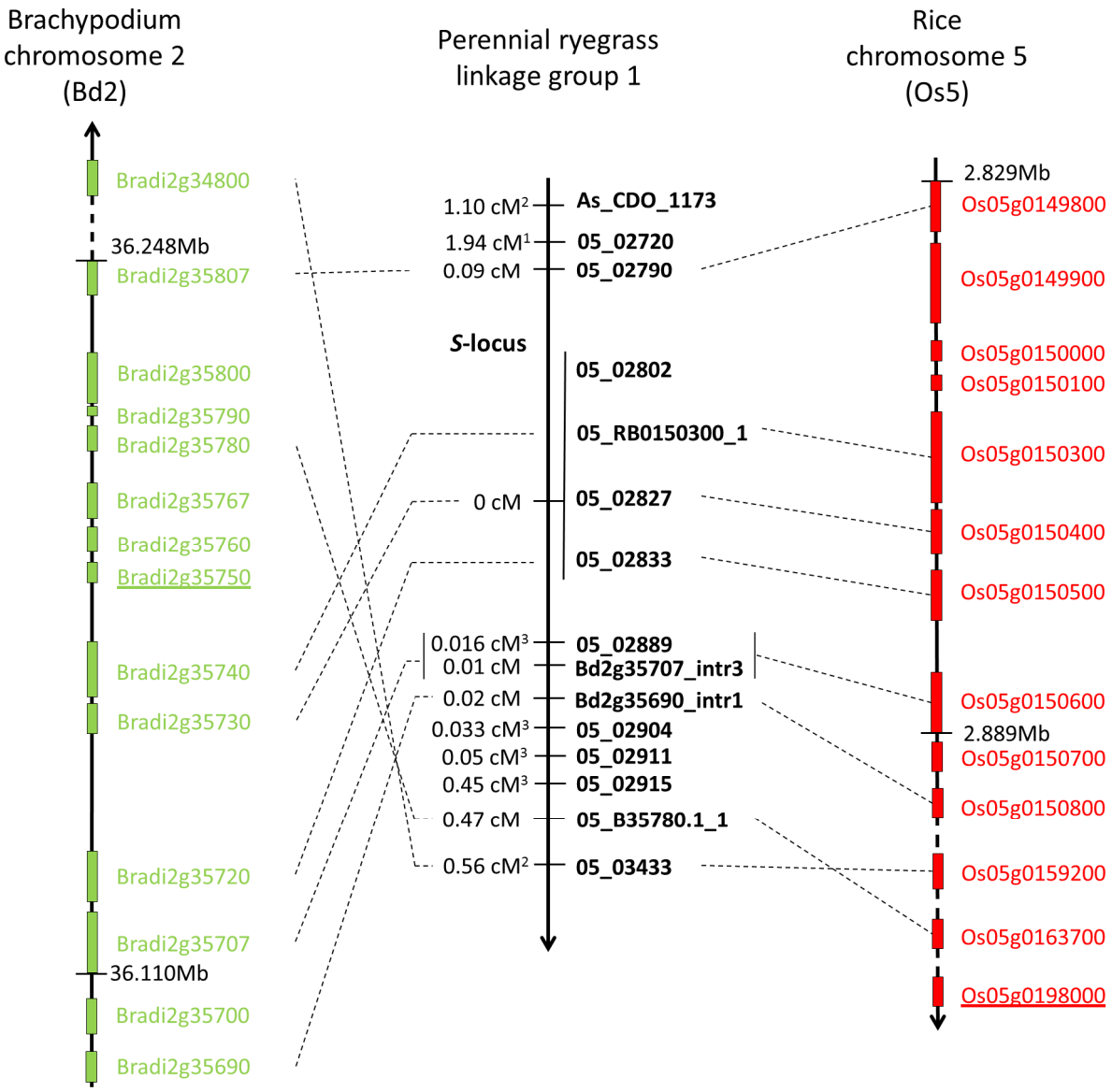

Figure 2. Genome synteny map between Brachypodium (*Brachypodium distachyon* L.), perennial ryegrass (*Lolium perenne* L.) and rice (*Oryza sativa* L.) at the *S*-locus region

