Expression of ten S class SLF-like genes in *Nicotiana alata* pollen and its implications for understanding the pollen factor of the S locus.

David Wheeler and Ed Newbigin

Plant Cell Biology Research Centre

School of Botany

University of Melbourne,

Victoria 3010, Australia

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. EF420251-EF4202510.
RUNNING HEAD: *SLF-like* genes from *Nicotiana alata*

KEY WORDS: Self-incompatibility, pollen S, S locus F-box protein, balancing selection, Solanaceae

For correspondence
Ed Newbigin
Plant Cell Biology Research Centre
School of Botany
University of Melbourne Victoria 3010 Australia
Email: edwardjn@unimelb.edu.au
Phone: (+61) 3 8344 4871
Fax: (+61) 3 9347 1071
ABSTRACT

The S locus of *Nicotiana alata* encodes a polymorphic series of ribonucleases (S-RNases) that determine the self-incompatibility (SI) phenotype of the style. The pollen product of the S locus (pollen S) in *N. alata* is unknown, but in species from the related genus *Petunia* and in self-incompatible members of the Plantaginaceae and Rosaceae, this function has been assigned to an F-box protein known as SLF or SFB. Here we describe the identification of 10 genes (designated DD1-10) encoding SLF-related proteins that are expressed in *N. alata* pollen. Because our approach to cloning the DD genes was based on sequences of SLFs from other species, we presume that one of the DD genes encodes the *N. alata* SLF ortholog. Seven of the DD genes were exclusively expressed in pollen and a low level of sequence variation was found in alleles of each DD gene. Mapping studies confirmed that all 10 DD genes were linked to the S locus and that at least three were located in the same chromosomal segment as pollen S. Finally, the different topologies of the phylogenetic trees produced using available S class sequences and those produced using S-RNase sequences suggests pollen S and the S-RNase have different evolutionary histories.
INTRODUCTION

Self-incompatibility (SI) is a genetic mechanism found in many flowering plants that acts to reduce the negative fitness effects associated with inbreeding (DE NETTANCOURT 2001). In the best understood examples, SI is controlled by a single genetic S locus with many S alleles: styles reject pollen grains when both express the same S specificity. Although SI systems have evolved independently several times during the diversification of flowering plants (MATTON et al. 1994), in three families, the Solanaceae, Plantaginaceae and Rosaceae, the SI phenotype of the style (but not of pollen) is controlled by an extracellular ribonuclease known as the S-RNase (ANDERSON et al. 1986; MCCLURE et al. 1989; SASSA et al. 1993; XUE et al. 1996), a feature that suggests a single evolutionary origin for RNase-based SI systems (IGIC and KOHN 2001). Mechanistically, S-RNases are thought to act as cytotoxins that can enter both compatible and incompatible pollen tubes (LUU et al. 2000; GOLDRAIJ et al. 2006;), retarding the growth of incompatible pollen tubes possibly by degrading the limited amount of ribosomal RNA they contain (MCCLURE et al. 1990).

The S locus pollen factor, known as pollen S, was until recently unknown. The first indication of its identity came from Lai and colleagues (2002), who sequenced a 63 kb region of Antirrhinum hispanicum (Plantaginaceae) genomic DNA that contained the S-RNase gene, and found within this sequence an F-box gene specifically expressed in pollen. F-box genes were soon identified at the S loci of Petunia (Solanaceae) (QIAO et al. 2004a; WANG et al. 2004a; TSUKAMOTO et al. 2005) and Prunus, Pyrus and Malus (Rosaceae) (ENTANI et al. 2003; USHIJIMA et al. 2003; SASSA et al. 2007), and
collectively these genes are now called the *S Locus F-box* genes (abbreviated as SLF for the Solanaceae and Plantaginaceae and as SFB for the Rosaceae). Functionally, F-box proteins are a component of one type of E3 ubiquitin-protein ligase, the enzyme responsible for transferring ubiquitin (Ub) from E2 Ub-conjugating enzyme to free Lys residues on a selected protein target (VIERSTRA 2003). This association with protein ubiquitination has led to a model in which compatible pollen tubes are protected from S-RNase cytotoxicity by a SLF-mediated mechanism involving Ub attachment and degradation (LAI et al. 2002; QIAO et al. 2004b).

