Supplementary note 11

Characterization of SI Loci in *P. inflata* and *P. axillaris* identifies multiple linked S-locus F-box genes.

Diwa Malla[a], Qinzhou Qi[a] Timothy P. Robbins[b] and Thomas L. Sims[a1]

[a]Department of Biological Sciences, Northern Illinois University, DeKalb IL 60115
[b]Department of Plant and Crop Sciences, University of Nottingham, Sutton Bonington, Leicestershire, UK LE12 5RD

[1]Address correspondence to tsims@niu.edu

Running title: S-locus sequences in *Petunia*

ABSTRACT

Petunia species exhibit S-RNase-based gametophytic self-incompatibility (GSI), in which plants are capable of recognizing and rejecting their own (self) pollen, while accepting pollen from a different (non-self) individual. Gametophytic self-incompatibility is found in both progenitor species (*Petunia inflata* and *Petunia axillaris* as well as in the cultivated hybrid *Petunia hybrida*). Genes critical for self versus non-self recognition are the S-locus ribonuclease (S-RNase), which is the style-recognition component of GSI, and S-locus F-box (SLF) genes, which encode the pollen-recognition component of GSI. Although any individual plant possesses at most two S-RNase alleles, multiple SLF genes appear to be involved in recognizing an overlapping range of non-self S-RNase proteins. Both S-RNase and SLF genes are thought to be tightly linked at the S-locus as pollen-recognition and style-recognition specificities do not recombine. Here, we have used BLASTn and BLASTp searches of the assembled sequences of *Petunia inflata* S6 and *Petunia axillaris* N to identify and characterize S-RNase and SLF genes in both species. Each genome contains a single S-RNase allele, as expected for these homozygous lines. The S-RNase allele in *Petunia inflata* appears to be a previously uncharacterized S-RNase allele, whereas the S-RNase allele in *Petunia axillaris* is identical to the previously reported S_1-RNase of *Petunia hybrida.* The latter observation confirms that self-incompatibility in *Petunia hybrida* can be inherited from both progenitor species. We have identified 29 putative S-Locus F-box genes in *Petunia inflata*, the majority of which are present as two closely related copies and 19 SLF genes in the genome of *Petunia axillaris*. The SLF genes in these two species represent at least 20 different SLF variants. In both species, multiple SLF genes are linked on the same scaffold, and in *Petunia axillaris* SLF10 is linked to the S_{ax1}-RNase. These data provide a valuable resource for future studies to assemble and characterize an entire S-locus, as well as to determine the molecular basis of self versus non-self recognition in GSI.

INTRODUCTION

Gametophytic self-incompatibility (GSI), a genetic mechanism that acts to prevent self-fertilization in many angiosperms, is based on the ability of the pistil to selectively inhibit growth of “self” pollen while allowing the growth of “non-self” pollen. During pollination, pollen grains germinate on the stigmatic surface, producing pollen tubes that enter the pistil and begin growth through the transmitting tract of the style. In incompatible pollinations in *Petunia*, pollen tube growth is arrested in the upper third of the style via the action of the S-RNase, an abundant, style-expressed ribonuclease that is imported into both incompatible and compatible pollen tubes. Current models for GSI propose that pollen-tube growth inhibition is due to the cytotoxic action of the S-RNase, which degrades pollen tube RNA, inhibiting protein synthesis. During a compatible pollination, the action of the S-RNase is inhibited by a SCF^{SLF} E3 ubiquitin ligase complex. This complex is proposed to contain one or more F-box proteins (SLF), a Culllin protein, the Skp1-like protein SSK1 and either Rbx1 or SBP1 RING-domain proteins (Zhao et al., 2010; Meng et al., 2011; Sims, 2012; Li et al., 2014).

Estimated to occur in up to three-quarters of eudicot families (Igic & Kohn 2001), S-RNase-based Gametophytic self-incompatibility (GSI) was first described in *Petunia hybrida*, by Darwin:

> “....protected flowers with their own pollen placed on the stigma never yielded nearly a full complement of seed; whilst those left uncovered produced fine capsules, showing that pollen from other plants must have been brought to them, probably by moths. Plants growing vigorously and flowering in pots in the green-house, never yielded a single capsule; and this may be attributed, at least in chief part, to the exclusion of moths.” (Darwin, 1891).

Petunia, primarily *Petunia hybrida, Petunia inflata* and *Petunia axillaris* has been one of the primary systems used to study the genetics and mechanism of S-RNase-based self-incompatibility ever since that first description (Sims & Robbins 2009). Researchers such as Mather (1943), Linskens (1975), and de Nettancourt (1977), determined that gametophytic SI in Petunia was governed by a single, multiallelic S-locus, and that recognition and rejection of self-pollen was controlled gametophytically by alleles expressed in pollen. Mutations unilaterally inactivating self-incompatibility in pollen (pollen-part mutations) have been identified in *Petunia inflata* and were associated with centric chromosomal fragments (Brewbaker and Natarajan, 1960). Tetraploid plants with diploid heteroallelic pollen have been shown to be self-compatible, due to "competitive interaction' in pollen. Shivanna and Rangaswamy (1969) demonstrated that pollination of immature styles could be used to overcome self-incompatibility, a phenomenon now understood to result from low-level expression of the S-RNase early in the development of the style. Ascher (1984) demonstrated quantitative variation in the strength of the self-incompatibility reaction, which he termed pseudo-self-compatibility. More recently, key genes involved in regulating recognition and rejection of pollen in GSI were either first discovered or functionally tested in *Petunia* (Clark et al., 1990; Ioerger et al., 1991; Sims & Ordanic 2001; Lee et al., 1994; Qiao et al., 2004; Sijacic et al., 2004; Kubo et al., 2010; Zhao et al., 2010; Li et al., 2014; Willams et al., 2014a, 2014b; Kubo et al., 2015).

To date, progress on investigating certain aspects of GSI function and mechanisms has been limited due to the lack of assembled and annotated genome and transcriptome sequences for *Petunia.* Kubo et al., (2010) and Williams et al., (2014a) demonstrated that multiple S-locus F-box (SLF) proteins acted to recognize and inhibit overlapping sets of non-self S-RNase. Although different approaches had previously identified up to 10 SLF variants, several questions regarding the role of SLF proteins cannot be answered without a high-quality assembled genome. Among these questions are: How may SLF genes are found in the *Petunia* genome? Are all SLF variants present and expressed in different S-haplotypes and/or *Petunia* species? Are all SLF variants linked to the S-locus and to each other? What is the sequence variability of different SLF alleles and SLF variants, and can that sequence variability be used to identify protein regions involved in binding to the S-RNase? Other questions about the organization and expression of GSI require a high-quality assembled genome. In previous work, Wang et al (2003, 2004) attempted to use chromosome walking approaches to characterize the S-locus of S_2-*Petunia inflata*. Although these authors estimated the size of the S-locus in *Petunia inflata* as at least 4.4 Mb, and were able to identify and partially sequence an 881 kb contig containing the S_2-SLF1 and S_2-RNase genes linked within 160 kb, they were unable to fully assemble the complete S-locus due to the presence of highly repetitive DNA sequences. The S-locus in the Solanaceae has been shown to be subcentromeric (Tanksley & Loaiza-Figueroa 1985; ten Hoopen et al., 1998; Entani et al., 2000) and is in a region of suppressed recombination, which in *Petunia*, is located on chromosome III. Thus, having a fully sequenced and assembled genome should aid in approaches to fully assemble and characterize a complete S-locus for *Petunia.*

We used BLASTn and BLASTp queries of the assembled genomic DNA sequences for *Petunia inflata* S6 and *Petunia axillaris* N followed by manual annotation and comparison to known sequences to identify S-RNase and SLF genes in both genomes. These approaches confirmed the existence of known SLF variants in *Petunia* and identified a number of new SLF variants. We identified 29 putative S-Locus F-box genes in *Petunia inflata* and 19 SLF genes in the genome of *Petunia axillaris*, representing at least 20 different SLF variants. In both species, multiple SLF genes are linked on the same scaffold, and in *Petunia axillaris* SLF10 is linked to the S_{ax1}-RNase. The

S-RNase in S6 *Petunia inflata* appears to be a previously uncharacterized S-RNase. In a surprising and fortuitous observation, we found that the S-RNase gene in *Petunia axillaris* N is identical to that in S_1-*Petunia hybrida* (Clark et al., 1990). This finding strongly suggests that the S-haplotypes are identical in these two lines, and indicates that self-incompatibility in *Petunia hybrida* can be inherited from both parents (Ando et al., 1998). Together with the characterization of other genes involved in GSI interactions, the sequence data obtained from the Petunia Genome Project will provide invaluable resources for further investigations into the organization, expression and function of genes governing self versus non-self recognition in gametophytic self-incompatibility.