The perennial ryegrass linkage group 1 (middle) is represented with the markers used for fine-mapping. The genetic distances were calculated from the recombination frequency out of 10,177 plants, for the exception of ([1]) calculated out of 1393 plants (DTS P235/59 population), ([2]) out of 6048 plants (all DTS populations) and ([3]) out of 8784 plants (P235/63, P235/64 and VrnA-S populations). The genome region syntenic to the perennial ryegrass *S*-locus region on Brachypodium chromosome 2 (left) and rice chromosome 5 (right) are represented with the genes and their physical distances. Genome synteny between the three species is indicated by thin dashed lines. Large distances on the chromosomes are represented by a dashed line and the chromosome orientation is shown by the arrow on the bar.

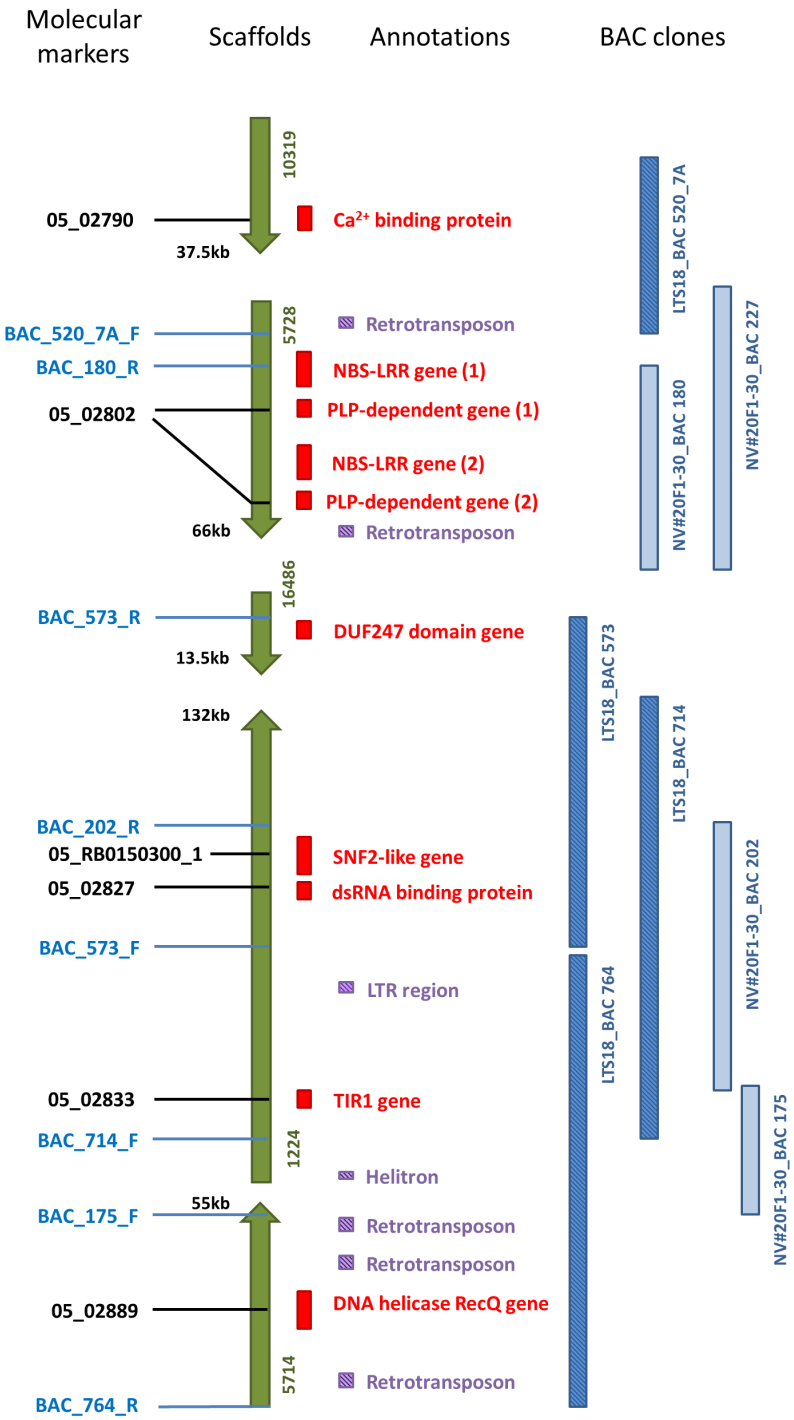

Figure 3. Graphic representation of the genome region at the *S*-locus in perennial ryegrass (*Lolium perenne* L.)