In each family, evidence exists to show that *SLF*/SFB encodes pollen S. In the Rosaceae, for instance, where producing transgenic plants is difficult, statistical tests of amino acid variation between different *Prunus SFB* alleles have identified two short regions of sequence that are under balancing selection (IKEDA et al. 2004; NUNES et al. 2006). SI is one of the best documented examples of balancing selection and the way it affects nucleotide sequence variation has previously been used to identify genes involved in SI (TAKEBAYASHI et al. 2003). Also consistent with a role for SFB in SI are truncations or deletions of this gene in self-compatible cultivars of fruit trees such as *Prunus avium* (sweet cherry) and *P. mume* (Japanese apricot) that carry mutations specifically affecting the SI phenotype of pollen (USHIJIMA et al. 2004; SONNEVELD et al. 2005;). For the Solanaceae and Plantaginaceae, transgenic experiments have been used to show that *SLF* encodes a protein that behaves in the manner expected of *pollen S* (QIAO et al. 2004a; SIJACIC et al. 2004). A *SLF* transgene does not cause pollen from a transgenic plant to express a new S allele identity, but instead causes a breakdown of SI in pollen if the *SLF* transgene is derived from an S allele different from the plant’s
own S alleles (QIAO et al. 2004a; SIJACIC et al. 2004). Thus, the SLF transgene derived from the Petunia inflata S2 allele (PiSLF2) caused loss of SI expression in pollen carrying either the S1 or S3 allele but not in pollen carrying the S2 allele (SIJACIC et al. 2004). The ability of two different pollen S alleles present in the same pollen grain to “competitively interact” with each other and thus allow the pollen tube to grow through an otherwise incompatible style, is well known, having previously been observed in tetraploids derived from self-incompatible diploids (e.g., CHAWLA et al. 1997). SI plants with pollen-part mutations (pollen-part mutants, PPMs) are another example of competitive interaction between pollen S alleles. PPMs are diploid plants with an extra copy of the S locus that is present either as a small additional chromosome called a centric fragment, as a translocated fragment attached to a non-homologous (i.e., non S-bearing) chromosome or as a fragment attached to a homologous (S-bearing chromosome) chromosome via unequal exchange (GOLZ et al. 1999, 2001).

Genes related to SLF/SFB but with no known role in SI, also occur in plant genomes, including those of self-compatible plants lacking an RNase-based SI mechanism (WANG et al. 2004b). These genes, referred to here as SLF-like genes, are members of a large family of F-box protein genes found only in plants (WANG et al. 2004b). Except for SLF/SFB, functions for most members of this family are unknown, even though it includes approximately 13% of the F-box protein genes (92/694) in Arabidopsis thaliana (WANG et al. 2004b). Phylogenetic analysis divides the SLF-like gene family into 5 different classes (A, B, C, M and S), with all SLF/SFB genes and some SLF-like genes belonging to class S (WANG et al., 2004b). The presence of SLF-like genes can complicate the process of identifying SLF/SFB genes, especially when SLF-like genes...
are pollen-expressed and linked to the S locus (Entani et al. 2003; Ushijima et al. 2003; Zhou et al. 2003; Wang et al. 2004a; Sassa et al. 2007).

In this paper we describe an analysis of the pollen-expressed class S SLF-like genes from the solanaceous self-incompatible plant Nicotiana alata. A motivation for this study was to broaden the range of solanaceous plants for which information on SLF genes is available, as all existing sequences are from Petunia (Qiao et al. 2004a; Sijacic et al. 2004; Tsukamoto et al. 2005). Using degenerate primers and a reverse transcriptase PCR (RT-PCR) approach with N. alata pollen RNA as the template, full-length sequences for 10 class S genes were obtained. Because potentially one of these 10 genes is the N. alata SLF ortholog, we gave them the temporary names DD1-10. Of the 10 DD genes, seven were exclusively expressed in pollen. Polymorphisms in 7 DD genes were used to show linkage to the S locus: placement of individual DD genes with regard to the S-RNase gene and surrounding S locus markers was done using the PPM plants described in Golz et al. (2001). Finally, phylogenetic analysis using DD and other class S sequences suggests that the SLFs in the Solanaceae and Plantaginaceae do not have a long history of co-evolving with the S-RNase gene. We discuss the implications this has for identifying pollen S in these families, and for understanding the evolutionary history of the pollen and stylar genes of the S locus.

MATERIALS AND METHODS

Plant material: SI lines of N. alata homozygous for S₁, S₂, S₃, S₆ or S₇ were maintained as described in Anderson et al. (1986). The collection of N. alata pollen-
part mutant plants (PPMs) homozygous for the S₆ allele and with mapped duplications containing the S₃ allele have been described previously (Golz et al. 1999, 2001). *N. alata* families segregating for 4 S alleles were produced by germinating S₁S₃ × S₆S₇ seed and S₂S₃ × S₆S₇ seed. S genotypes were determined by PCR amplification of S-RNase genes as described by Li et al. (2000).

**Primer design and PCR:** The primers D2 5’-YTIATIGGICITGYRAYGG-3’ (forward) and D4 5’-CICCRTAYTSIWTCATNAYCC-3’ (reverse) were designed based on regions conserved in *SLF* genes from the Solanaceae and Plantaginaceae. PCR was performed with approximately 50 ng of genomic DNA or 1 µl of cDNA template. Reactions were carried out in a final volume of 20 µl of 1× PCR buffer (Invitrogen) containing template DNA, 0.2-0.5 µM of each primer, 0.2 mM dNTPs, 1.5 mM MgCl₂ and 2 units of Taq polymerase (Invitrogen) on a GeneAmp 2700 thermal cycler (Applied Biosystems). Cycling conditions were as follows: 94°C 2 min, then 32 cycles (94°C, 30 s; 45°C-65°C, 30 s; 72°C, 60 s). RACE PCR was performed using the Smart RACE kit (Clonetech), as described in the manufacturer’s instructions.