RESULTS

Identification of the S-RNase allele in *Petunia axillaris*

BLASTn and BLASTp searches of the *Petunia axillaris* assembled genome gave a single hit that showed 100% DNA sequence identity to the S_1-RNase mRNA of *Petunia hybrida* (Clark et al., 1990). Comparison of this genomic DNA region with the sequenced S_1-RNase gene of *Petunia hybrida* (U07362) showed nearly 100% DNA sequence identity (3119 of 3221 bases) across the entire coding region, and 99% identity when 5' and 3' flanking regions are included (Figure 1). The position of the intron and splice sites were identical between the S_1-RNase gene of *Petunia hybrida* and its homolog in *Petunia axillaris.* In spite of these minor differences, the predicted amino acid sequence for S_1-*Petunia hybrida* and the S-RNase of *Petunia axillaris* are identical (Figure 2). We have therefore tentatively named the S-RNase gene in *Petunia axillaris* N as S_{ax1}-RNase to illustrate this identity. Because the S-RNase genes in S_1 *Petunia hybrida* and S_{ax1} *Petunia axillaris* are identical, this finding demonstrates that self-incompatibility in *Petunia hybrida* can be inherited from both of the progenitor species (e.g., Ando et 1998). This finding also suggests that the S-locus of S_1 *Petunia hybrida* and that of S_{ax1} *Petunia axillaris* encode the same haplotype. Indeed, PCR primers based on SLF gene sequences from *Petunia axillaris* (see below) have been successfully used to clone multiple SLF variant genes from a pollen cDNA library of S_1S_1 *Petunia hybrida* (Sims and Ordanic 2001; Qi & Sims, in preparation). Figures 1 and 2 show DNA and protein sequence alignments of the S_{ax1} and S_1-RNases.

```
Sax1    1   AAGCTTTGCACTTTCTGTTGACTGTATGCCATTGTTACCACCTTAAAGGAGCTATCCATT  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1    1   AAGCTTTGCACTTTCTGTTGACTGTATGCCATTGTTACCACCTTAAAGGAGCTATCCATT  60

Sax1   61   TTTTCCATGAACATGCTGAGAAAAGGACCAAATGAAGGAAACTGGACAAGATCAGAGTCA  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1   61   TTTTCCATGAACATGCTGAGAAAAGGACCAAATGAAGGAAACTGGACAAGATCAGAGTCA  120

Sax1  121   AAGGTGTATTACAAAAGATTATCCTTCTCTCCTGCTATGGACTTCCATCTCAGGAAAAAG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  121   AAGGTGTATTACAAAAGATTATCCTTCTCTCCTGCTATGGACTTCCATCTCAGGAAAAAG  180

Sax1  181   ATAGAGCTTTGTTCATGTACTCACATCCCAACTCTAGAAAATGTTCCAACTTAGTTGTAA  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  181   ATAGAGCTTTGTTCATGTACTCACATCCCAACTCTAGAAAATGTTCCAACTTAGTTGTAA  240

Sax1  241   ACCTGATATAGTTTACAAATTATCGTATTGAATATTCCCTCAGTGATACTTCATTATCTT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  241   ACCTGATATAGTTTACAAATTATCGTATTGAATATTCCCTCAGTGATACTTCATTATCTT  300

Sax1  301   CTCTATGTAGTTGACCGACCATTTAATTGTAAAAGTATTATGTTTCAATGAAAGGTGAAG  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  301   CTCTATGTAGTTGACCGACCATTTAATTGTAAAAGTATTATGTTTCAATGAAAGGTGAAG  360

Sax1  361   AGTTAAATTCTTGTACAGGATGTCGAGGGAGGAAAATGATTCATCTCCCATAGCAACTCA  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  361   AGTTAAATTCTTGTACAGGATGTCGAGGGAGGAAAATGATTCATCTCCCATAGCAACTCA  420

Sax1  421   TGAAGCCTTGAAATTTATCGAGTAAATATAttttttctttttttACTTTGTGCTGATCAATT  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  421   TGAAGCCTTGAAATTTATCGAGTAAATATATTTTTCTTTTTTACTTTGTGCTGATCAATT  480

Sax1  481   AACATTTTCTTAATCTTGTATAATCTACTACCACTCATATTATGTGTAAATGTTTCTATT  540
            |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
PhS1  481   AACATTTTCTTAATCTTGTATAA-CTACTACCACTCATATTATGTGTAAATGTTTCTATT  539
                                  *
Sax1  541   CTAAATCTTAAAATTCTCGTGTTGCACGGGCATCAAAAATATTAATTTATTTTTCTAAAT  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  540   CTAAATCTTAAAATTCTCGTGTTGCACGGGCATCAAAAATATTAATTTATTTTTCTAAAT  599

Sax1  601   TTACATAATTTAATCATGCTACAACCATTCGTAACTTTCTATAATGATTTACAatatata  660
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
PhS1  600   TTACATAATTTAATCATGCTACAACCATTTGTAACTTTCTATAATGATTTACAATATATA  659
                                        *
Sax1  661   tatatatatatatatatatata----------------------tatatatatatata  696
            ||||||||||||||||||||||                      ||||||||||||||
PhS1  660   TATATATATATATATATATATATATATATATATATATATATATTTATATATATATATA  719

Sax1  697   tatttatatatatataaaacatactatctattcaatatatatatgtgatatttttaaatt  756
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  720   TATTTATATATATATAAAACATACTATCTATTCAATATATATATGTGATATTTTTAAATT  779



Sax1  757   gaatatatatatttatataatatttttaacatattgtatatatagtatttattcaattaa  816
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  780   GAATATATATATTTATATAATATTTTTAACATATTGTATATATAGTATTTATTCAATTAA  839
```

```
Sax1  817   tatttttaacatactatatattcaatatttatttaatattttaacacaatatatcacttt  876
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  840   TATTTTTAACATACTATATATTCAATATTTATTTAATATTTTAACACAATATATCACTTT  899

Sax1  877   ttgtatatgttgtgtattCAACATTTATTTACATATGTTTTCTTCCCACACAGCTAAATT  936
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  900   TTGTATATGTTGTGTATTCAACATTTATTTACATATGTTTTCTTCCCACACAGCTAAATT  959

Sax1  937   GTAAGGTATTTGAAATAAAATCAGGGCCATCTGTTGTCGGCAACTGAGCCATCCACCTGT  996
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  960   GTAAGGTATTTGAAATAAAATCAGGGCCATCTGTTGTCGGCAACTGAGCCATCCACCTGT  1019

Sax1  997   CTACAACAACAACAACATACCTAGTGTAATTCCATAAGTGGGTTTAGGAAACTGAGATGT  1056
            |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
PhS1  1020  CTACAACAACAACAACATACCTAGTGTAATTCCATAAGTGGATTTAGGAAACTGAGATGT  1079
                                                    *
Sax1  1057  ACGCAGAACTTACCCCACCAGAATGAAGAGATTGTTTCCGAAAGACCCTCGGCTaaaaaa  1116
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1080  ACGCAGAACTTACCCCACCAGAATGAAGAGATTGTTTCCGAAAGACCCTCGGCTAAAAAA  1139

Sax1  1117  aCATATTTGAAAttttttttAAGACAAACCAAATATTTTGaaaaaagataaaaacgaata  1176
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1140  ACATATTTGAAATTTTTTTTAAGACAAACCAAATATTTTGAAAAAAGATAAAAACGAATA  1199

Sax1  1177  agttaaaaaaTACCATTAATGCTCAAGTTCTCATATAAAAACTACATAACCAGGGAAAAC  1236
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1200  AGTTAAAAAATACCATTAATGCTCAAGTTCTCATATAAAAACTACATAACCAGGGAAAAC  1259

Sax1  1237  ACAAGACGTCAATAGCGATAAGGACAAGATAAAACTACTATGCATAAGAATACTACCGCT  1296
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1260  ACAAGACGTCAATAGCGATAAGGACAAGATAAAACTACTATGCATAAGAATACTACCGCT  1319

Sax1  1297  AAAATGTCAACAATCAATCGTCTTCTACCTAACCTTCTACCATAATCCCAGACCTCCACG  1356
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1320  AAAATGTCAACAATCAATCGTCTTCTACCTAACCTTCTACCATAATCCCAGACCTCCACG  1379

Sax1  1357  CTTTCCTGTCAATGGTCATGTCCTCGGTGATCTAGATGTGAGTCATATCATGTCGAATCA  1416
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1380  CTTTCCTGTCAATGGTCATGTCCTCGGTGATCTAGATGTGAGTCATATCATGTCGAATCA  1439

Sax1  1417  CCTCGCCCCAATTCTTCTTTGGTCTACCTCTACCTCTCTGCAGACCTAGCACAACCAGCC  1476
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1440  CCTCGCCCCAATTCTTCTTTGGTCTACCTCTACCTCTCTGCAGACCTAGCACAACCAGCC  1499

Sax1  1477  TCTCACACCTCCTCACTGGCGCATCGGTGCCCCTTCTCTTTGCATGGCCGAACCATCGCA  1536
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1500  TCTCACACCTCCTCACTGGCGCATCGGTGCCCCTTCTCTTTGCATGGCCGAACCATCGCA  1559

Sax1  1537  ATCTCACTTCTCGCATCTTGTCCACTACTGAGGCCACTCTCACCTTATCACGAATGGCCT  1596
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1560  ATCTCACTTCTCGCATCTTGTCCACTACTGAGGCCACTCTCACCTTATCACGAATGGCCT  1619

Sax1  1597  CATTCCTAATCTTATCCAACTCGTGTGCCCACACAGCCATCTAAGCTGTTAACATACATA  1656
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1620  CATTCCTAATCTTATCCAACTCGTGTGCCCACACAGCCATCTAAGCTGTTAACATACATA  1679

Sax1  1657  TTATTTTAACATATCCAACCTCATGTGTCCACACATCCACGTGTCTCATATATATTATTT  1716
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1680  TTATTTTAACATATCCAACCTCATGTGTCCACACATCCACGTGTCTCATATATATTATTT  1739
```

```
Sax1  1717  TTAAAAGGCTGTTAACACTATTTAAGTGCCCTAATCCACTTCTTCTAGTACAGATTCTAG  1776
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1740  TTAAAAGGCTGTTAACACTATTTAAGTGCCCTAATCCACTTCTTCTAGTACAGATTCTAG  1799

Sax1  1777  CTTGaaaaaaaaaTTAGTTAGGTTAGTGAAGTGATAAGTCCTATATTAACCCATTCCACT  1836
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1800  CTTGAAAAAAAAAATTAGTTAGGTTAGTGAAGTGATAAGTCCTATATTAACCCATTCCACT  1859

Sax1  1837  GAAAATACGATTATTATATATTGCCATATAGCAAAAGGAAGGAACTAAACATGAGTTGTT  1896
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1860  GAAAATACGATTATTATATATTGCCATATAGCAAAAGGAAGGAACTAAACATGAGTTGTT  1919

Sax1  1897  CAAACTTTAGAATGTTCAAGTTACAGCTGGCGTCAGTTTTATGTGTTTTTCTTTTTGCTT  1956
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1920  CAAACTTTAGAATGTTCAAGTTACAGCTGGCGTCAGTTTTATGTGTTTTTCTTTTTGCTT  1979

Sax1  1957  GCTCTCCAATTTCTGGGTCTTTCGACCACTGGCAACTCGTTTTAACATGGCCTGCAGGTT  2016
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  1980  GCTCTCCAATTTCTGGGTCTTTCGACCACTGGCAACTCGTTTTAACATGGCCTGCAGGTT  2039

Sax1  2017  ATTGCAAAGTTAAAGGTTGTCCGAGACCAGTAATTCCGAACGACTTTACTATTCATGGTC  2076
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2040  ATTGCAAAGTTAAAGGTTGTCCGAGACCAGTAATTCCGAACGACTTTACTATTCATGGTC  2099

Sax1  2077  TTTGGCCAGATAGCATTTCCGTCATAATGAATAACTGCGATCCGACTAAAACGTTTGTGA  2136
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2100  TTTGGCCAGATAGCATTTCCGTCATAATGAATAACTGCGATCCGACTAAAACGTTTGTGA  2159

Sax1  2137  CGATCACTGTAAGTTTATAACATTATCTTCTTAAGCGATTGTAAttttttttttctcatt  2196
            |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
PhS1  2160  CGATCACTGTAAGTTTATAACATTATCTTCTTAAGCGATTGTAAATTTTTTTTTCTCATT  2219
                                                        *
Sax1  2197  tattgttttgctttttccttttcttttttatttttgtttcttgAATAACCTGCAGCCTAATGT  2256
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2220  TATTGTTTTGCTTTTTCCTTTCTTTTTATTTTGTTTCTTGAATAACCTGCAGCCTAATGT  2279

Sax1  2257  TTATAGGAAATAAATCAAATAACCGAACTGGAGAAGCGCTGGCCTGAATTGACTACTACC  2316
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2280  TTATAGGAAATAAATCAAATAACCGAACTGGAGAAGCGCTGGCCTGAATTGACTACTACC  2339

Sax1  2317  GCACAATTTGCTTTAACGAGTCAATCTTTCTGGAGATATCAATACGAAAAGCATGGAACA  2376
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2340  GCACAATTTGCTTTAACGAGTCAATCTTTCTGGAGATATCAATACGAAAAGCATGGAACA  2399

Sax1  2377  TGTTGTTTTCCTGTCTACAGTCAATCAGCATATTTTGATTTTGCTATAAAATTAAAAGAC  2436
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2400  TGTTGTTTTCCTGTCTACAGTCAATCAGCATATTTTGATTTTGCTATAAAATTAAAAGAC  2459

Sax1  2437  AAGACTGATCTGTTGAGTATTCTCAGAAGTCAAGGTGTTACTCCGGGATCAACTTATACT  2496
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2460  AAGACTGATCTGTTGAGTATTCTCAGAAGTCAAGGTGTTACTCCGGGATCAACTTATACT  2519

Sax1  2497  GGAGAAAGAATCAACAGTTCCATCGCGTCAGTAACCCGAGTGAAACCTAACCTCAAGTGC  2556
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2520  GGAGAAAGAATCAACAGTTCCATCGCGTCAGTAACCCGAGTGAAACCTAACCTCAAGTGC  2579
```

```
Sax1  2557  CTTTATTATCGAGGCAAATTGGAATTAACTGAGATAGGAATATGTTTTGACCGAACGACA  2616
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2580  CTTTATTATCGAGGCAAATTGGAATTAACTGAGATAGGAATATGTTTTGACCGAACGACA  2639

Sax1  2617  GTTGCTATGATGTCGTGTCCTCGGATTAGTACGTCATGCAAATTCGGGACAAATGCGAGG  2676
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2640  GTTGCTATGATGTCGTGTCCTCGGATTAGTACGTCATGCAAATTCGGGACAAATGCGAGG  2699

Sax1  2677  ATTACGTTTCGACAGTGAGAAACGTTCGATTTCATGTTCTTTCTTTCTAATTTTATGCAG  2736
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2700  ATTACGTTTCGACAGTGAGAAACGTTCGATTTCATGTTCTTTCTTTCTAATTTTATGCAG  2759

Sax1  2737  AGTATAATAAAGGAGGTTTTACACTGTATACCCAAATTTTATAAAATATTACTATATATT  2796
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2760  AGTATAATAAAGGAGGTTTTACACTGTATACCCAAATTTTATAAAATATTACTATATATT  2819

Sax1  2797  TTTACTCATTTTAAACTTCATTTCACTTCATGCCCATTTTGACATAAATATTTATACTCC  2856
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2820  TTTACTCATTTTAAACTTCATTTCACTTCATGCCCATTTTGACATAAATATTTATACTCC  2879

Sax1  2857  ATGTCCATTTTGACATAAATATTTACACAACTATCCATGGTACAAAAAGTTCACCTTTTA  2916
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2880  ATGTCCATTTTGACATAAATATTTACACAACTATCCATGGTACAAAAAGTTCACCTTTTA  2939

Sax1  2917  GCCACGATGAACCGTAGATAGTGCCACCGTGGCTACCGGTCGCCATTAGCCACGATTAAT  2976
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  2940  GCCACGATGAACCGTAGATAGTGCCACCGTGGCTACCGGTCGCCATTAGCCACGATTAAT  2999

Sax1  2977  CATGGCTAATGGCCGCGTTTAGCCGTGGTCTTACAATAGGTTCTGGAACCGAGGCTAACA  3036
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  3000  CATGGCTAATGGCCGCGTTTAGCCGTGGTCTTACAATAGGTTCTGGAACCGAGGCTAACA  3059

Sax1  3037  ACCTATAGATTTCATGACCCAATTAACATTTACCGTGTAAGACCCATGTCTAatttaatt  3096
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  3060  ACCTATAGATTTCATGACCCAATTAACATTTACCGTGTAAGACCCATGTCTAATTTAATT  3119

Sax1  3097  tttatattttactatttttatattaATCACGGTTCAACCGTGGCTTAGTAACCCCTTCTT  3156
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PhS1  3120  TTTATATTTTACTATTTTTATATTAATCACGGTTCAACCGTGGCTTAGTAACCCCTTCTT  3179

Sax1  3157  aaaaaaaaTTTAATATTTAAATGAAAGACAGTTGCTGAATTC  3198
            |||||||||||||||||||||||||||||||||||||||||||
PhS1  3180  AAAAAAAATTTAATATTTAAATGAAAGACAGTTGCTGAATTC  3221
```