The *S*-locus region is represented with five perennial ryegrass genome scaffolds (green bars) onto which the BAC clones (blue bars) pulled out with the markers used for fine-mapping (black) were aligned. The genes (red bars) contained at the *S*-locus were placed on the scaffolds as well as the repetitive elements (purple bars) and the BAC-end markers (blue). The sizes of the scaffolds are indicated in black in kilobase pairs (kb).

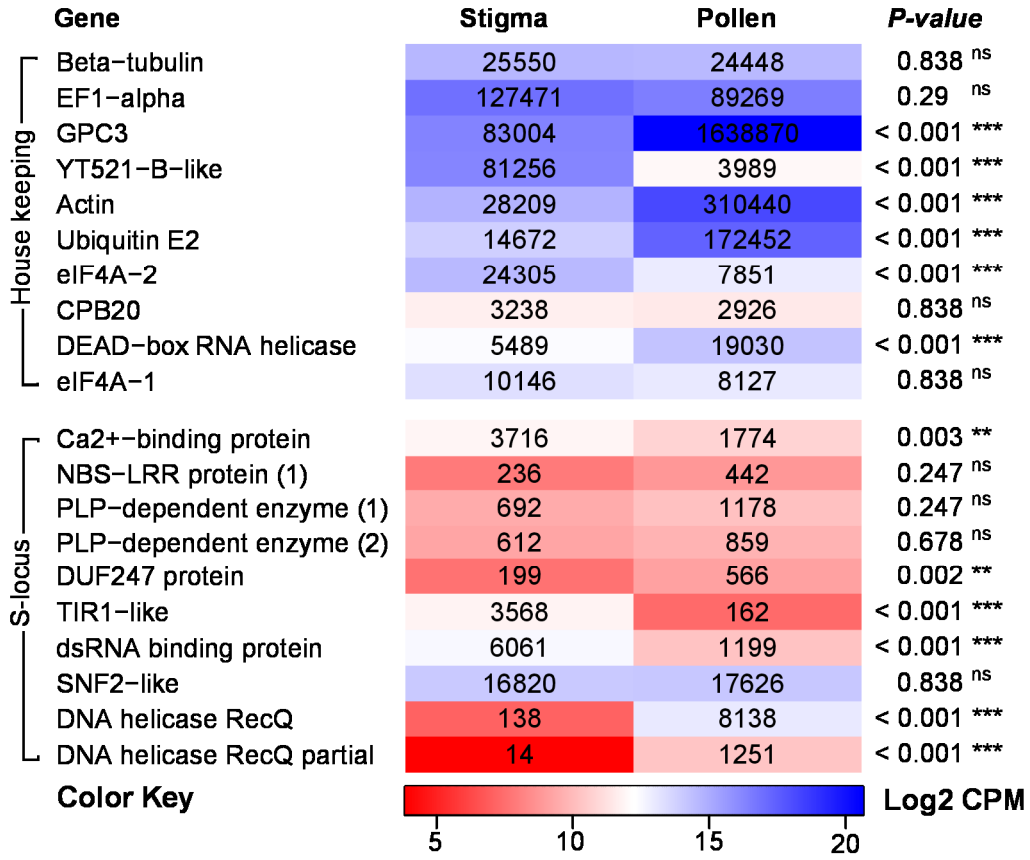

Figure 4. Heatmap of the gene expression at the *S*-locus for pollen and stigma tissue samples of perennial ryegrass (*Lolium perenne* L.)

The figure is representing the expression for each gene annotated at the *S*-locus as well as ten housekeeping genes for both pollen and stigma tissues. For each annotated gene and tissue, the expression is represented by the number of reads per millions (CPM) and translated into a colour key using $\log_2$CPM.

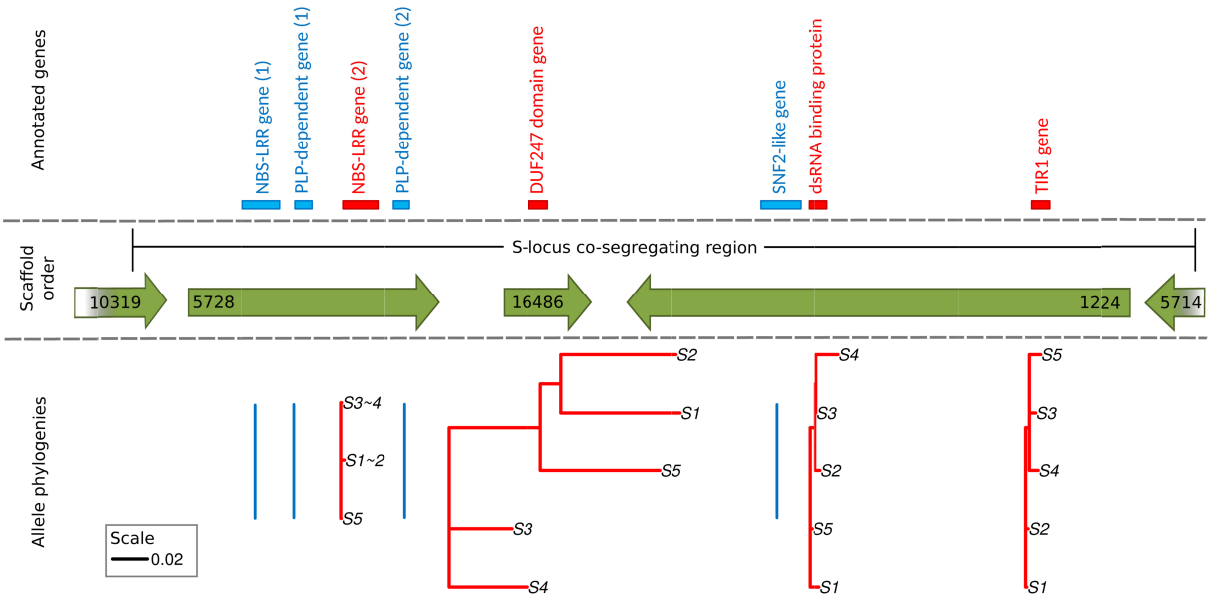

Figure 5. Gene phylogenies in the genome region co-segregating with the *S*-locus in perennial ryegrass (*Lolium perenne* L.)

The annotated genes co-segregating with the *S*-locus in perennial ryegrass and their relative positions are given based on ordered genome scaffold information as shown in Figure 3. The lower section shows gene phylogenies constructed for each gene using coding sequences of the alleles S_1 to S_5. Genes given in blue indicate the absence of allelic diversity, represented by a single blue bar. Genes with more than one allele are shown in red, their corresponding phylogney is shown below. The scale shows the number of nucleotide changes per site.

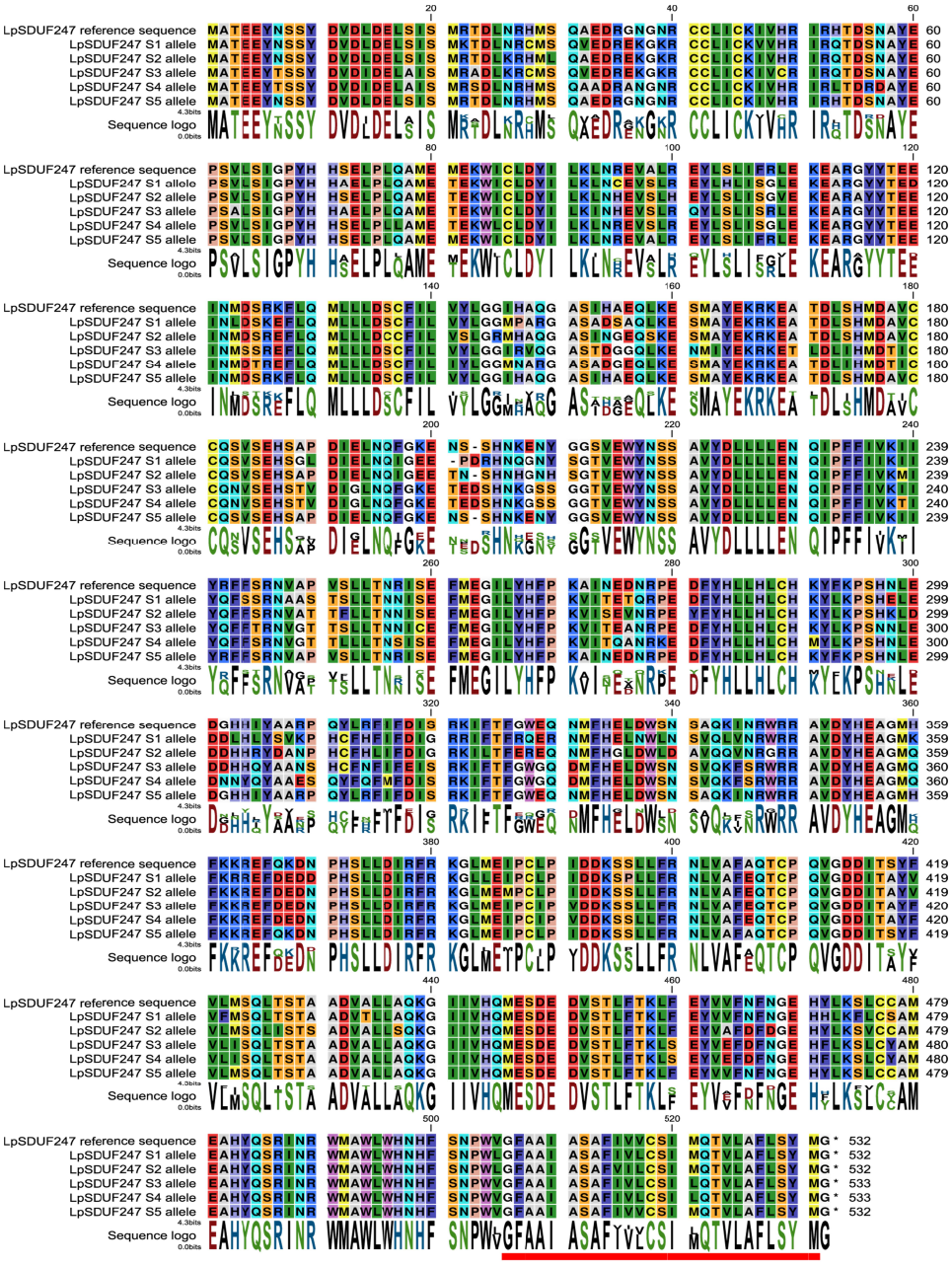

Figure 6. Protein sequence alignment of different LpSDUF247 alleles including the protein predicted from the perennial ryegrass reference genome sequence (LpSDUF247_ref)

The protein sequences were reconstructed from RNA sequencing data of different genotypes: LpSDUF247_S_1 from the genotype P235/58/3, LpSDUF247_S_2 from the genotype P235/59/3 and P235/59/21, LpSDUF247_S_3 and LpSDUF247_S_4 from the genotype F1-30 and LpSDUF247_S_5 from the LTS18-BAC clone 573. The predicted transmembrane domain of the LpSDUF247 protein is underlined in red.

Table 1. Number and distribution of recombination events between DNA markers at the *S*-locus region in the different mapping populations.