**Cloning and DNA sequencing:** PCR products were purified using the QIAquick kit (Qiagen) and ligated into the pGEM-T easy vector (Promega). Electrocompetent *E. coli* (DH10B) cells were transformed and recombinant clones were selected for a PCR-based screen using standard vector specific primers (T7/Sp6). Resulting products were digested with 1 unit of either Hae III or Rsa I (Promega) and separated by gel electrophoresis. Plasmid DNA from clones with unique restriction digestion patterns was isolated and sequenced commercially (Macrogen, South Korea). Sequences of the
N. alata DD genes have been deposited in GenBank with the accession numbers EF420251-EF4202510. Protein domain analysis was done using the SMART search tool (LETUNIC et al. 2006) (http://smart.embl-heidelberg.de/). All evolutionary analyses were performed using the Phylip package of programs (FELSENSTEIN 2004). DNA alignments were generated using TRANALIGN (http://bioweb.pasteur.fr/docs/EMBOSS/tranalign), based on ClustalW (v1.8) (HIGGINS and SHARP 1988) protein alignments of the corresponding sequences. MP (maximum parsimony) trees were estimated from the DNA alignment using DNAPARS with default settings. Neighbor-Joining (SAITOU and NEI 1987) distance trees were generated using NEIGHBOR, with a DNA distance matrix computed using DNADIST. Bootstrap analyses (FELSENSTEIN 1985) were carried out using 1000 bootstrap pseudo-replicates.

DNA blot analysis: Genomic DNA was extracted from leaf material using the plant DNAeasy kit (Qiagen). Leaf DNA (10 µg) was digested to completion with Eco RI (Promega), fractionated on an agarose gel and transferred to Hybond N+ (Amersham) membrane using the alkaline blot procedure described by the manufacturer. DNA fragments were radiolabeled with $^{32}$P dCTP using the Primagene kit (Promega). Hybridisations were performed overnight at 65°C in 5× SSPE, 5× Denhardt’s solution, 0.5% SDS and 100 µg/ml denatured herring sperm. After hybridisation membranes were washed twice in 2× SSC, 0.1% SDS and exposed to film.

RT-PCR analysis: Total RNA was prepared using Trizol reagent (Invitrogen). DNA contamination was removed by treating isolated RNA (1 µg) with 2 units of DNase I
First strand cDNA synthesis was carried out using an oligo dT$_{17}$ primer and 200 units of Superscript III (Invitrogen). Sequences of the gene-specific primers used in this experiment are shown in supplementary Table 1.

RESULTS

Cloning and sequence analysis of candidate S class SLF-like protein genes from N. alata pollen: Partial cDNAs for S class SLF-like genes were isolated from N. alata pollen by a degenerate PCR approach. Primers D2 and D4 were designed to bind to sequences conserved in Antirrhinum and Petunia SLFs. Pollen RNAs isolated from N. alata plants homozygous for the S$_1$, S$_2$, S$_3$, S$_6$ or S$_7$ allele were reverse transcribed and used as templates. PCR products were cloned and sequenced, and conceptual translations of all 6 frames were compared to GenBank using BlastP. After several rounds of screening, 10 partial cDNAs with high amino acid sequence similarity to Petunia SLF and SLF-like sequences were identified and given the provisional names DD1-DD10. Sequence differences between the 10 partial cDNAs were 30-40% and there was 5% or less difference in the sequence of the same DD cDNA amplified from different pollen RNA templates (data not shown). As the DD cDNAs represented candidates for N. alata SLF, Rapid Amplification of cDNA Ends (RACE) was used to obtain additional 5' and 3' sequence information. Full-length sequences for DD1 and DD5-8 were obtained by RACE using S$_2$S$_2$ pollen cDNA as the template. Full-length sequences for DD2 (S$_6$S$_6$), DD3/DD4 (S$_1$S$_1$) and DD9 (S$_3$S$_3$) were obtained from the
indicated pollen cDNAs, and a near full-length sequence for DD10 (missing the initiator Met codon and 5’ UTR) was obtained from S6S6 pollen cDNA.

Figure 1 shows conceptual translations of the 10 DD cDNAs aligned with representative SLF and SLF-like proteins from P. inflata (Solanaceae), A. hispanicum (Plantaginaceae) and Prunus mume (Rosaceae). DD5 and DD8 were the most closely related of the N. alata DD proteins with 88% amino acid similarity (78% identity), and DD3 and DD7 the least related with 50% amino acid similarity (35% identity). The protein similarity between the DD sequences and P. inflata SLF-S3 ranged from 53% (DD7) to 80% (DD3). SMART and PFAM analysis identified an N-terminal F-box domain in all N. alata DD proteins, consistent with them being members of the F-box protein superfamily. Other F-box proteins have diverse regions outside the F-box domain that bind to specific substrate proteins (GAGNE et al. 2002). Potential substrate binding motifs present in the DD proteins include the related F-box associated domains type 1 and type 3 (FBA_1, PF07734; FBA_3, PF08268) (Figure 1). WANG et al. (2004b) incorporated these domains into the four conserved C-terminal motifs (C1-C4) used to define the class S SLF-like proteins. C1-C4 are present in the SLF/SFB proteins and in all 10 DD proteins (Figure 1).

Nucleotide sequences of the N. alata DD genes were aligned with sequences of the SLF and class S SLF-like genes from P. inflata, P. axillaris and P. hybrida (Solanaceae) and A. hispanicum (Plantaginaceae). Sequences of SFBs from species in the Rosaceae were also included. Figure 2 shows the MP tree produced from the DNA alignment. Bootstrap analysis provides strong support for almost all nodes and distance trees
generated using the same alignment had an identical topology (results not shown). The Solanaceae SLF and SLF-like sequences are in a single cluster, as are the Antirrhinum sequences, with nodes at the base of the Solanaceae and Plantaginaceae clusters (nodes 1 and 2) having strong bootstrap support (100%). DD7 and DD6 are at the base of the Solanaceae cluster and sister to three well-supported smaller clusters (nodes 3-5). The P. inflata SLF sequences, PiSLF-S1, PiSLF-S2 and PiSLF-S3, group together in a cluster (node 6) with PaF1, an SLF-like sequence from P. axillaris (TSUKAMOTO et al. 2005). These sequences are sister to the Nicotiana sequences DD1, DD3 and DD4 (node 5). The second small cluster contains P. hybrida and P. inflata SLF-like sequences and DD2 (node 4), and the third cluster comprises only DD sequences (node 3).

**Expression of the DD genes in various N. alata S-genotypes and tissues:** To determine whether expression of a DD gene was restricted to pollen of a particular S genotype, PCR experiments were done with gene-specific primers and S 1S1, S2S2, S1S3, S6S6 and S7S7 pollen cDNA templates (Figure 3). In each case, DNA sequence analysis and / or restriction enzyme digests were used to confirm that the correct products had been amplified (data not shown); and pollen RNA (-RT control) was used to show that the products were derived from cDNA and not genomic DNA. Transcripts for two of the 10 DD genes were amplified from all five pollen templates tested (DD4 and DD6) and transcripts for another six DD genes (DD1-DD3, DD7, DD8 and DD10) were amplified from between 2 and 4 templates. For instance, DD1 transcripts were present in all pollen cDNAs except S1S1, and DD2 transcripts were detectable in S1S1, S6S6 and S7S7 cDNAs, but not S2S2 and S1S3 cDNAs. DD3 was detected in S1S1 and S1S3 pollen cDNAs but could not be amplified from S3S3 cDNA (data not shown). The remaining
two \textit{DD} genes were more restricted in their expression and transcripts were detectable in only one of the cDNAs; \textit{S}_2\textit{S}_2 \text{ template for } \textit{DD}5 \text{ and } \textit{S}_1\textit{S}_3 \text{ template for } \textit{DD}9.

The tissue-specific expression patterns of the \textit{DD} genes were investigated by RT-PCR using cDNA synthesized from \textit{N. alata} pollen, style, leaf and petal RNA. Figure 4 shows that \textit{DD}2, \textit{DD}3 and \textit{DD}5-9 were expressed only in pollen; that \textit{DD}1 and \textit{DD}4 were expressed in pollen and style; and that \textit{DD}10 was expressed in pollen and petal. The \textit{DD} genes were not expressed at detectable levels in non-floral tissues such as leaf or root (Figure 4 and data not shown).

**Determining linkage between the \textit{DD} genes and the \textit{S} locus:** Evidence of linkage between the \textit{S} locus and individual \textit{DD} genes was obtained using two families of \textit{N. alata} plants in which 4 \textit{S} alleles were segregating. \textit{S} genotypes of progeny plants resulting from a \textit{S}_1\textit{S}_3 \times \textit{S}_6\textit{S}_7 \text{ cross or a } \textit{S}_2\textit{S}_3 \times \textit{S}_6\textit{S}_7 \text{ cross were determined using allele-specific PCR of the } \textit{S-RNase} \text{ gene (results not shown). Cleavable amplified polymorphic sequence (CAPS) markers were designed for 8 of the } \textit{N. alata} \textit{DD} \text{ genes to test for co-segregation with an individual } \textit{S-RNase} \text{ allele (e.g., } \textit{S}_3). \text{ For } \textit{DD}3-\textit{DD}10 \text{ analysis of 18-20 plants detected no recombination between the } \textit{DD} \text{ marker and } \textit{S}_3-\textit{RNase} \text{ (supplementary Table 2), indicating that these } \textit{DD} \text{ genes are less than 5 cM from the } \textit{S} \text{ locus.}

Mapping of the \textit{DD} genes was also done using a collection of PPM \textit{N. alata} plants that are homozygous for the \textit{S}_6 \text{ allele and carry chromosomal duplications of varying lengths that contain an } \textit{S}_3 \text{ allele (} \textit{dS}_3). \text{ The duplications are present in the genomes of these}
plants as small extra chromosomes (centric fragments), segments that have been added to other chromosomes (translocations) or segments that are linked to an $S_6$ allele via an unequal crossover (Figure 5B). This approach to mapping genes to the *N. alata* $S$ locus is quicker than the standard linkage-based approach (supplementary Table 2) and provides information about relative gene order, but is only suitable for genes with $S_3$- and $S_6$-associated polymorphisms. Suitable polymorphisms were detected for $DD1-3$, $DD6$, $DD7$, $DD9$ and $DD10$. DNA blot analysis indicated that the $DD$ genes were all single-copy sequences in the *N. alata* genome (Figure 5A and data not shown).

Figure 5 illustrates how the PPMs were used to determine map locations for the $DD$ genes. Four PPM lines were selected for this analysis (M1-1, M1-2, M1-5 and M1-7) with Figure 5B showing the previously described order of markers ($CP100$, $48A$, $S$-RNase and 167A) on $dS_3$ in each line. *CP100*, $48A$ and $S$-RNase all map without recombination to the $S$ locus and 167A is 0.9 cM away (Li *et al.* 2000). On blots of genomic DNA from $S_3S_3$, $S_6S_6$, M1-1, M1-2, M1-5 and M1-7 (Figure 5A), the $DD1$ probe detected a restriction fragment-length polymorphism (RFLP) between $S_3S_3$ ($DD1-3$) and $S_6S_6$ ($DD1-6$). As expected, $DD1-6$ was present in all M1 plants ($S_6S_6dS_3$) but $DD1-3$ was only present in M1-7, placing $DD1$ to the right of marker 167A (Figure 5C). Similarly the $DD2$ probe hybridized to $DD2-3$ and $DD2-6$ in $S_3S_3$ and $S_6S_6$, respectively (Figure 5A). $DD2-3$ was present in all M1 plants, placing this gene within the $S$ locus in a chromosomal segment delimited by the left-hand border of the translocation in M1-7 and the right-hand border of the unequal exchange in M1-5 (Figure 5C). The other genes known to lie in this segment are $48A$ and pollen $S$. Figure 5C shows $DD2$ grouped with $48A$ on the centromeric side of the S locus, although the placement of
these markers with respect to each other and pollen $S$ is arbitrary. Figure 5C summarizes the PPM mapping data. $DD7$ and $DD10$ are in the same chromosomal segment as $DD2$, $DD9$ lies between the $S$-RNase and $167A$, and $DD1$, $DD3$ and $DD6$ are further from the $S$ locus than $167A$.

**DISCUSSION**

This paper describes the identification and initial characterization of 10 genes expressed in *N. alata* pollen ($DD1$-$10$) that code for class S SLF-like proteins, a class that includes all the known SLFs of *Antirrhinum* and *Petunia*, the SFBs of *Prunus*, and other clearly related proteins in these and other species with no obvious role in SI (WANG et al. 2004b). The $DD$ genes were all present as single-copy sequences and the 10 $DD$ genes appears to represent the full complement of class S SLF-like genes expressed in *N. alata* pollen that can be recovered using the RT-PCR-based approach described here. Because none of the $DD$ genes could unambiguously be identified as encoding the *N. alata* ortholog of SLF (see below), we propose to continue using this provisional naming scheme until such time as functions are identified and individual $DD$ genes can be given more descriptive names.

A surprising aspect of this work was the linkage detected in small-scale experiments between 8 of the 10 $DD$ genes ($DD3$-$DD10$) and the $S$ locus (supplementary Table 2). This potentially indicated the existence of a single cluster of $DD$ genes. Clusters of *F-box protein* genes have been described in Arabidopsis by GAGNE et al. (2002), where
35.9% of the 694 F-box protein genes are in arrays of 2-7 members. Instead of a single cluster of DD genes, it could equally be true that there are many separate SLF-like genes, all subject to the suppressed recombination affecting the S locus and precluding the production of well-resolved genetic maps (Li et al. 2000; Wang et al. 2003). The relative order of 7 of the DD genes with respect to each other and surrounding S locus markers could, however, be determined using a small number of PPMs with duplicated $S_3$ allele segments of differing sizes (Golz et al. 2001). This analysis showed that 4 DD genes ($DD2$, $DD7$, $DD9$ and $DD10$) exist as single genes (or possibly a cluster of a few genes) on either side of the S-RNase gene and within the S locus. Three of the DD genes ($DD1$, $DD3$ and $DD6$) were further from the S-RNase gene than the marker $167A$ and accordingly must be at least 0.9 cM from the S locus. In total, all 10 DD genes were linked to the S locus, although a lack of suitable polymorphisms for $DD4$, $DD5$ and $DD8$ meant that the position of these genes relative to other S locus markers and each other could not be determined.

Even though all the class S SLF-like genes identified in this study were expressed in N. alata pollen, they clearly cannot all be involved in SI. Some, for instance, were expressed in tissues other than pollen: specifically $DD1$ and $DD3$, which were also expressed in styles, and $DD10$, which had some expression in petals. Since all SLFs characterized to date are only expressed in pollen, these genes can be excluded from consideration as potential SLF orthologs. The remaining 7 DD genes, however, appear to be expressed only in pollen and thus warrant further study. Although all the DD genes are at or near the N. alata S locus, fine-scale mapping indicates that only three, $DD2$, $DD7$ and $DD10$, are in the same region of the chromosome as pollen S. Since
expression in petal means that DD10 can be excluded, only DD2 and DD7 can still be considered possible SLF orthologs, with the status of DD4, DD5 and DD8 being uncertain.

Interestingly, of these 5 ‘candidate’ genes, DD5 transcripts were only amplified from S2S2 pollen cDNA, and DD2, DD7 and DD8 transcripts were only amplified from 2 of the 5 pollen cDNA templates tested (Figure 3 and data not shown). DD4 transcripts were amplified from all 5 pollen templates. While this could potentially reflect a level of S allele specificity, amplification of transcripts from some cDNA templates but not others was also seen with DD genes that are clearly not pollen S (e.g., DD9 was only amplified from S1S3 pollen cDNA).

Both the number of class S SLF-like genes at or near the S locus and their pollen expression are worthy of comment. No such clustering of class S SLF-like genes is apparent in Arabidopsis, where the 7 genes of this type are distributed across 3 of the 5 chromosomes (WANG et al., 2004b). Predominant or exclusive expression of 10 class S SLF-like genes in N. alata pollen also contrasts with the situation in Arabidopsis, where six of the seven genes are constitutively expressed and only one, AtSFL79, is strictly pollen expressed (WANG et al. 2004b). SLF-like genes have, however, been found at or near the S locus in the Rosaceae (e.g. see ENTANI et al. 2003; USHIJIMA et al. 2003; SASSA et al. 2007), the Plantaginaceae (ZHOU et al. 2003), and in the solanaceous species Petunia inflata (WANG et al. 2003). In each case, these S locus-linked F-box protein genes are predominantly or exclusively expressed in pollen, just like the DD genes. We speculate that the number of class S SLF-like genes at the S locus and their
pollen expression is not due to chance but is an evolutionary outcome arising from the state of permanent heterozygosity that SI imposes.

In a process analogous to that leading to a loss of active genes on the male-determining Y chromosome of animals (see SKALETSKY et al. 2003), permanent heterozygosity at the S locus will reduce the effectiveness of selection against deleterious mutations occurring in genes with unrelated functions that are embedded within it (UYENOYAMA 1997, 2005). As each individual S allele experiences a small effective population size, deleterious mutations in genes tightly linked to the S locus are highly likely to replace functional wild-type alleles, a process that should, over time, lead to the degeneration of most genes within the S locus region except those with SI-specific functions. Any gene expressed in a haploid cell like a pollen grain will escape this effect because there are no genes associated with a second S allele to provide \textit{in trans} the functions lost by mutation. Hence these genes will still be subject to selection against deleterious mutations, especially if their products contribute to the reproductive fitness of pollen. Selection may also favor expanding the number of these genes at the S locus through gene duplication in cases where higher expression levels enhance male fitness. Thus, this evolutionary scenario suggests that the cluster of \textit{DD} genes at the \textit{N. alata} S locus arises because enforced heterozygosity favors genes that contribute to male fitness and that through pollen expression avoid gene erosion processes.

The phylogenetic tree in Figure 2 highlights several puzzling aspects of the S class SLF-like proteins. Most notably, the SLF/SFB proteins of the Plantaginaceae, Solanaceae and Rosaceae do not form a monophyletic clade to the exclusion of other class S SLF-
like proteins with no role in SI, such as many of the DD proteins. Contrast can be made here to trees of S-RNases and the related S-like RNases, which are not involved in SI. Phylogenetic analysis places S-RNases and S-like RNases into one of three classes: a single monophyletic clade that contains all S-RNases from the Solanaceae, Plantaginaceae and Rosaceae, and two separate clades of S-like RNases (IGiC and KOHN 2001; STEINBACHS and HOLSINGER 2002). The apparent homology of the S-RNases leads to the conclusion that the three known RNase-based SI systems share a single evolutionary origin. In Figure 2, however, S class SLF-like proteins are grouped not according to function but on the basis of the taxonomic relationships of the organisms from which they were derived. Thus the conclusion that the S-RNases of the Solanaceae, Rosaceae and Plantaginaceae are homologous is not borne out by this analysis of the SLFs and SFBs. Assuming that the SLFs and SFBs are involved in SI, then the most parsimonious interpretation of Figure 2 is that their encoding genes have been independently recruited to roles in RNase-based SI in the Rosaceae, Plantaginaceae and Solanaceae. Lack of homology implies that different mechanisms may be used to achieve the rejection of incompatible pollen in each family.

A second notable aspect of the phylogenetic reconstruction in Figure 2 is the well-resolved clade arising at node 5 that contains all the *Petunia* SLFs, proteins such as PaF1 (*P. axillaris* F-box protein 1) that are presumed not to be at the *Petunia* S locus (TSUKAMOTO *et al*., 2005) and two DD proteins (DD1 and DD3) that are not considered candidates for the *N. alata* SLF ortholog because their encoding genes are at least 0.9 cM from the S locus (Figure 5). Numerous phylogenetic analyses of the solanaceous S-RNases place these sequences in a well-resolved monophyletic clade that excludes S-
like RNases (e.g., IGIC and KOHN, 2001; STEINBACHS and HOLSINGER, 2002). Any RNase sequence falling into the S-RNase clade not involved in self-incompatibility (e.g., one obtained from a self-compatible plant) is presumed to be derived from a functional S-RNase (relic S-RNases; GOLZ et al. 1998). For the clade in Figure 2 to be a clade of solanaceous SLFs resembling the clade of solanaceous S-RNases, DD1, DD3 and PaF1 must all be relic SLFs. Since these proteins are from self-incompatible species, it therefore becomes problematic to assign the name SLF to any protein sequence within this clade unless its function in SI has been verified experimentally. It is also notable that 4 of the remaining 5 SLF candidates, DD2, DD5, DD7 and DD8, are not in this SLF-containing clade.

The final puzzling feature of the tree is the shortness of terminal branches for SLFs in the Plantaginaceae and Solanaceae, reflecting the low levels of sequence polymorphism among SLF alleles in these two families. Notably, this feature is not seen with SFBs from the Rosaceae. The 4 Antirrhinum SLF alleles are identical at 97% or more of their amino acid positions and the two most divergent Petunia SLF alleles, PiSLF-S3 and PaSLF-S17, are 87% identical. These % identity values are closer to those of genes with functions unrelated to self-incompatibility that also lie within the S locus (such as 48A; TAKEBAYASHI et al. 2003) than they are to the S-RNases, where pair-wise amino acid identities of 40-50% are common (CLARK and KAO 1994). The pair-wise amino acid identity in partial cDNAs of DD1, DD2, DD4, DD6, DD7, DD8 and DD10 from different S allele backgrounds ranges from 1-5%.
SI imposes a very intense form of balancing selection on the S locus, as a consequence of which high numbers of S allele lineages are maintained over long periods of time (reviewed by Clark and Kao 1994; Richman 2000). The effect balancing selection has had on the phylogenetic history of the S-RNase gene, particularly on extending the phylogenies of allelic lineages over time scales that are vastly longer than those expected under neutrality, has been demonstrated many times (see Ioeger et al. 1990; Richman et al. 1996). In the Rosaceae identifying SFBs as the pollen S factor is in part based on agreement between the expected phylogeny of a gene under balancing selection and the observed SFB phylogenies (Ikeda et al, 2004; Nunes et al, 2006; Sassa et al, 2007). Conversely, genes have also been excluded from consideration as pollen S if their phylogeny did not show the expected features (e.g., Takebayashi et al, 2003).

Little evidence of this predicted evolutionary history is seen in the Petunia and Antirrhinum SLFs, however: the clades containing these sequences are marked by short terminal branches that quickly converged on the common ancestor of all sampled lineages. Because in the Rosaceae SFBs, sites identified as being positively selected largely overlap with regions of highest sequence variability (Ikeda et al, 2004; Nunes et al, 2006), we visually inspected an alignment of SLF proteins from the Solanaceae for highly variable sites (i.e., sites occupied by many different amino acids; see supplementary Figure 7). As only one such site was found it is probable that the sequences also lack positively selected sites likely to be responsible for defining pollen specificities.
The marked lack of concordance between S-RNase trees, which are largely consistent with well-established theories as to how genes evolve under balancing selection (Clark and Kao, 1994), and the class S SLF-like tree for the Solanaceae leads to the unsettling conclusion that the SLFs and S-RNases are not stably co-evolving partners, something that previously had been assumed would be true of the pollen and stylar factors of the S locus. Two possible explanations of this situation exist: either there is no obligate relationship between the S-RNase and a particular class S F-box protein; or the SLFs do not fulfill the role of pollen S in the Solanaceae and Plantaginaceae.

Although phylogenetic reconstruction of the class S SLF-like proteins provides little indication that the SLFs have a long history of evolving under balancing selection, other evidence is consistent with the expected behavior of pollen S. Apart from their genetic location and expression pattern, this evidence includes the binding of recombinant SLF proteins to S-RNases and the pollination behavior of transgenic plants that express SLFs in their pollen (e.g., see Sijacic et al., 2004; Qiao et al., 2004b). It should however be noted that so far it has not been possible to demonstrate allele specificity for the SLFs (McClure 2004). Recombinant SLF proteins bind S-RNases non-specifically, and SLF expression in transgenic plants results in a loss of SI rather than in the gain of a new S allele specificity. Since other proteins can bind S-RNases (Sims and Ordanic 2001), and since mutations in other genes can lead to self-compatible phenotypes (e.g., At et al. 1991), it remains formal possible that the properties currently used to define pollen S in the Solanaceae and Plantaginaceae are not exclusive to this factor.
The alternative explanation is that the functions of pollen S are transient and able to be performed by any one of the many paralagous class S SLF-like genes within the S locus. SASSA et al. (2007) recently made a similar suggestion, based on their observation that the S locus in Malus and Pyrus (Rosaceae) contained two or more copies of the SFB gene (‘SFB brothers’ or SFBBs). If paralagous SLF genes act as proposed, then pollen S lineages might be expected to turn over far more rapidly than S-RNase lineages. Although this suggestion currently lacks experimental support, it is obvious that identifying which if any of the DD genes codes for pollen S in N. alata will help refine ideas as to why the SLF and S-RNase genes have such remarkably divergent evolutionary histories.

ACKNOWLEDGEMENTS

We thank Bruce M’Ginness for assistance in the glass house and Poh Ling Koh for experimental help. This work is supported by a grant from the Australian Research Council.

LITRITURE CITED


FELSENSTEIN, J., 2004 PHYLIP (Phylogeny Inference Package) version 3.6, pp. Distributed by the author.


LAI, Z., W. MA, B. HAN, L. LIANG, Y. ZHANG et al., 2002 An F-box gene linked to the self-incompatibility (S) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. Plant Mol. Biol. **50**: 29-42.


Qiao, H., H. Wang, L. Zhao, J. Zhou, J. Huang et al., 2004b The F-box protein AhSLF-S2 physically interacts with S-RNases that may be inhibited by the ubiquitin/26S proteasome pathway of protein degradation during compatible pollination in Antirrhinum. Plant Cell 16: 582-595.


SIMS, T. L., and M. ORDANIC, 2001 Identification of a S-ribonuclease-binding protein in


Figure 1: Amino acid sequences encoded by the 10 *N. alata* DD cDNAs aligned with representative SLF-like sequences from *A. hispanicum*, *P. inflata*, and *P. mume*. Primer binding sites for degenerate primers D2 and D4 are shown by the arrows. The F-box and four conserved motifs identified in Wang *et al.* (2004b) are indicated by solid lines. The C-terminal F-box associated domain (FBA 1 and FBA 3) is indicated by the dashed line underneath the alignment. Accession numbers of sequences from species other than *N. alata*: AhSLFS1 (AJ515535), PiSLFS1 (AY500390), PiSLFS2 (AY136628), PiSLFS3 (AY500392), PiS3A134 (AY363975), PmSFBS1 (BAD08320).

Figure 2: Maximum parsimony tree (5087 steps) derived from aligned SLF/SFB and SLF-like DNA sequences. The 3 *Petunia* sequences identified with pollen *S* are indicated in bold type. Bootstrap values (1,000 pseudoreplicates) for numbered nodes are: 1=100%, 2=100%, 3=96%, 4=100%, 5=99%, 6=100%. Additional accession numbers of sequences shown in the tree: PaSLFS17 (AY766153), PaSLFS19 (AY766154), PhSLFS3A (AY639403), PhSLFS3B (AY639402), PaF1 (AY766155), PiS3A113 (AY363972), PiS1A134 (AY363973), PiS2A134 (AY363974), PiS2A113 (AY363971), AhSLFS5 (AJ515536), AhSLF-S1E (AJ515535), AhSLF-S4D (AJ515534), AhSLFS4A (AJ515534), PmSFBS7 (AB092622), PdSFBa (AB092966), MdSFBB9b (AB270794), MdSFBB9a (AB270793), PpFBB4a (AB270797), PpSFBB4b (AB270798), PavSFB1 (PavSFB1), PbSFBb (AB081648).

Figure 3: PCR carried out using primers specific for each DD gene shown on the right, with pollen cDNA templates of the indicated S genotype shown at the top of the gel.
No reverse transcriptase controls are indicated by the minus sign and the size of each PCR product in base pairs (bp) is shown to the left of each gel.

Figure 4: Spatial expression pattern of DD genes. RT-PCR performed on N. alata tissue indicated at the top of each gel. Primers used were specific for the gene shown to the right, and the cDNA genotype tested is indicated in parentheses. No reverse transcriptase controls are indicated by the minus sign, and actin-specific primers were used to confirm the presence of template. PCR product sizes in bp are indicated to the left.

Figure 5: Linkage of the DD genes to the S locus. A. Representative southern blots using probes specific for the indicated DD gene. Source of genomic DNA is shown at the top of each lane and the different alleles are indicated to the right. B. Maps of dS$_3$ in each of the indicated PPMs. A black dot indicates the centromere. PPMs used in this experiment all have the genotype $S_6S_6dS_3$. C. Map of the N. alata S locus showing relative placement of the 7 DD genes mapped using this approach. Placement of genes in the segment between CP100 and S-RNase with respect to each other is arbitrary.

Supplementary Table 1: Sequences of oligonucleotide primers used in this study.

Supplementary Table 2: Summary of pooled data showing linkage between the indicated DD gene and the S$_3$ allele in families arising from $S_1S_3 \times S_6S_7$ or $S_2S_3 \times S_6S_7$ crosses.
Supplementary Figure 6: *N. alata* genomic DNA isolated from S₃S₃, S₆S₆ and four PPMs, M1-1, M1-2, M1-5, M1-7, was used as template in the indicated PCRs. Each PCR used primers that specifically amplified a region of the gene of interest, with the size of the resulting products shown in base pairs to the left of the figure. For *DD6*, digestion with Hae III was used to distinguish the S₃S₃ product (*DD6*-3) and the S₆S₆ product (*DD6*-6). Digestion of the *DD6*-3 product with Hae III produces a diagnostic fragment of 320 bp. These results demonstrate that *DD6* is on the distal side of the marker *167A* and that DD9 is between *167A* and the *S-RNase* gene (see Figure 5C), locations that are inconsistent with either gene encoding pollen S.

Supplementary Figure 7: Alignment of deduced amino acid sequences of three *SLF* alleles from *Petunia inflata* (SIJACIC *et al.* 2004). Amino acid differences between any two sequences are between 10.3% and 11.6%, with 44 of the 389 positions being variable. Replacements at 12 of the 44 sites are conservative, however, and at only one site (position 75) are there three alternative amino acids in the alignment. At a second site (position 261) there are two different amino acids and a gap. For comparison, IKEDA *et al.* (2004) identified 34 of the 384 sites in the *Prunus* SFBs as being highly variable.