Figure 1. Alignment of the genomic DNA sequences of the S_1-RNase gene of *Petunia hybrida* (PhS1) and the homologous region for S_{ax1}-*Petunia axillaris* (Sax1). Start and stop codons for the S_1-RNase are shown in boldface and the exon regions of the S_1-RNase are underlined. Individual base differences are shown with an asterisk.

```
Sax1    1  MFKLQLASVLCVFLFACSPISGSFDHWQLVLTWPAGYCKVKGCPRPVIPNDFTIHGLWPD  60
           MFKLQLASVLCVFLFACSPISGSFDHWQLVLTWPAGYCKVKGCPRPVIPNDFTIHGLWPD
PhS1     1  MFKLQLASVLCVFLFACSPISGSFDHWQLVLTWPAGYCKVKGCPRPVIPNDFTIHGLWPD  60

Sax1   61  SISVIMNNCDPTKTFVTITEINQITELEKRWPELTTTAQFALTSQSFWRYQYEKHGTCCF  120
           SISVIMNNCDPTKTFVTITEINQITELEKRWPELTTTAQFALTSQSFWRYQYEKHGTCCF
PhS1   61  SISVIMNNCDPTKTFVTITEINQITELEKRWPELTTTAQFALTSQSFWRYQYEKHGTCCF  120

Sax1  121  PVYSQSAYFDFAIKLKDKTDLLSILRSQGVTPGSTYTGERINSSIASVTRVKPNLKCLYY  180
           PVYSQSAYFDFAIKLKDKTDLLSILRSQGVTPGSTYTGERINSSIASVTRVKPNLKCLYY
PhS1   121  PVYSQSAYFDFAIKLKDKTDLLSILRSQGVTPGSTYTGERINSSIASVTRVKPNLKCLYY  180

Sax1   181  RGKLELTEIGICFDRTTVAMMSCPRISTSCKFGTNARITFRQ  222
           RGKLELTEIGICFDRTTVAMMSCPRISTSCKFGTNARITFRQ
PhS1   181  RGKLELTEIGICFDRTTVAMMSCPRISTSCKFGTNARITFRQ  222
```

Figure 2. Alignment of the predicted amino acid sequence from S_{ax1}-*Petunia axillaris* and the known amino acid sequence for S_1-*Petunia hybrida*

Identification of the S-RNase allele in *Petunia inflata*

BLASTp and BLASTn queries of the assembled *Petunia inflata* genome gave a single strong hit, as expected for a homozygous line. Identification of open reading frames and comparison with known S-RNase alleles demonstrated that the identified sequence region encoded a S-RNase that was most closely related (91% amino acid sequence identity) to the S_3-RNase of *Petunia hybrida* (Clark et al., 1990; Genbank AAA60466). Alignment of the protein sequences of the S6-*Petunia inflata* S-RNase with that of S_3-RNase of *Petunia hybrida* (Figure 3) showed that the majority of amino acid sequence differences were in the HVa and HVb regions demonstrated to be sufficient to distinguish S-RNase alleles in some cases (Matton et al., 1997, 1999). Thus, the S-RNase in this line of *Petunia inflata* likely represents a previously uncharacterized S-locus haplotype.

```
                                        C1                                C2
PinfS6   1    MVRLQLLSALFILLFSLSPVSANFDYFQLVLTWPASFCYPKNKCQRRSNNFTIHGLWPEK  60
              M RLQL+SA FILLFSLSPVSANFDYFQLVLTWPASFCYPKNKCQRRSNNFTIHGLWPEK
PhybS3   1    MFRLQLISAFFILLFSLSPVSANFDYFQLVLTWPASFCYPKNKCQRRSNNFTIHGLWPEK  60

                          HVa                       HVb             C3
PinfS6   61   KRFRLEFCTGDEYARFLKEDSIINDLERHWIQMRFDEKYAKDKQPLWEHEYTKHGICCSN  120
              KRFRLEFCTGD+Y RFL+ED+IIN LERHWIQMRFDE YA  KQPLWEHEY +HGICC N
PhybS3   61   KRFRLEFCTGDKYKRFLEEDNIINVLERHWIQMRFDETYANTKQPLWEHEYNRHGICCKN  120

                          C4
PinfS6   121  LYKQREYFLLAMRLKDKLDLLTILRNHGITPGTKHTFGEIQKAIKTVTNNNDPDLKCVEN  180
              LY Q+ YFLLAMRLKDKLDLLT LR HGITPGTKHTFGEIQKAIKTVT+NNDPDLKCVEN
PhybS3   121  LYDQKAYFLLAMRLKDKLDLLTTLRTHGITPGTKHTFGEIQKAIKTVTSNNDPDLKCVEN  180

                     C5
PinfS6   181  IKGVMELNEIGICYTPAADRFDRCRHSNTCDETSSTKILFRG  222
              IKGVMELNEIGICYTPAADRFDRCRHSNTCDETSSTKILFRG
PhybS3   181  IKGVMELNEIGICYTPAADRFDRCRHSNTCDETSSTKILFRG  222
```

Figure 3. Alignment of the predicted amino acid sequence for the S-RNase of S6 *Petunia inflata* and the S_3-RNase of *Petunia hybrida.* Conserved domains C1 through C5 are underlined and labeled. The two hypervariable domains are labeled and shaded in yellow.

Identification and Characterization of SLF genes in *Petunia axillaris*

Multiple sequence alignment of previously identified SLF sequences was used to produce consensus sequences for SLF1 through SLF8 variants (Figure 4). Those consensus sequences along with previously identified SLF and SLFL DNA sequences were used as BLASTn queries to identify putative SLF homologs on numerous scaffolds. Queries with SLF2, SLF3, SLF5, SLF6, SLF8 and S_2-SLFx showed strong homology to only a single gene region on different scaffolds. Manual annotation and BLASTp queries of the NCBI protein database confirmed that each of these regions corresponded to previously identified SLF variants from *Petunia*. Conversely, SLF1, SLF4, SLF7, and SLF-S3B queries identified from two to five separate gene sequences on different scaffolds. Manual annotation and characterization of these hits demonstrated that all appeared to be true SLF genes. Criteria included the presence of a single open reading frame with no introns (SLF genes lack introns), the presence of a well-defined N-terminal F-box domain, the absence of any other protein interaction domains and significant amino acid sequence identity to previously identified SLF proteins.

SLF1 **AAGAATAAAGGATGGCGAATGGTATTTTAAAGAAATTGCC**

SLF2 **CTTTCAGATGTTTATTGGGATCCTCCTA**

SLF3 **GATTAATATATTATTTAGGATTCCCGTGAAATCTCT**

SLF4 **ACACCATTCTCCAAAGTGCAATGAAATTATATTGTAAAGAATACAA**

SLF5 **GCCGTTACAGTGCCAATATTATGAAGATGCCA**

SLF6 **GAGATGAATATATTCTGTTAAAGCGTTGCTTATACAAGAAAACAACCAAT**

SLF7 **AGCACAGAGACTTTTCGCAATATGAAAATGCCGGATGCGTGTCATTTCAAA**

SLF8 **TGAATACGCGAGGAATAAGCTTTTGCGTCAAACTCAAAGGATC**

Figure 4: Consensus sequences from aligned SLF gene variants used as BLASTn probes. Multiple sequence alignment of previously-identified SLF variants SLF1 through SLF8 were used to produce a consensus sequence region unique to each SLF. These consensus sequences were used as the initial probes for BLASTn queries of the assembled genomes.

Table 1. Summary of SLF and S-RNase BLAST queries in *Petunia axillaris*. Scaffold identities and coordinates and best BLASTp hits are shown for identified SLF and S-RNase genes for "N/S26" *Petunia axillaris*. Linked regions are color-coded

Query	Scaffold	Size	Coordinates	BLASTp Hit	BLASTp %ID	Class
SLF1-1	00326	940666	508809-509978	AGL76530, S5-SLF1	99	SLF1
SFL1-2	00715	718562	578400-579572	ADD21612, S1-SLF1	99	SLF1
SLF2	00070	1471679	699640-700806	BAJ24857, S19-SLF2	95	SLF2
SLF3	00326	940666	596118-597275	BAJ24858, S5-SLF3	98	SLF3
SLF4-1	00671	726586	152315-153526	BAJ24865, S7-SLF4	94	SLF4
SLF4-2*(SLF12)	00326	940666	531044-532225	BAJ24865, S7-SLF4	78	SLF12
SLF5	03085	28205	6171-7340	BAJ24871, S7-SLF5	95	SLF5
SLF6	00172	1583766	619079-620257	BAJ24881, S19-SLF6	96	SLF6
SLF7-1(SLF14)	00172	1583766	708846-710139	ABR18788, S2-DD8	77	SLF14
SLF7-2(SLF16)	01162	482111	154759-155919	ABR18785, S2-DD5	76	SLF16
SLF8	00326	940666	464377-465552	ABX82525, S2-SLF8	90	SLF8
SLF-S3B-1 (SLF10)	00336	1804933	636123-637262	ABR15914,A-134	98	SLF10
SLF-S3B-2 (SLF18)	00671	726586	549469-550611	ABR15914, A-134	74	SLF18
SLF-S3B-3**(SLF19_1)	75437	1274	113-1273	ABR18782, S1-DD2	70	SLF19_1
SLF-S3B-4** (SLF19_2)	01590	192725	30722-31882	ABR18782, S1-DD2	70	SLF19_2
SLF-S3B-5 (SLF20)	01394	65038	63647-64564	ABR18782, S1-DD2	69	SLF20
S_2-SLFx (SLF13)	00326	940666	686239-687402	AHF49538, S2-SLFx	99	SLF13
SLF11	00426	634794	185428-186405	BAQ18962	95	SLF11
SLF15	01310	413452	331303-332496	AIK66473	94	SLF15
SLF17	01162	482111	154759-155919	AIK66479	99	SLF15
S-RNase	00336	1804933	449226-450052	AAA60465, S_1-RNase	100	N/A

*SLF4-2 sequence is in FBA_1 superfamily but lacks a clear N-terminal F-box domain.
**Identical sequences

Several of the identified genes appeared to be new SLF variants, including SLF4_2, SLF7-1, SLF7-2, SLF-S3B-1, SLF-S3B-2, SLF-S3B-3/4, SLF-S3B-5 and S2-SLFx. These have tentatively been assigned new SLF variant numbers by comparison with assignments suggested by Williams et al (2014b) and are listed in Table 1. Together these data indicate that there appear to be a minimum of 17 SLF variants present in *Petunia axillaris.* Several of the SLF genes are linked on the same scaffold assembly, as shown in Table 1 and Figure 5. Scaffold 00326 encodes five SLF gene sequences including SLF1_1, SLF3, SLF8, SLF13 and SLF12. SLF4 and SLF18 are also linked on a separate scaffold, as are SLF6 and SLF14. Finally, the S_{ax1}-RNase of *Petunia axillaris* is linked to SLF10 at a distance of approximately 187 kb. Two hits on different scaffolds (00326 and 00715) showed extremely strong (99%) BLASTp identity to previously identified SLF1 variants from *Petunia hybrida*, S5-SLF1 and S1-SLF1. These two SLF proteins differ by 11 out of 389 amino acids. Interestingly, *Petunia axillaris* N is completely self-compatible whereas S_1-*Petunia hybrida* is completely self-incompatible. As one common cause of the breakdown of self-incompatibility is a partial duplication of the S-locus that results in heteroallelic pollen expressing different SLF alleles (Golz et al, 2000; Sijacic et al., 2004) it is possible that these sequences reflect a partial duplication of the S-locus in this line. In fact, a naturally occurring example of SLF gene duplication associated with self-incompatibility has been described previously in *Petunia axillaris* (Tsukamoto et al., 2005).

Identification and Characterization of SLF genes in *Petunia inflata*

BLAST queries of the *Petunia inflata* genome assembly, carried out in the same manner as for *Petunia axillaris*, gave an even greater number of putative SLF hits. At least 29 identified genes appear to encode true SLF proteins, using the same criteria as above. Three gene regions (scaffold 01586, SLF7-3; scaffold 01751, SLF12_2 and scaffold 01501, SLF15_2) are apparent pseudogenes, as none encode a functional open reading frame. In the case of SLF7_3 a single nucleotide insertion at position 360 results in a frame-shift. As was the case for *Petunia axillaris*, several of the SLF genes are linked on the same scaffold. We have identified five linkage clusters, including SLF1-1 with SLF8-1 and SLF12_2(ψ), SLF1-2 with SLF12_1, SLF5-2 with SLF14_1, SLF7-3(Ψ) with SLF16_2 and SLF16_1 with SLF13_1.

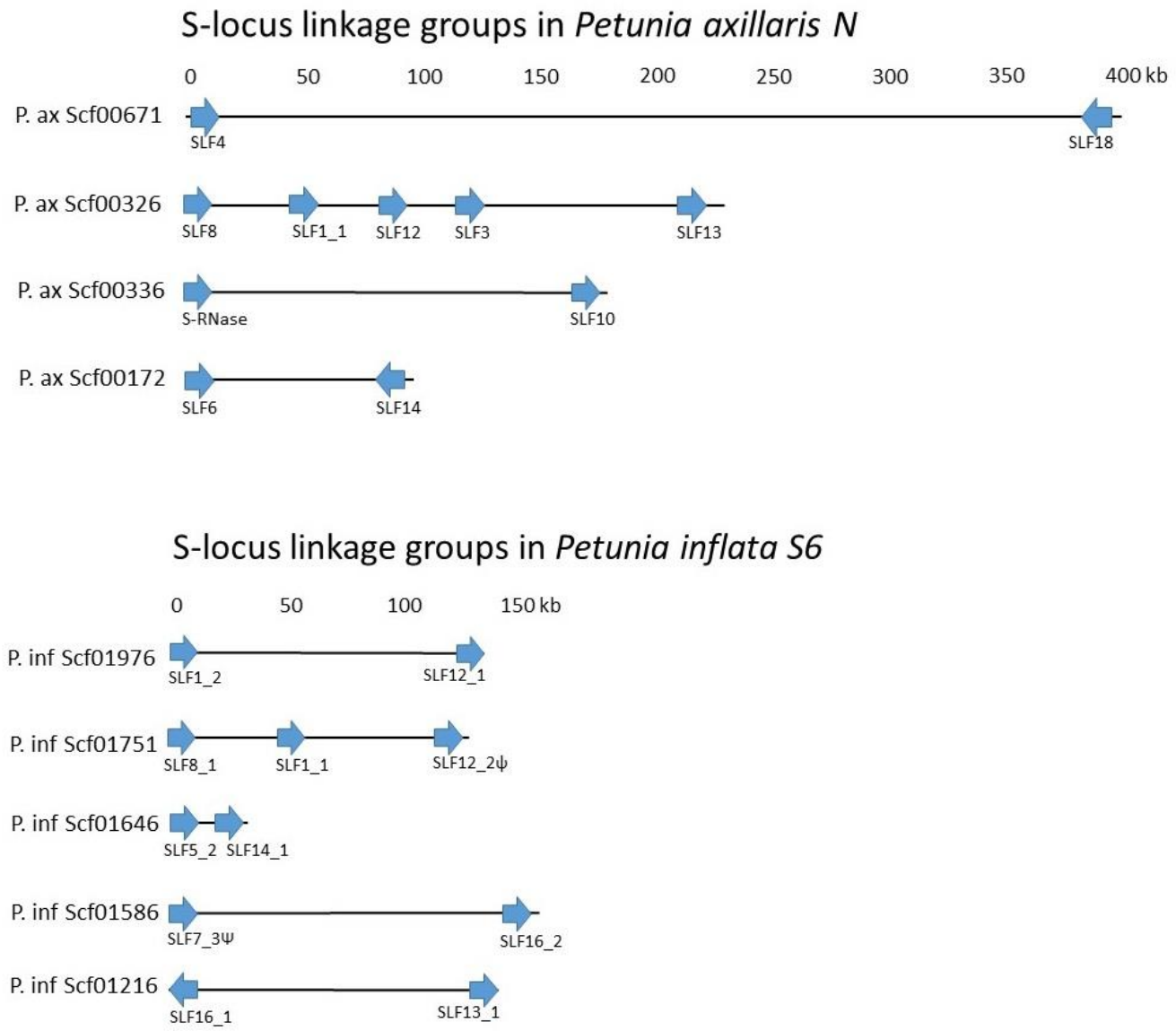

Figure 5. Linkage of SLF variants in *Petunia axillaris* N and *Petunia inflata* S6. The diagrams show a schematic of linked SLF variants found on the same scaffolds for *Petunia axillaris N* and for *Petunia inflata* S6. Individual scaffolds are numbered with the identity and orientation of different SLF variants found on the scaffolds shown by arrowheads. Depiction of the size of individual SLF variants is not to scale. Approximate distances between SLF variants is depicted by the scale at the top of the figure. Only that portion of individual scaffolds containing SLF variants is shown.

Table 2. Summary of SLF and S-RNase BLAST queries in *Petunia inflata*. Scaffold identities and coordinates and best BLASTp hits are shown for identified SLF and S-RNase genes for "S6" *Petunia inflata*. Linked regions are color-coded.

