PERSPECTIVE

RNase-Based Self-Incompatibility: Puzzled by Pollen S

Ed Newbigin,a Timothy Paape,b and Joshua R. Kohnb

aSchool of Botany, University of Melbourne, VIC 3010, Australia
bSection of Ecology, Behavior, and Evolution, Division of Biological Sciences, University of California at San Diego, La Jolla, California 92093

Many plants have a genetically determined self-incompatibility system in which the rejection of self pollen grains is controlled by alleles of an S locus. A common feature of these S loci is that separate pollen- and style-expressed genes (pollen S and style S, respectively) determine S allele identity. The long-held view has been that pollen S and style S must be a coevolving gene pair in order for allelic recognition to be maintained as new S alleles arise. In at least three plant families, the Solanaceae, Rosaceae, and Plantaginaceae, the style S gene has long been known to encode an extracellular ribonuclease called the S-RNase. Pollen S in these families has more recently been identified and encodes an F-box protein known as either SLF or SFB. In this perspective, we describe the puzzling evolutionary relationship that exists between the SLF/SFB and S-RNase genes and show that in most cases cognate pairs of genes are not coevolving in the expected manner. Because some pollen S genes appear to have arisen much more recently than their style S cognates, we conclude that either some pollen S genes have been falsely identified or that there is a major problem with our understanding of how the S locus evolves.

In three plant families, the Solanaceae, Rosaceae, and Plantaginaceae (formerly Scrophulariaceae; Albach et al. 2005), self-incompatibility (SI) is controlled by extracellular ribonucleases (S-RNases) in the style that are the products of an S locus characterized by large numbers of alleles (Takayama and Isogai, 2005). For many years the pollen counterpart of the S-RNases (pollen S) was unknown. However, a series of recent articles identified candidate genes in all three families (Lai et al., 2002; Entani et al., 2003; Ushijima et al., 2003; Ikeda et al., 2004; Sijacic et al., 2004) and provided evidence for involvement of these genes in the SI response (Ikeda et al., 2004; Sijacic et al., 2004; Qiao et al., 2004a, 2004b; Ushijima et al., 2004; Sonneveld et al., 2005). The studies implicated genes encoding an F-box protein, involved in selective targets for ubiquitination, as being pollen S in all three families. In the Solanaceae and Plantaginaceae, this F-box protein is called SLF and in the Rosaceae it is called SFB by some groups (Entani et al., 2003) and SFB by others (Ushijima et al., 2003; Ikeda et al., 2004). For simplicity, however, we will refer to the pollen S proteins from the Solanaceae and Plantaginaceae as SLF and from the Rosaceae as SFB. Commentaries heralding this exciting development appeared in many scientific journals, including this one (McClure, 2004), and it appeared that we were closer than ever to understanding RNase-based SI.

However, many more questions than answers have been raised by our as yet limited understanding of pollen S. Rather than leading to elucidation of a single conserved SI mechanism among families that use S-RNases, subsequent studies have identified mechanistic differences between the Rosaceae and Solanaceae in the way incompatible pollen is rejected (Sonneveld et al., 2005; Hauck et al., 2006). Uncertainty surrounds the mechanism of SLF action in the Solanaceae (McClure, 2006), and even the components that comprise the ubiquitination complex of which SLF is a part appear different between the Solanaceae and Plantaginaceae (Hua and Kao, 2006; Huang et al., 2006). In this perspective, we examine another aspect of the SLF/SFB story: the puzzling evolutionary relationships between SLF/SFB alleles and their putative cognate S-RNases. In at least some cases, the putative pollen S genes are evolving far differently than expected for this self-recognition locus. Either some pollen S genes have been falsely identified, or there is a major problem with our understanding of how pollen S evolves.

POPULATION GENETICS OF THE S LOCUS

To understand why currently identified pollen S genes are puzzling requires an appreciation of the evolutionary processes that act on the S locus and explain its unusually high levels of allelic polymorphism. In the SI systems described here, haploid pollen grains are rejected when they express an S allele that matches either of those expressed in the diploid style. This simple genetic system prevents self-fertilization and limits cross-fertilizations among related plants, avoiding the deleterious effects of close inbreeding. As Wright (1939) pointed out, a rare S allele will experience strong selection to increase in frequency.
because it has increased mating opportunities relative to common S alleles. This type of selection, called frequency-dependent selection, is responsible for the maintenance of dozens of S alleles in populations of SI plants (reviewed in Lawrence, 2000).