	DTS populations			VrnA-S populations					
Marker name	**P235/59** (1393 plants)	**P235/63** (1769 plants)	**P235/64** (2886 plants)	**VrnA-S216** (1425 plants)	**VrnA-S264** (964 plants)	**VrnA-S324** (746 plants)	**VrnA-S404** (994 plants)	**Total recombinants** (10177 plants)	***S*-locus region**
As_CDO_1173	-	13	27	22	10	11	14	97[b]	
05_02720	27	-	-	-	-	-	-	-	
05_02790	3	2	2	1	1	0	0	9	
05_02802	0	0	0	0	0	0	0	0	*S*
05_RB0150300_1	0	0	0	0	0	0	0	0	*S*
05_02827	0	0	0	0	0	0	0	0	*S*
05_02833	0	0	0	0	0	0	0	0	*S*
05_02889	1	0	0	M	M	M	M	1[a]	
Bd2g35707_intr3	1	0	0	0	0	0	0	1	
Bd2g35690_intr1	1	0	0	0	0	0	1	2	
05_02904	2	0	0	M	M	M	M	2[a]	
05_02911	3	0	0	M	M	M	M	3[a]	
05_02915	21	4	2	M	M	M	M	27[a]	
05_B35780_1	21	14	5	0	6	0	2	48	
05_03433	-	14	5	0	6	0	3	28[b]	

Fifteen polymorphic markers covering the *S*-locus in perennial ryegrass (*Lolium perenne* L.) were genotyped in the biparental mapping populations using high resolution melting curve analysis. For each population, the number of recombinants is indicated at each marker locus unless the marker is monomorphic (M). The markers co-segregating with the *S*-locus are indicated with an *S*. "[a]" indicates the number of recombinants out of 6048 plants from the DTS populations only, "[b]" indicates the number of recombinants out of 8784 plants from the P235/63, P235/64 and VrnA-S populations only and "–" indicates missing data.

Table 2: Summary of the annotation of the genome region at the *S*-locus in perennial ryegrass (*Lolium perenne* L.).

	Annotation position		Gene orthologs				
Scaffold name	Start (bp)	End (bp)	*Oryza sativa*	*Brachypodium distachyon*	Gene description (NCBI)	*Lolium* gene names	*S*-locus region
10319	31,476	26,240	Os05g0149800	Bradi2g35807	Ca^{2+} binding protein (EF-Hand superfamily)		
5728	22,894	26,919	Os12g0481400 Os12g0481700	Bradi2g35767	Putative NBS-LRR disease resistance protein RGA4-like (1)		*S*
5728	33,238	31,481	Os05g0150000	Bradi2g35760	Pyridoxal 5-phosphate (PLP) dependent enzyme (1)		*S*
5728	43,359	48,613	Os12g0481400 Os12g0481700	Bradi2g35767	Putative NBS-LRR disease resistance protein RGA4-like (2)	*LpNBS-LRR(2)*	*S*
5728	54,849	53,139	Os05g0150000	Bradi2g35760	Pyridoxal 5-phosphate (PLP) dependent enzyme (2)		*S*
16486	8,928	7,372	Os05g0198000	Bradi2g35750	Protein of unknown function DUF247	*LpSDUF247*	*S*
1224	92,615	83,803	Os05g0150300	Bradi2g35740	Chromatin-remodelling complex ATPase chain, transcription factor SNF2-like		*S*
1224	78,596	83,083	Os05g0150400	Bradi2g35730	Protein containing a double-stranded RNA binding motif	*LpdsRNA*	*S*
1224	19,871	24,673	Os05g0150500	Bradi2g35720	Transport inhibitor response 1 (TIR1) containing F-box domain, part of the SCF ubiquitin ligase complex	*LpTIR1*	*S*
5714	29,223	22,792	Os05g0150600	Bradi2g35707	ATP-dependent DNA helicase, RecQ family protein		
5714	30,447	29,792	Os05g0150600	Bradi2g35707	ATP-dependent DNA helicase, RecQ family protein (partial)		

The table details the annotation of the ryegrass genome scaffolds building the *S*-locus region. For each scaffold, the position of the predicted genes is given as well as the gene orthologs from rice (*Oryza sativa* L.) and Brachypodium (*Brachypodium distachyon* L.). A description of the function, derived from the rice ortholog (NCBI), is given for each gene. *S* indicates a gene co-segregating with the *S*-locus.