Gene	Scaffold	Size	Coordinates	Best BLASTp Hit	BLASTp %ID	Class
SLF1-1	01751	453812	144819-146092 (+)	AIK66522, S6a-SLF1	99	SLF1
SLF1-2	01976	214490	4694-5969 (+)	ADD21612, S1-SLF1	95	SLF1
SLF2-1	01991	197077	278634-279788 (+)	AIK66498, S3-SLF2	90	SLF2
SLF2-2	01973	296567	153541-154695 (-)	BAQ19037, Sm-SLF2	83	SLF2?
SLF3	01923	159365	5575-6582 (+)	BAQ18957, S10-SLF3	96	SLF3
SLF4-1	00922	184426	64033-65244 (-)	BAJ24865, S7-SLF4	96	SLF4
SLF4-2	05340	224001	129825-131051 (+)	BAJ24865, S7-SLF4	93	SLF4
SLF5-1	01516	657630	644401-645570 (+)	BAQ19039, Sm-SLF5	96	SLF5
SLF5-2	01646	115327	3200-4369 (+)	BAQ19039, Sm-SLF5	96	SLF5
SLF6-1	01506	144469	61311-62492 (+)	BAJ24881, S19-SLF6	95	SLF6
SLF6-2	00996	461697	107014-108195 (-)	BAJ24881, S19-SLF6	96	SLF6
SLF7-1	00080	221173	3893-5071 (-)	ABX82526, S2-SLFLa	99	SLF7
SLF7-2	00526	752603	259704-260882 (-)	BAQ18970, S11-SLF7	98	SLF7
SLF7-3Ψ	01586	859704	334991-336167 (+)	BAQ19025, S0m-SLF7	79	pseudogene
SLF8-1	01751	453812	50755-51939 (+)	BAQ18935, S7-SLF8	97	SLF8
SLF8-2	03408	108618	69122-70303 (+)	BAQ18935, S7-SLF8	88	SLF8?
SLF9-1	01995	82553	62151-63281 (-)	BAQ18945, S9-SLF9	97	SLF9
SLF9-2	01799	514228	140243-141373 (-)	BAQ18945. S9-SLF9	98	SLF9
SLF10	01608	234268	115426-116565 (+)	AAR15915, S2-A134	99	SLF10
SLF11-1	04466	242965	204665-205837 (+)	AIK66454, S6a-SLF11	99	SLF11
SLF11-2	00431	165049	14066-15178 (+)	BAQ18974, S11-SLF11	95	SLF11
SLF12-1	01976	214490	142523-143704 (+)	AIK66458, S6a-SLF12	100	SLF12
SLF12-2Ψ	01751	453812	177587-179429 (+)	AIK66458, S5-SLF12	71	pseudogene
SLF13-1	01216	514760	360226-361392 (+)	AIK66464, S6a-SLF13	99	SLF13
SLF13-2	02268	142811	121932-123095 (-)	BAQ19014, S22-SLF13	96	SLF13
SLF14-1	01646	115327	30658-31851 (+)	AIK66469, S6a-SLF14	99	SLF14
SLF14-2	01246	26952	6131-7324 (-)	AIK66469, S6a-SLF14	99	SLF14
SLF15-1	16932	4714	563-1723 (+)	AIK66475, S12-SLF15	96	SLF15
SLF15-2Ψ	01501	1043979	485372-486154 (-)	AIK66472, S3-SLF15	84	pseudogene

SLF16-1	01216	514760	214995-216155 (-)	AIK66479, S6a-SLF16	100	SLF16
SLF16-2	01586	859704	501814-502983 (+)	BAQ18940, S7-SLF16	77	SLF16
SLF17	00041	2522227	1710310-1711476 (+)	BAQ18941, S7-SLF17	99	SLF17
S6-RNase	01010	272190	153855-154627 (+)	AAA60466, S3-RNase	91	S6-RNase

Phylogenetic comparison of SLF gene sequences

To analyze similarities among different SLF variants, we used MEGA6 (Tamura et al., 2013) to carry out a phylogenetic comparison of all of the SLF variants identified in *Petunia inflata* S6 and *Petunia axillaris* N. This analysis (Figure 6) identified 6 clades supported by bootstrap values greater than 72% of 1000 bootstrap replicates. The identified clades contained the following SLF varants. Clade 1: SLF 7, SLF14, SLF15, SLF16, SLF17. Clade 2: SLF3, SLF11, SLF13. Clade 3: SLF9, SLF10, SLF18, SLF19, SLF20. Clade 4: SLF1, SLF2. Clade 5: SLF8. Clade 6: SLF4, SLF5, SLF6.

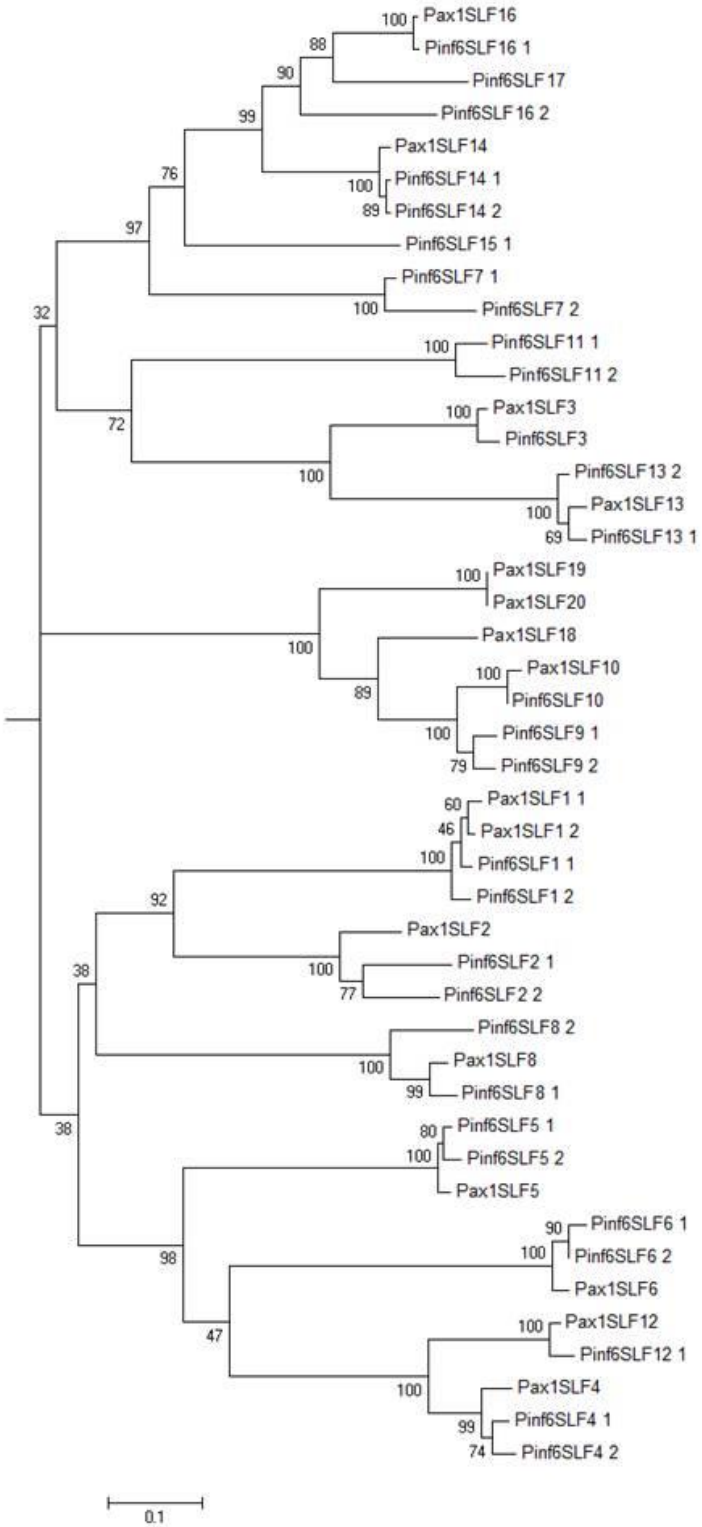

Figure 6. Maximum likelihood phylogenetic relationships between SLF variants from *Petunia axillaris N* and *Petunia inflata* S6 were produced using MEGA6. The scale bar at the bottom shows the number of substitutions per site. Bootstrap values are shown for individual nodes.

DISCUSSION

Having completely assembled and annotated genomic DNA sequences for *Petunia inflata* and *Petunia axillaris* will provide an abundance of information that should enhance investigations on the organization and expression of the S-locus and on the functional mechanisms involved in self versus non-self recognition in gametophytic self-incompatibility. In a preliminary analysis of the sequence assemblies for the two genomes, we carried out BLAST queries to identify S-RNase and SLF genes presumed to be encoded and linked to the S-locus. Numerous SLF genes (19 in *Petunia axillaris* and 29 in *Petunia inflata*) were identified. These included sequences nearly identical to previously identified SLF genes as well as several new SLF genes which likely encode additional SLF variants beyond those currently identified. One of the challenges of GSI research will be to determine if all of the putative SLF genes are expressed and if some or all of them function in self-incompatibility recognition (via interactions with S-RNase proteins). As predicted from current models for the structure of the S-locus, many of the SLF genes identified were linked to each other; at least one SLF gene in *Petunia axillaris* was found on the same DNA scaffold as the S-RNase gene in this species. Although the quality of the sequence assembly for these two genomes did not allow assembly of a complete "virtual" S-locus, it is likely that a complete S-locus can be assembled using these data combined with screens of BAC libraries. Recently, an analysis of SLF genes in the completed tomato and potato genomes has indicated that the approximate size of the S-locus is 14.5 Mb and 17.9 Mb, respectively (Kubo et al., 2015). Adding together all the SLF-bearing scaffold sizes in *Petunia axillaris* gives a total minimum size of approximately 8.0 Mb. A similar analysis of SLF-bearing scaffold sizes in *Petunia inflata* gives a total minimum size of approximately 11.3 Mb; this is consistent with a significant duplication in *Petunia inflata*. Until the scaffolds are connected, it is not possible to accurately compare the scale of these S-loci, but in *Petunia* it would appear to be of comparable size to that observed in *Solanum*.

Analysis of S-RNase-encoding genes in *Petunia inflata* S6 and *Petunia axillaris* N showed single copies of the S-RNase, as expected from the fact that both of these lines were homozygous. The S-RNase in *Petunia inflata* appears to be an as yet uncharacterized S-RNase. The most closely related characterized S-RNase to the one identified is the S_3-RNase of *Petunia hybrida* (Q40875, Okuley and Sims, 1994). As these two S-RNases differ by 20 amino acids, with 10 of the differences occurring in the HVa and HVb hypervariable regions known to play a major role in determining allele specificity (Matton et al, 1997), it is extremely likely that this S-RNase encodes a different recognition specificity than any previously published and characterized S-RNase. By contrast, we found that the S-RNase in N *Petunia axillaris* was nearly identical (99% nucleotide sequence identity across the entire gene and 5' and 3' flanking regions) to that of the S_1-RNase of *Petunia hybrida*. That result shows that self-incompatibility in the cultivated garden petunia can be inherited from both of the progenitor species to *Petunia hybrida*, *Petunia inflata* and *Petunia axillaris*. Because S_1-*Petunia hybrida* is completely self-incompatible and never sets seed on self-pollination, whereas the sequenced *Petunia axillaris* N line is completely self-compatible, setting large seed capsules on selfing (Sims, unpublished), it will be instructive to closely compare the structure and expression of the S-loci in these two lines to determine the molecular basis for GSI breakdown.

We identified at least 20 different SLF variants in the two species analyzed, but not all SLF variants were present in each species. For example, *Petunia axillaris* N lacked SLF7, SLF9 and SLF11 whereas *Petunia inflata* S6 lacked SLF18, SLF19 and SLF20. Whereas the SLF variants in *Petunia* axillaris were mostly single genes, nearly all of the SLF variants in *Petunia inflata* were found as two closely-related copies that mapped to different scaffolds. This suggests that the S-locus in *Petunia inflata* S6 may have undergone a recent duplication event compared with

Petunia axillaris. Supporting that suggestion is the finding that the estimated size of S-locus-containing scaffolds in *Petunia inflata* S6 exceeds 11 Mb whereas that for *Petunia axillaris* N spans just over 8 Mb of DNA.

METHODS

Plant Material

Seeds of *Petunia inflata* line S6 and *Petunia axillaris* line N were obtained from the Free University of Amsterdam. Plants used for DNA extraction were grown axenically in tissue culture containers (Ball Horticulture). Mature plants (leaves and stems) were harvested, flash-frozen in liquid nitrogen and stored at -80C until used for DNA extractions.

DNA Extractions

Plant material (approximately 15 g) was extracted using a modification of methods designed to isolate high molecular weight DNA from nuclei (Fisher and Goldberg, 1982; Carrier et al., 2011). Briefly, frozen plant material was homogenized in a blender with liquid nitrogen until a fine powder was obtained. Powdered material was thawed in 1X SEB plus mercaptoethanol (10 mM Tris pH 8.0, 100 mM KCl, 10 mM Na_2EDTA, 0.5 M sucrose, 4 mM spermidine, 1 mM spermine, 0.13% carbamic acid, 0.25% PVP-40, 0.2% β-mercaptoethanol), then filtered through Nitex mesh. Triton X-100 was added to 0.5% and nuclei isolated and washed by repeated low speed centrifugation and washing with SEB. Nuclei were lysed by adding an equal volume of NLB (2% Sodium N-lauryl sarcosine, 40 mM Na_2EDTA, 0.1 M Tris-HCl pH 8.0 and 1mg/ml proteinase K) followed by incubation at 55 C for 1 hour. Cesium chloride was added to 50% w/w along with ethidium bromide to a final concentration of 0.4%. DNA gradients were centrifuged in a 70.1 Ti rotor at 40,000 rpm for 36 hours followed by re-banding in a VTi 65.2 rotor at 60,000 rpm for 6 hours. Ethidium bromide was removed by extraction with SSC-saturated isopropanol and the remaining solution dialyzed against TNE (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.1 mM EDTA) for 24 hours. DNA was precipitated, washed with 70% ethanol, dried and resuspended in EB (10 mM Tris pH 8.0) to a final concentration of 150 µg/ml.

DNA Sequencing

DNA libraries (two 1000 bp paired-end, two 8 kb and two 15 kb mate-pair) were constructed at the University of Illinois Roy J. Carver Biotechnology Center. Paired-end libraries were sequenced on two lanes of HiSeq 2000 and the mate-pair libraries sequenced on a single HiSeq 2500 lane.

DNA Assembly and Annotation

DNA sequences for both *Petunia inflata* and *Petunia axillaris* were combined with previous raw reads produced by BGI-Shenzen for the Petunia Genome Project and assembled using SOAPdenovo2 (Luo et al., 2010). For *Petunia axillaris*, the Illumina data were combined with PacBio data to produce the final genome assembly. MAKER (Cantarel et al 2008) was used to annotate predicted transcript and protein sequences.

Identification and manual annotation of S-RNase and SLF gene sequences

Known S-RNase sequences (S_1-RNAse, S_3-RNase) from *Petunia hybrida* (Clark et al., 1990) were used in BLASTn and BLASTp (Altschul et al., 1990) searches of the assembled scaffolds for *Petunia axillaris* and *Petunia inflata*. Individual scaffold regions were used in pairwise alignment against known S-RNase mRNA and gene sequences and analyzed for open reading frames. Multiple approaches were used to identify potential SLF coding genes. Initially, all known SLF genes of a particular variant class (SLF1 to SLF8, Kubo et al., 2010) were aligned to identify conserved regions unique to a particular SLF variant class (Figure 4). Those conserved sequences were then used

as BLASTn probes against the assembled genomes to identify scaffold regions homologous to a particular SLF variant. In addition, other known SLFL genes (Wang et al 2003), DDX genes from *Nicotiana alata* (Wheeler and Newbigin 2000), SLF-S3B (Qiao et al 2004) and S2-SLFx (Li et al 2014) were used as BLASTn probes. BLASTn queries used an Expect value of 10. Scaffold regions identified by BLASTn queries were used as input into the Translate program of ExPASy to identify open reading frames. Translated protein sequences were then used in BLASTp queries of GenBank to identify the most closely-related proteins. Conserved domains were identified using a combination of BLASTp and InterPro Scan 5 (Jones et al. 2014).

Phylogenetic comparison of SLF variants

Manually annotated SLF variants from *Petunia axillaris* N and *Petunia inflata* S6 were used as input to the MEGA6 phylogenetics analysis program (Tamura et al., 2013). Sequences were aligned using ClustalW and the alignment used as input to the MEGA6 maximum likelihood phylogenetics analysis program, using the bootstrap method with a value of 1000 iterations.

REFERENCES

Altschul, S.F., Gish, W., Miller., W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403-410

Ando, T., Tsukamoto, T., Akiba, N., Kokubin, H., Watanabe, H., Ueda, Y. and Marchesi, E. (1998). Differentiation in the Degree of Self-incompatibility in *Petunia axillaris* (Solanaceae) Occurring in Uruguay. Acta. Phytotax. Geobot. 49, 37-47.

Ascher, P.D. (1984). Self-incompatibility. In K.C. Sink, Ed., *Petunia: Monographs on theoretical and applied genetics Vol. 9.* Berlin, Springer-Verlag 92-109

Brewbaker, J.L., Natarajan, A.T. (1960). Centric fragments and pollen-part mutation of incompatibility alleles in Petunia. Genetics 45, 699-704

C**antarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A.S., and Yandell, M.** (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research 18, 188-196.

C**arrier, G., Santoni, S., Rodier-Goud, M., Canaguier, A., de Kochko, A., Dubreuil-Tranchant, C., This, P., Boursiquot, J.-M., and le Cunff, L.** (2011). An efficient and rapid protocol for plant nuclear DNA preparation suitable for next generation sequencing methods. American Journal of Botany e13-315.

Clark, K.R., Okuley, J., Collins, P.D. and Sims, T.L. (1990). Sequence variability and developmental expression of S-alleles in self-incompatible and pseudo-self-compatible *Petunia.* The Plant Cell 2, 815-826.

Darwin, C. The effects of cross and self fertilisation in the vegetable kingdom. 2nd ed. New York (NY), USA: D. Appleton and Company; 1898

De Nettancourt, D. (1977). *Incompatibility in angiosperms*. New York USA: Springer-Verlag.

Entani, T., Iwano, M., Shiba, H., Takayama, S., Fukui, K. and Isogai, A. (1999). Centromeric localization of an S-

RNase gene in *Petunia hybrida* Vilm. Theoretical and Applied Genetics 99: 391-397

Fischer, R.L. and Goldberg, R.B. (1982). Structure and flanking regions of soybean seed protein genes. Cell 29, 651-660.

Golz, J.F., Su, V., Clarke, A.E., Newbigin, E. (1999). A molecular description of mutations affecting the pollen component of the *Nicotiana alata* S locus. Genetics 152, 1123-1135.

Hua, Z. and Kao, T.-H. (2006). Identification and characterization of components of a putative *Petunia* S-locus F-box-containing E3 ligase complex involved in S-RNase-based self-incompatibility. The Plant Cell 18, 2531-2553.

Igic, B. and Kohn, J.R. (2001). Evolutionary relationships among self-incompatibility S-RNases. Proceedings of the National Academy of Sciences USA 98, 13167-13171

Ioerger, T.R., Gohlke, J.R., Xu, B., Kao, T.-H. (1991). Primary structural features of the self-incompatibility protein in Solanaceae. Sexual Plant Reproduction 4, 81-87.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R. and Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236-1240.

Kubo, K., Entani, T., Takara, A., Wang, N., Fields, A.M., Hua, Z., Toyoda, M., Kawashima, S., Ando, T., Isogai, A., Kao, T.-H. and Takayama, S. (2010). Collaborative Non-Self Recognition System in S-RNase-Based Self-Incompatibility. Science 330, 796-799.

Kubo, K., Paape, T., Hatakeyama, M., Entani, T., Takara, A., Kajihara, K., Tsukahara, M., Shimizu-Inatsugi, R., Shimizu, K.K. and Takayama, S. (2015). Gene duplication and genetic exchange drive the evolution of S-RNase-based self-incompatibility in *Petunia*. Nature Plants 1. doi:10.1038/nplants.2014.5

Lee, H.S., Huang, S., Kao, T.-H. (1994). S proteins control rejection of incompatible pollen in *Petunia inflata* Nature 367, 560-563

Li, R., Zhu, H,, Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Chan, G., Kristiansen, K., Li, S., Yang, H., Wang, J. and Wang, J. (2010). De novo assembly of human genomes with massively parallel short read sequencing. Genome Research 20, 265-272.

Li, S., Sun, P., Stephen-Williams, J. and Kao, T.-H. (2014). Identification of the self-incompatibility locus F-box protein-containing complex in *Petunia inflata.* Plant Reproduction 27, 31-45.

Linskens, H.F. (1975). Incompatibility in *Petunia*. Proceedings of Royal Society London Biological Sciences 188, 299-311.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, X., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, S., Yu, C., Wang, B., Lu, Y., Han, C., Shi, Y., Liu, S., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S,-M,, Peng, S., Xiaoqian, Z., Liu, G., Lia, X., Li, Y., Yang, H., Wang, J., Lam, T.-W. and Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short read *de novo* assmbler. Giga Science 1, 18-23.

Mather, K. (1943). Specific differences in *Petunia* I. Journal of Genetics 45, 215-235

Matton, D.P., Maes, O., Laublin, G., Xike, Q., Bertrand, C., Morse, D. and Cappadocia, M. (1997). Hypervariable domains of self-incompatibility RNases mediate allele-specific pollen recognition. The Plant Cell 9, 1757-1766.

Matton, D.P., Luu, D.T., Xike, Q,, Laublin, G., O'Brien, M., Maes, O., Morse, D. and Cappadocia, M. (1999). Production of an S-RNase with dual specificity suggests a novel hypothesis for the generation of new S alleles. The Plant Cell 11, 2087-2097.

McClure, B., Cruz-García, F. and Romero, C. (2011). Compatibility and incompatibility in S-RNase-based systems. Annals of Botany 108, 647-658

Meng, X., Sun, P. and Kao, T.-H. (2011). S-RNase-based self-incompatibility in *Petunia inflata*. Annals of Botany 108, 637-646.

Qiao, H., Wang, F., Zhao, L., Zhou, J.L., Lai, Z., Zhang, Y.S., Robbins, T.P. and Xue, Y. (2004). The F-box protein AhSLF-S_2 controls the pollen function of S-RNase-based self-incompatibility. Plant Cell 16, 2307-2322.

Shivanna, K.R. and Rangaswamy, N.S. (1969). Overcoming self-incompatibility in *Petunia axillaris* I. Delayed pollination, pollination with stored pollen and bud pollination. Phytomorphology 19, 372-380.

Sijacic P., Wang, X., Skirpan, A.L., Wang, Y., Dowd P.E., McCubbin, A.G., Huang, S., Kao, T.-H. (2004). Identification of the pollen determinant of S-RNase-mediated self-incompatibility. Nature 429, 302-305.

Sims, T.L. and Ordanic, M. (2001). Identification of a S-ribonuclease binding protein in *Petunia hybrida*. Plant Molecular Biology 47, 771-783.

Sims, T.L., Patel, A. and Shrestha, P. (2010). Protein interactions and subcellular localization in S-RNase-based self-incompatibility. Biochemical Society Transactions 38, 622-626.

Sims, T.L. and Robbins, T.P. (2009). Gametophytic Self-Incompatibility in *Petunia*. in Petunia: Evolutionary, Developmental and Physiological Genetics. Gerats T and Strommer J (eds.) ISBN 978-0-387-84795-5, Springer Life Sciences pp.85-106

Sims, T.L. (2012). Protein Interactions in S-RNase-based Gametophytic Self-Incompatibility. In Jianfeng Cai ed. Protein Interactions. In Tech Web www.intechopen.com

Tamara, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). MEGA6 , Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30 , 2725-2729.

Tanksley, S.D., Loaiza-Figueroa, F. (1985). Gametophytic self-incompatibility is controlled by a single major locus on chromosome 1 in *Lycopersicon peruvianum.* Proceedings of the National Academy of Sciences USA 82 , 5093-5096.

ten Hoopen, R., Harbord, R.M. Maes, T., Nanninga, N. and Robbins, T.P. (1998). The self-incompatibility (S) locus in *Petunia hybrida* is located on chromosome III in a region syntenic for the Solanaceae. Plant Journal 16, 729-734

Tsukamoto, T., Ando, T., Watanabe, H., Marchesi, E. and Kao, T.-H. (2005) Duplication of the S-locus F-box gene is associated with breakdown of pollen function in an S-haplotype identified in a natural population of self-incompatible *Petunia axillaris.* Plant Molecular Biology 57,141–153

Wang, Y., Wang, X., McCubbin, A.G. and Kao, T.-H. (2003) Genetic mapping and molecular characterization of the self-incompatibility (S) locus in *Petunia inflata* Plant Molecular Biology 53, 565-580.

Wang, Y., Tsukamoto, T., Yi, K,-W., Wang, X., Huang, A., McCubbin, A.G. and Kao, T.-H. (2004) Chromosome walking in the *Petunia inflata* self-incompatibility (S-) locus and gene identification in an 881-kb contig containing S2-RNase. Plant Molecular Biology 54, 727-742.

Wheeler, D. and Newbigin, E. (2007) Expression of 10 S-Class SLF-*like* Genes in *Nicotiana alata* Pollen and Its Implications for Understanding the Polen Factor of the S Locus. Genetics 177, 2171-2180

Williams, J.S., Natale, C.A., Wang, N., Li, S., Brubaker T.R., Sun, P. and Kao, T.-H. (2014a) Four previously identified *Petunia inflata* S-locus F-box genes are involved in pollen specificity in self-incompatibility. Molecular Plant 7, 567-569.

Williams, J.S., Der, J.P., dePamphilis, C.W. and Kao, T.-H. (2014b). Transcriptome analysis reveals the same 17 S-Locus F-box genes in two haplotypes of the self-incompatibility locus of *Petunia inflata*. Plant Cell 26, 2873-2888.

Zhao, L., Huang, J., Zhao, Z., Li, Q., Sims, T.L. and Xue, Y. (2010) The Skp1-like Protein SSK1 is Required for Cross-Pollen Compatibility in S-RNase-Based Self-Incompatibility. Plant Journal 65, 52-63.