A further consequence of frequency-dependent selection is that S alleles persist in populations for periods of time much longer than those of alleles of more standard genes (Charlesworth and Guttmann, 1997). Allelic polymorphism persists because any allele that becomes rare in a population through random genetic drift is likely to subsequently increase in frequency due to selection, rather than to decline further in frequency toward extinction. Frequency-dependent selection acting on the S locus is expected to preserve polymorphism over tens of millions of years, far longer than the average lifetime of a species (Takahata, 1990; Vekemans and Slatkin, 1994).

The predicted effects of frequency-dependent selection on the S locus have largely been confirmed from analysis of S-RNase sequences. For instance, in addition to their large numbers, S-RNase alleles from the Solanaceae are highly polymorphic: amino acid identity of alleles drawn from the same obligately heterozygous individual is often <50% (loeger et al., 1990; Richman et al., 1995). Much of this diversity is found in hypervariable regions. Various tests of the magnitude and nature of the changes in these regions have shown it to be due to selection favoring amino acid replacement, as one would expect if selection favors the creation of new specificities (loeger et al., 1991; Ishimizu et al., 1998; Takebayashi et al., 2003; Igic et al., 2007; Vieira et al., 2007). Moreover, there is direct transgenic evidence in the Solanaceae showing that changing amino acids within hypervariable regions alters the allelic identity of an S-RNase (Matton et al., 1999). However, there is abundant polymorphism throughout the S-RNase gene, save for a relatively few sites that are conserved to protect enzymatic function or proper protein conformation. So while some of the extreme sequence polymorphism is due to selection for different specificities, much of it has accrued due to the great age of allelic lineages. Another piece of evidence for the great age of S-RNase alleles is shared ancestral polymorphism. An allele from one species is often more closely related to an allele from another species or genus than to other alleles within its own species (loeger et al., 1990). This is evidence that the common ancestor of these species already possessed much of the extant allelic diversity and passed it down to its descendents. For instance, most of the S-RNase allele polymorphism in the Solanaceae was present in a common ancestor that is estimated to have existed 30 to 40 million years ago (Paape et al., 2008). Furthermore, genealogies also point to a single evolutionary origin for the S-RNases in the Rosaceae, Solanaceae, and Plantaginaceae, suggesting that RNase-based SI was the ancestral state in the majority of eudicots (Igic and Kohn, 2001; Steinbachs and Holsinger, 2002).

In all SI systems that have been characterized at the molecular level, the determinants of pollen and stylar specificity are encoded by a pair of genes at the S locus, with one specifying stylar and the other pollen specificity. To maintain the functional relationship between any pair of pollen and stylar genes, recombination is suppressed around the S locus. Accordingly, the creation of a new S allele requires coordinate changes to both genes. How this happens is unknown, although some plausible evolutionary pathways have been described (Newbigin and Uyenoyama, 2005). Whatever the route, the strong expectation is that the pollen and stylar genes should coevolve. That is, SFL or SFB alleles that control pollen recognition are expected to show the same signature of long-term frequency-dependent selection seen in the S-RNases: abundant sequence polymorphism, evidence for positive selection, long coalescence times of extant alleles, and shared ancestral polymorphism. In addition, the genealogies of cognate pollen and stylar alleles are expected to be largely concordant, reflecting a long coevolutionary history. In the Brassicaceae, which possess an SI system not involving S-RNases, these theoretical expectations of pollen and stylar loci have largely been confirmed: both pollen and style genes at the S locus show extreme polymorphism, long coalescence times, and evidence of positive selection, and there is good concordance in the genealogies of pairs of pollen and style alleles (Sato et al., 2002; Takebayashi et al., 2003).

**WHAT’S PUZZLING ABOUT THE SLF AND SFBs?**

Perhaps the clearest example of where the SLFs have failed to meet these evolutionary expectations is from Antirrhinum hispanicum (Plantaginaceae), a wild relative of the snapdragon. This example is significant because the first SLF allele to be described came from this family. This allele is Ah SLF-S2, the SLF allele associated with the S2 allele (Lai et al., 2002). Ah SLF-S2 was identified based on its proximity to the A. hispanicum S2-RNase gene and pollen-dominated expression pattern. Later it was shown that Ah SLF-S2 and S-RNase interact in yeast two-hybrid and communoprecipitation assays and that expressing Ah SLF-S2 in transgenic plants produces a pollination phenotype that is consistent with some models of how pollen S functions (Qiao et al., 2004a, 2004b). However, there were two major surprises when the sequences of the Ah SLF alleles associated with the S1, S4, and S5 alleles became available (Zhou et al., 2003). The first was the extremely high (97 to 99%) pairwise amino acid identity among the four SLF alleles (Figure 1A). The second was that the few amino acid differences that did exist were scattered throughout the sequence, not clustered into a hypervariable region as they are in the S-RNases. By comparison, the S-RNases associated with these four Antirrhinum S alleles appear just as expected, with pairwise amino acid identities ranging from 36 to 51%. Levels of nonsynonymous (amino acid–changing, Ka) and synonymous (silent, Ks) substitutions in the DNA sequence are both more than an order of magnitude lower at the Antirrhinum SLF locus than at the cognate S-RNase locus (Table 1: Xue et al., 1996; Zhou et al., 2003; Sassa et al., 2007; Wheeler and Newbigin, 2007). The high level of identity among the Antirrhinum
SLF alleles is indicative of polymorphisms with a much shorter evolutionary history than that of the S-RNase (Figure 1A). Equating the Antirrhinum SLF gene with pollen S means that in Antirrhinum, pollen S must evolve in a radically different way than the S-RNase it interacts with.

By contrast, the Rosaceae SFBs provide some of the best evolutionary and other evidence of a gene that is coevolving with an S-RNase gene and that regulates pollen SI phenotype. Many SFB alleles have been cloned from species in the genus Prunus (apricots, peaches, and almonds), which belongs to the Amygdaloideae subfamily of the Rosaceae, and sequence variability among them is quite similar to that of the corresponding S-RNases (Table 1, Figure 1B). In addition, statistical tests find strong evidence that sequence polymorphism among SFB alleles is due to selection (Ikeda et al., 2004; Nunes et al., 2006). While the amount of divergence among pollen and style alleles is similar, there is unfortunately little phylogenetic structure to either the Prunus S-RNase or SFB alleles. Therefore, no strong inference can be made as to whether style and pollen genes share the same evolutionary history (Nunes et al., 2006). Finally, loss-of-function mutations in Prunus SFB are associated with pollen-part self-fertility (Ushijima et al., 2004; Sonneveld et al., 2005, Hauck et al., 2006).

Intriguingly, Prunus also provides examples of S-RNases from different species that are identical or nearly identical in sequence but are associated with SFBs that differ by as many as 12 amino acid replacements (Šurbanovski et al., 2007).

The generally encouraging story to emerge from analyses of SI in Prunus becomes a bit more curious when recent evidence from another subfamily of Rosaceae, the Maloideae (apples and pears, among others), is considered. Whereas each Prunus S allele has just one linked SFB gene, in the Maloideae, each allele has two to three related SFB genes (Sassa et al., 2007). Called SFBB, for S locus F-box brothers, levels of polymorphism across

<table>
<thead>
<tr>
<th>S-RNase</th>
<th>SLF/SFB/SFBB</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. hispanicum</td>
<td>1.118 0.506 0.034 0.0092</td>
</tr>
<tr>
<td>P. inflata</td>
<td>0.386 0.152 0.159 0.0502</td>
</tr>
<tr>
<td>Maloideae</td>
<td>0.251 0.218 0.233 0.1560</td>
</tr>
<tr>
<td>P. avium</td>
<td>0.239 0.138 0.247 0.1138</td>
</tr>
</tbody>
</table>

Loci compared are from A. hispanicum (Plantaginaceae), P. inflata (Solanaceae), subfamily Maloideae (Rosaceae; two haplotypes from Malus domestica and two from Pyrus pyrifolia), and P. avium (Rosaceae, subfamily Amygdaloideae). For the Maloideae, values are averaged across SFBB sequences (see text). Ka (replacement) and Ks (silent) values were estimated using DNASP v4.20 (Rozas et al., 2003). Supplemental Table 1 online contains accession numbers for all the sequences used.
Figure 2. Maximum Likelihood Tree of SLF and SLF-Like Genes from the Plantaginaceae and Solanaceae.

Plantaginaceae sequences are from *A. hispanicum* (Ahis) and *A. majus* (Amaj), and Solanaceae sequences are from *Nicotiana alata* (Nala), *Petunia axillaris* (Paxi), *P. hybida* (Phyb), and *P. inflata* (Pinf). Genes considered to encode pollen S are indicated in boldface. The phylogeny was constructed using PAUP (Swofford, 2000), and ModelTest v3.0 (Posada and Crandall, 1998) was used to select the most likely nucleotide substitution model (GTR + G + I). The results from MrBayes 3.1 (http://mrbayes.csit.fsu.edu/manual.php/) produced a topology consistent with the maximum likelihood tree (GTR + G + I nucleotide substitution model; the analysis was run for 1,000,000 generations, sampling every 500th generation). Numbers on branches represent bootstrap support (PAUP; Swofford, 2000) and Bayesian posterior probabilities (calculated using MrBayes3; Ronquist and Huelsenbeck, 2003), respectively. Both methods yielded trees with similar topologies. Supplemental Table 1 online contains accession numbers for all the sequences used.
these genes are similar to each other and only somewhat lower than those of their cognate S-RNases. It is thus unclear which SFBB locus is involved in pollen specificity, and Sassa et al. (2007) suggest that they may act in concert to determine mating type. Testing this hypothesis will be difficult, however, because all of the Maloideae that have been examined are trees. Transgenic approaches aimed at testing the effects of loss-of-function and gain-of-function mutations on pollination phenotype, while feasible (Broothaerts et al., 2004), may take some time.

Fitting in between the Rosaceae SFBs and the Plantaginaceae SLFs in terms of nucleotide variation are the Solanaceae SLFs, which are all from the genus Petunia (Figure 1A; Sijacic et al., 2004). Despite the fact that S-RNases of the Solanaceae generally have higher levels of nucleotide polymorphism, implying a greater age, than those of Prunus or the Maloideae, levels of nucleotide polymorphism among the Petunia SLFs are lower than those of the SFBs (Table 1). As with the Antirrhinum SLFs, variable amino acid positions are not clustered into distinct hypervariable regions but distributed over the sequence. Similar types of binding and transgenic experiments to those performed with Ah SLF-S2 have also been done with the Petunia SLF alleles and have produced results that are consistent with expectations for pollen S (Sijacic et al., 2004; Hua and Kao, 2006; Hua et al., 2007).

Attempts at identifying SLF orthologs in solanaceous species other than Petunia have highlighted one further evolutionary difference between this gene and the S-RNase gene (Wheeler and Newbigin, 2007). Phylogenetic clustering of S-RNases relative to nearly all other plant RNases points to a single common origin for S-RNase-based SI (Igic and Kohn, 2001; Steinbachs and Holsinger, 2002). In particular, aside from a few genes that appear to be derived from the S locus itself (relic S-RNases; Golz et al., 1998), S-RNases from the Plantaginaceae and Solanaceae are each other’s closest known relatives, as are S-RNases from the two subfamilies of Rosaceae from which S-RNases have been cloned. By contrast, the SLF or SFB genes in each taxa are more closely related to other F-box protein genes in the same taxon, usually also found linked to the S locus, than to the SLF and SFB genes of other taxa (Figure 2; Entani et al., 2003; Ushijima et al., 2003; Wheeler and Newbigin, 2007). This introduces a further complication into the SLF/SFB story: embedded in the S locus are many pollen-expressed F-box protein genes that are related to the SLFs and SFBs. These genes are called SLF-likes. Although some SLF-likes are closely related to the SLFs (Figure 2), they are generally thought not to function in SI because alleles from different S genotypes show little or no sequence variation, and when expressed in transgenic plants these genes do not alter the pollination phenotype (Entani et al., 2003; Ushijima et al., 2003; Wang et al., 2003; Zhou et al., 2003; Hua et al., 2007; Sassa et al., 2007; Wheeler and Newbigin, 2007). The close intrafamily relationships between the SLF/SFB and SLF-like genes again implies an origin for the SLF/SFB alleles that is much younger than that of the corresponding S-RNase alleles.

CONCLUSIONS

If the molecular data supporting the role of SLF/SFB genes as encoding pollen S are accepted, then it is obvious that we must abandon, at least in some taxa, our expectations for how frequency-dependent selection will affect sequence polymorphism and coevolution with the S-RNase locus. Given that coevolution of pollen and stylar genes occupies a central place in theories of the S locus and that conclusions based wholly or in part on this expectation are to be found in much of the literature on RNase-based SI systems, abandoning this theory is not going to be easy.

An obvious first consequence of uncoupling the histories of pollen and style genes is that our methods for finding pollen S in additional taxa will need to be revised. Most attempts at identifying pollen S have been based on analyzing the sequences of pollen-expressed genes at the S locus. Those genes with high levels of polymorphism are considered pollen S candidates and less polymorphic genes are rejected. For instance, Takebayashi et al. (2003) rejected the S locus–linked, pollen-expressed gene 48a as a candidate for pollen S based on its low level of polymorphism and lack of evidence of positive selection. However, as the Antirrhinum example shows, abundant sequence polymorphism can no longer be considered a reliable diagnostic feature of pollen S. Compounding the problem is the fact that most taxa have many pollen-expressed F-box protein loci at or near the S locus (Sassa et al., 2007; Wheeler and Newbigin, 2007), all of which must be considered candidates until functional tests demonstrate otherwise.

There are several possible ways to remove the requirement for coevolution of pollen and stylar determinants and allow pollen S lineages to be much younger than S-RNase lineages. Potentially, it is the existence of paralogous genes, as is the case for the SLF and SFB loci but not the S-RNase locus, that is responsible for eradicating evidence of frequency-dependent selection. For instance, it is possible that gene conversion between SLF alleles or among related F-box protein loci homogenizes these sequences except for whatever variation is needed to specify allelic identity. Alternatively, any of the F-box protein paralogs at the S locus (SLFs, SFBs, and SLF-likes) may also be able to encode pollen S, so that over time new genes are recruited to this function. Although the selective constraints that permit these and other conceivable scenarios to take place are not readily apparent, it should at least be possible to distinguish them by the predictions they make about how sequence variation is distributed over SLF and SFB alleles and their SLF-like paralogs.

**Supplemental Data**

The following material is available in the online version of this article.

**Supplemental Table 1.** Gene Names and Accession Numbers for the F-Box Proteins (SLF, SLF-Like, SFB, and SFBB) Used in the Analyses Shown in Table 1 and Figures 1 and 2.
ACKNOWLEDGMENTS

This work is supported by grants from the Australian Research Council to E.N. and the U.S. National Science Foundation (DEB 0639984) to J.R.K.

REFERENCES


**RNase-Based Self-Incompatibility: Puzzled by Pollen S**

Ed Newbigin, Timothy Paape and Joshua R. Kohn

*Plant Cell* 2008;20;2286-2292; originally published online September 5, 2008;
DOI 10.1105/tpc.108.060327

This information is current as of July 1, 2017

<table>
<thead>
<tr>
<th>Supplemental Data</th>
<th>/content/suppl/2008/08/19/tpc.108.060327.DC1.html</th>
</tr>
</thead>
<tbody>
<tr>
<td>References</td>
<td>This article cites 47 articles, 22 of which can be accessed free at: /content/20/9/2286.full.html#ref-list-1</td>
</tr>
<tr>
<td>eTOCs</td>
<td>Sign up for eTOCs at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a></td>
</tr>
<tr>
<td>CiteTrack Alerts</td>
<td>Sign up for CiteTrack Alerts at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a></td>
</tr>
<tr>
<td>Subscription Information</td>
<td>Subscription Information for <em>The Plant Cell</em> and <em>Plant Physiology</em> is available at: <a href="http://www.aspb.org/publications/subscriptions.cfm">http://www.aspb.org/publications/subscriptions.cfm</a></td>
</tr>
</tbody>
</table>

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY