Characterisation of the Self-Incompatibility Related F-box Proteins of *Nicotiana alata*

Submitted by

Poh Ling Koh

in fulfillment of the requirement for the Degree of Doctor of Philosophy

School of BioSciences

University of Melbourne

8th November 2015

Declaration

This thesis comprises only my own original work towards the PhD degree except for the contributions mentioned in the Acknowledgements. I have acknowledged all materials used and the thesis is less than 100, 000 words in length exclusive of tables, figures and references.

Poh Ling Koh

Student number: 244308

Abstract

Self-incompatibility is a genetically encoded barrier to self-fertilisation found in some plant species. In solanaceous plants such as *Nicotiana alata* self-incompatibility is controlled by a highly allelic single locus known as the *S* locus (Newbigin et al., 1993). Fertilisation is prevented when the *S* allele expressed by the haploid pollen grain matches either of the S alleles expressed by the diploid female reproductive tissue or style. The only known products of the *N. alata* S locus are a style-specific extracellular ribonuclease called the S-RNase and a family of 10 pollen-expressed F-box protein genes called the *DD* genes (Wheeler and Newbigin 2007). S-RNases determine the allelic identity of the style and in *Petunia inflata*, another self-incompatible member of the Solanaceae, a family of *DD*-related genes called S locus F-box or SLF genes regulates S allele identity in pollen (Kubo et al., 2010). Although it now appears that pollen identity in solanaceous plants is determined through the action of several genes, as described by Kubo et al's (2010) collaborative non-self recognition model, when this thesis began it was thought that only one gene, encoding a determinant factor known as pollen S, was involved (McCubbin et al., 1997; Newbign et al., 2008). As the hypothesis was that *N. alata* pollen S was encoded by a *DD* gene or related sequence, this thesis began with the aim of discovering whether there were any more *DD* genes expressed in *N. alata* pollen and which of the 10 (or more) *DD* genes encoded the pollen S determinant.

Chapter 1 reviews the literature relevant to this study and sets out the thesis aims. Chapter 2 describes the use of next generation sequencing of *N. alata* pollen grain RNA to identify *DDs* and other RNase-based SI related transcripts reported by other studies. As the *N. alata* genome has not been sequenced an RNA-Seq bioinformatics pipeline was developed for *de novo* transcriptome assembly. The assembled pollen grain transcriptome was validated using bioinformatic and molecular approaches and searched using the sequences implicated by other studies in the self-incompatibility response of solanaceous plants (Hua and Kao, 2006; Zhao et al., 2010). F-box proteins are a component of the SCF (Skp1-Cullin-F-box protein) complex that takes ubiquitin from the E2 ubiquitin-conjugating enzyme and transfers it to a substrate protein or target. F-box proteins such as DD/SLF bind to the substrate protein and give the SCF complex its unique specificity (Petroski and Deshaies, 2005; Vierstra 2009). Two novel DD/SLF transcripts were identified in the transcriptome, as well as transcripts of other genes suggested to encode components of the SLF-containing SCF complex. Some of the work described in this chapter contributed to the paper by Lampugnani et al. (2013).

Because SLFs are F-box proteins their main function in the self-incompatibility response is presumably to bind to and ubiquitylate the S-RNase, leading to degradation by the 26S proteasome

pathway (Sijacic et al., 2004). Chapter 3 reports on experiments designed to identify the *N. alata* pollen S determinant using recombinant DD/SLF and S-RNase proteins produced in *E. coli*, and an in vitro binding assay similar to that described by Hua and Kao (2006) for *Petunia* SLFs. A major problem for the binding assay was obtaining recombinant proteins in a largely soluble and intact or full-length form – the proteins were usually insoluble and had undergone degradation – but even so the in vitro assay failed to consistently detect binding between SLF and S-RNase. An alternative assay was developed that used antibodies specific for one of the protein partners to immunoprecipitate the other partner (the Co-IP assay). As reproducible binding of SLF and S-RNase could be detected in a Co-IP assay that used recombinant SLF proteins in *E. coli* extracts and native S-RNases from *N. alata* stylar extracts, further experiments aimed at production of enriched, soluble SLF/DD proteins were planned.

Chapter 4 describes work focussed on refolding the insoluble DD and SLF proteins into a soluble and enriched form for use in the Co-IP assay and biophysical characterization. Protein folding is not a routine procedure but many studies have successfully rescued misfolded proteins by this approach. This chapter reports the production of soluble, enriched SLF/DD proteins from which the N-terminal F-box motif had been deleted. These truncated SLF/DD proteins could still interact specifically with *N. alata* S-RNases, suggesting that F-box motif is not necessary for this interaction. Biophysical characterization of SLF/DD proteins by circular dichroism analysis suggested they had a large component of β-sheet, which is consistent with the six-bladed beta-propeller structure predicted from theoretical modeling. However the high content of β-sheet was also consistent with the presence of amyloids, misfolded globular proteins composed of intermolecular arrays of parallel β-sheets (Nelson et al., 2005). As analytical ultracentrifugation of the resolubilised SLF/DD proteins detected polydispersed oligomers, it seems likely that refolding had resulted in amyloid formation that were nonetheless still functional, based on their ability to bind to S-RNases.

Chapter 5, the final chapter, summarises the work in the rest of the thesis and suggests some productive areas for further research.

Acknowledgements

I first would like to thank my supervisors, Dr Ed Newbigin and Dr Terry Mulhern, for their guidance and support throughout the course of my study. They have always been there for me and have imparted to me knowledge and skills without reservation. Without their support, the completion of this thesis would not have been possible and it has been an invaluable lifetime experience working with them that will certainly benefit me in my future scientific career.

I would also like to thank School of Botany (now part of the School of BioSciences) and University of Melbourne for awarding me the international fee remission scholarships and science faculty postgraduate scholarship. Without this vital financial support, I would not have been able to finish my project.

I aknowledge the generous contributions and technical assistance to this thesis provided by Andrew Cassin (Australian Centre for Plant Functional Genomics, University of Melbourne), Alexander Ray and Dr Mok Yee Foong (Bio21 Institute, University of Melbourne) and Dr David Wheeler (School of Botany, University of Melbourne). The nature of their contributions is described in detail elsewhere.

I would like to thank the people who shared with me the good and difficult times of my PhD. During the course of our studies, we experienced the ups and downs together, offering each other mutual support and also having lots of fun too. So many of you were there for me but I'd especially like to mention Dr Isabel Moller, Dr Shane Emanuelle, Dr Tony Chin, Dr Arianna Hernandez-Sanchez, Dr Joshua Koh, Dr Lili Song, Dr Aimin Wu, Dr Wei Zeng, Dr Edwin Lampugnani and Erinna Pak. I appreciate and thank all the others who have helped me along the way, especially Dr Monika Doblin and Andrew Casin.

Lastly, I'd like to thank my family for the continuous support throughout the years. Without their support, I will not have the chance to study aboard for so many years with no big worries. They are always there for me no matter what happens and without them I would not have completed this journey.

Table of Contents Page

1.1. Overview of self-incompatibility in flowering plants

The archetypal angiosperm flower consists of sepals, petals, stamens and carpels, with sepals and petals in the two outer whorls of sterile organs and stamens and carpels in the two inner whorls of reproductive organs. So a typical flower is a hermaphrodite containing both male (stamen) and female (carpel) reproductive organs. As reproduction in flowering plants begins with the transfer of pollen (male gametophyte) from stamen to the start of the female reproductive tract (stigma), the overall architecture of such a hermaphrodite flower, with its close proximity of male and female organs, seems ideally suited to the production of offspring (seeds) following autogamy or self-pollination. That roughly half of all angiosperm species do not produce seeds by this method (e.g., see Igic and Kohn, 2006) reflects the various adaptations flowering plants have evolved that favour the production of outbred over inbred progeny. These adaptations include mechanisms that favour pollen distribution and allogamy (cross-pollination), as well as those that disfavour seed production after autogamy. Among the latter types of adaptation are the self-incompatibility systems, the subject of this chapter.

Self-incompatibility (SI) has been defined as "the inability of a fertile hermaphrodite seed plant to produce zygotes after self-pollination" (Lundqvist, 1965) and as such refers specifically to those systems where "self" pollen tubes are prevented from reaching the ovules through the action of a pre-zygotic barrier within the female reproductive tract (stigma and style). Although SI systems generally are the topic of this review, Lundqvist's definition potentially excludes some of these systems, such as the so called "late acting" or ovarian SI systems, where the growth of self and cross pollen tubes within the stigma and style is identical up to the point of ovule penetration. Rejection of the incompatible gamete occurs at a later stage – from micropylar entry to failure to fuse with the egg cell, or post-zygotically with failure of the zygote to divide or an endosperm to develop (Seavey and Bawa, 1986; Ford and Wilkinson, 2012). Although gamete rejection in some late acting systems may be prezygotic (e.g., in *Thryptomene calycina* (Myrtaceae), where the growth of self pollen tubes is apparently arrested near the micropyle; (Beardsell et al., 1993)), as a class the late-acting systems are poorly described and will not be discussed further. Readers are directed to Sage et al., (1994) for one of the few comprehensive reviews available of the ovarian systems, and to Allen and Hiscock (2008) for a discussion of the distribution of these systems in flowering plants.

In SI systems where a clearly defined pre-zygotic barrier exists within the female reproductive tract further classification is possible based on the morphology of the flower. Heteromorphic SI systems are characterised by physical differences in floral structures (heterostyly) that divide the population up into different mating types or morphs. Cases of heterostyly have been reported in 28 families

with a familiar example being that of *Primula* (Primulaceae), where variation in the lengths of the stamen and style identifies the two morphs known as pin and thrum: the pin morph has short anthers and a long style and thrum morph has the reverse (Barrett, 2002). Although this variation physically prevents self-pollination there is also an incompatibility mechanism that blocks cross-fertilisation between plants of the same morph (pin and pin or thrum and thrum). In *Primula*, floral morphology and intra-morph incompatibility are determined by two alleles at a genetic locus named the *S* locus, a dominant *S* allele and recessive *s* allele (McCubbin, 2008). Short-styled thrum plants are *Ss* heterozygotes and long-styled pin plants are *ss* homozygotes. The *Primula S* locus has been described as a supergene (Ernst, 1955) as the single locus controls both morphological traits such as gynoecium development (style height), pollen size and stamen filament length, as well as the associated SI barrier. Currently work is ongoing to isolate genes at both the *Primula S* locus (e.g., see Li et al., 2011) and from the *S* locus of heterostylous *Fagopyrum* (common buckwheat; Polygonaceae), where a putative transcription factor related to the *Arabidopsis thaliana* protein EARLY FLOWERING 3 appears to be involved in control of the short-styled thrum phenotype (Yasui et al., 2012).

The other type of SI system where a clearly defined pre-zygotic barrier exists is the homomorphic SI system. The hermaphrodite flowers of different mating types in a homomorphic system are physically identical to one another and there is no physical barrier to self-pollination. Like the heteromorphic systems, rejection of self pollen is genetically controlled by a single *S* locus (most commonly), with a distinguishing feature of homomorphic *S* loci being the presence of numerous *S* alleles (Newbigin et al., 1993). Because of this the different *S* alleles in a homomorphic system are numbered (S_1, S_2, S_3, etc). Pollen grains are rejected by styles when both express the same *S* allele and accepted when the *S* alleles present are different. Homomorphic system can be further divided based on the way the *S* allele identity of pollen grains is determined. In sporophytic SI (SSI) systems, the *S* alleles expressed by the haploid pollen grain are those of the diploid parent or sporophyte, whereas in gametophytic SI (GSI) systems, the pollen grain expresses the single *S* allele present in its own haploid genome. Figures 1.1 and 1.2 illustrate the genetics of the two SI systems. Although pollen grains in a sporophytic system can carry the products of two *S* alleles, some *S* alleles in this system are dominant over others so that only a single *S* allele is expressed. Dominance and recessive *S* alleles can also be expressed by the stigma with this relationship being independent of the dominance relationships in pollen. A result of *S* allele dominance is the very complex patterns of compatibility and incompatibility seen between plants in a population (Hiscock and Tabah, 2003; Mable et al., 2003). By contrast the two *S* alleles expressed by the style in a GSI plant are co-dominant. Of the two systems, GSI is taxonomically the more widespread having been reported in

the Solanaceae, Rosaceae, Plantaginaceae, Poaceae, Fabaceae, Onagraceae, Campanulaceae and Papaveraceae among others (Allen and Hiscock 2008). SSI is known in six families, the Asteraceae, Betulaceae, Brassicaceae, Caryophyllaceae, Convolvulaceae and Polemoniaceae. Additionally there are SI systems where instead of a single locus, pollen rejection is controlled by multiple loci. An example of this is found in the Poaceae (grasses) where a two-locus (*S* and *Z*) GSI system prevents fertilisation when identical *S* and *Z* alleles are expressed by pollen and stigma (Klaas et al., 2011).

By regulating the availability of compatible mates and reducing the negative consequences of inbreeding, SI systems have profound effects on the fitness, structure and distribution of plant populations (e.g., Baker's law; Baker, 1955; Barrett, 2002). Indeed, the effects of an SI system are so profound that some have suggested this adaptation was an essential factor in the emergence and rapid diversification of angiosperms during the Cretaceous period (Whitehouse, 1950; Bell, 1995). However, instead of the single ancestral SI system they proposed, molecular studies, where they have been done, have shown that these systems are mostly unrelated to each other, leading to the conclusion that SI has arisen many times during the evolution of flowering plants and that most systems have converged on control by a single locus (Read et al., 1995). A notable exception to this are the three ribonuclease-based systems found in the Rosaceae, Solanaceae and Plantaginaceae, which evidence suggests are descended from a common ancestral system present in the majority of dicots (Igic and Kohn, 2001; Steinbachs and Holsinger, 2002). Another convergent feature of most *S* loci is the presence of two separate genes for the factors regulating the SI phenotypes of pollen and style. Lewis (1949) first showed the existence of separate pollen and stylar factor genes through mutational experiments on *Oenothera organensis* (Onagraceae), a species with a single locus GSI system. Lewis (1949) concluded that in *Oenothera* the *S* locus must be a composite of two mutational units "held together in such a way at meiosis that no crossing over takes place between them"; later the *S* locus of *Prunus avium* (Rosaceae) was found to have the same two-component structure (Lewis and Crowe, 1954). Throughout this thesis the two mutational units that comprise the *S* locus will be referred to as the *style S* and *pollen S* genes.

As single genes of large effect and also because their role in self/non-self recognition could be compared to the vertebrate immune system (e.g., see Heslop-Harrison, 1975), the cell and molecular biology of many SI systems have been intensively studied over the past few decades. These studies will not be described here in detail as comprehensive reviews have appeared in the literature at regular intervals (Tantikanjana et al., 2010; Wheeler et al., 2010; McClure et al., 2011; Iwano and Takayama, 2012) and only short summaries of the essential features of each system will be provided for comparative purposes.

The only SSI system for which the *style S* and *pollen S* genes have been characterised is the one found in the Brassicaceae, although some work has been done on the SSI systems of *Senecio squalidus* (Asteraceae; Allen et al., 2011) and *Ipomoea trifida* (Convolvulaceae; Rahman et al., 2007). In the *Brassica* SSI system *style S* encodes a plasma membrane-associated receptor kinase expressed by stigmatic papilla cells called the S LOCUS RECEPTOR KINASE (SRK: Stein et al., 1991) and *pollen S* is a small cysteine rich protein expressed by anthers that is known as either S LOCUS PROTEIN 11 (SP11) or S LOCUS CYS-RICH PROTEIN (SCR: Schopfer et al., 1999; Suzuki et al., 1999). Signaling in this system is initiated through an allele-specific interaction between the SP11/SCR male determinant and SRK female determinant following contact between the pollen grain and stigmatic papillae cell. Binding of SP11/SCR to its cognate SRK is *S* allele specific and leads to autophosphorylation of SRK as well as the phosphorylation of intracellular targets (e.g., the Armadillo-repeat-containing protein, ARC1) to initiate a signalling cascade that suppresses pollen tube germination by disrupting the delivery of stigmatic factors that are required for pollen hydration (Figure 1.3; Takayama et al., 2001; Stone et al., 2003; Samuel et al., 2009). With its involvement of a small protein ligand and a membrane-bound protein kinase, the *Brassica* SSI system is similar to many other ligand-receptor systems in plants (see Matsubayashi and Sakagami, 2006).

The *S* locus *style S* and *pollen S* genes associated with GSI have been identified for the Papaveraceae, Rosaceae, Solanaceae and Plantaginaceae. Work in the Papaveraceae, or poppy, family has almost entirely been done with *Papaver rhoeas*, the corn poppy (Lawrence, 1975). *Papaver rhoeas* style S, called PrsS for *P. rhoeas* stigma S determinant, is a small (~15 kDa) protein secreted by stigmatic papilla cells (Foote et al., 1994) and pollen S, called PrpS for *P. rhoeas* pollen S determinant, is a ~20 kDa transmembrane protein expressed by pollen grains and pollen tubes (Figure 1.4; Wheeler et al., 2009). SI in *P. rhoeas* is initiated by the interaction of PrpS and PrsS from the same *S* allele. This interaction leads to an increase in levels of Ca^{2+} in the pollen tube cytoplasm, deploymerisation of the actin cytoskeleton, and ultimately in death of the pollen tube by an apoptotic mechanism (Wheeler et al., 2010). PrsS belongs to a large, plant-specific protein family named SPH (S-Protein Homologue; Ride et al., 1999) but there are no obvious homologues of PrpS, which lacks kinase or other known catalytic domains but is suggested to be a novel class of receptor proteins (Wheeler et al., 2010).

Although each has its own unique features, self/non-self discrimination in the *Brassica* SSI and *Papaver* GSI systems is mediated by plasma membrane receptors perceiving secreted protein ligands, a fairly typical means of intercellular communication in both plants and animals (e.g., De Smet et al., 2009). By contrast GSI in the Solanaceae, Plantaginaceae and Rosaceae represents a

novel cell-signalling paradigm as perception does not occur at the plasma membrane and the style S determinant in each family is an extracellular ribonuclease known as the S-RNase (Newbigin et al., 1993). For this reason, these SI systems are known as the RNase-based systems. The next sections of the thesis will focus on the general appearance of pollen tube rejection in these systems, characterisation of the *style S* and *pollen S* determinants and the biochemical models they suggest, and discuss recent evidence of the diverse functions that S-RNase homologs play in the biology of eukaryotes and human diseases.

1.2. Pollen rejection in the RNase-based SI system

This section provides an overview on the biology of pollen tube rejection in *Nicotiana alata* and *Pyrus pyrifolia* or Japanese pear. A study in *N. alata* showed that incompatible (rejected) pollen tubes remain viable; in contrast studies in *Pyrus pyrifolia* showed that incompatible pollen tubes undergo programmed cell death (PCD) instead. Interestingly, depolymerisation of F-actin is observed in incompatible (rejected) pollen tubes from both species, indicating some similarities despite the different final outcomes (Lush and Clarke, 1997; Liu et al., 2007; Wang et al., 2009; Wang et al., 2010).

Nicotiana alata, a species from the Solanaceae or nightshade family, is self-incompatible because of an RNase-based SI system and has a long history of use in studies of RNase-based SI (e.g., Bredemeijer and Blass, 1981; Anderson et al., 1986). Characterisation of *N. alata* pollen tubes growth within compatible and incompatibly pollinated styles showed that within the first seven hours post-pollination, the growth rate of compatible and incompatible pollen tubes is similar but growth rate for the latter slows dramatically after seven hours showing that it takes time for full rejection to occur (Lush and Clarke 1997). The growth rate of incompatible pollen tubes was reduced to almost six fold in comparision to compatible pollen tubes and incompatible pollen tubes seldom grow beyond the top half section of the style. However a grafting experiment showed that incompatible pollen tubes remain viable. Style grafting involves taking a pollinated style and transfering the upper section of the cut style (scion) onto a lower section of another cut style (stock). The growth rate of incompatible pollen tubes growing in an incompatible style scion increases once the pollen tube crosses over into a compatible style stock section (Lush and Clarke, 1997). Although it is clear that incompatible pollen tubes remain viable and given sufficient time should be able to fertilise ovules, in practice this rarely happens because the flower would have senesced before fertilisation is possible (Lush and Clarke, 1997).

In contrast, *in vitro* studies performed in *P. pyrifolia* suggest that incompatible pollen tubes in this species undergo PCD. In *P. pyrifolia* it was observed S-RNase induces F-actin depolymerisation and

other changes in pollen tubes such as nuclear DNA degradation, alteration in mitochondrial structure and disruption of reactive oxidative species (ROS) balance at pollen tube tip (Liu et al., 2007; Wang et al., 2009; Wang et al., 2010). As these are PCD markers observed in the well-characterised GSI system present in *Pavaper rhoeas* (Franklin-Tong and Gourlay, 2008), it appears that incompatible *P. pyrifolia* pollen tubes also undergo cell death.

P. pyrifolia pollen tubes germinated *in vitro* and treated with self or non-self S-RNases were labelled with phallodin to visualise F-actin integrity (Liu et al., 2007). Depolymerisation of F-actin into punctuate foci was observed only in pollen tubes treated with self S-RNase. A separate study detected DNA degradation, the release of cytochrome c into the cytosol and changes in mitochondrial structure in *in vitro* germinated pollen tubes treated with self S-RNase but not pollen tubes treated with non-self S-RNase (Wang et al., 2009). Pollen viability was also higher in *in vitro* germinated pollen tubes treated with self S-RNase. Reactive oxidative species (ROS) are required for pollen tube tip growth but addition of self S-RNase but not non-self S-RNase to *in vitro* germinated pollen disrupted the ROS balance (Wang et al., 2010). When ROS balance was disrupted, nuclear DNA degradation was also observed *in vitro* and the degree of DNA degradation in *in vivo* incompatibly pollinated style was stronger than in a compatibly pollinated style (Wang et al., 2010). The disruption of ROS balance at pollen tube tips, F-actin depolymerisation, DNA degradation and alteration of mitochondrial structure by self S-RNase are characteristics similar to those seen in the GSI system present in *P. rhoeas*, where the final outcome of incompatible pollen tubes is PCD (Wang et al., 2010).

The different outcome observed *in vivo* in *N. alata* and *in vitro* in *P. pyrifolia* suggest differences in the RNase-based SI mechanism present in the Rosaceae (*P. pyrifolia*) and Solanaceae (*N. alata*) and that the Rosaceae SI system is more similar to the GSI mechanism present in the Papaveraceae. But transcriptome profiling studies of *semi in vivo* grown pollen tubes of *Arabidopsis thaliana* suggest another explanation (Qin et al., 2009). In a *semi in vivo* system, pollen tubes grow through a cut pollinated style explant and into a culture medium where the actively growing tips can be harvested. Qin et al., (2009) showed that *Arabidopsis* pollen tubes that came into contact with style tissue in a *semi in vivo* system contained numerous transcripts that were not present in pollen tubes that only grew in an *in vitro* system, suggesting that pollen tubes can sense and respond to the external environment. Of the transcripts unique to pollen tubes in the *semi in vivo* system, many were annotated as being involved in cell signalling, transcription and pollen tube elongation (Qin et al., 2009). In addition, tubes growing *in vivo* respond to guidance cues secreted by style and ovule that accurately direct them in the direction of an unfertilised ovule (Kessler and Grossniklaus, 2011;

Takeuchi and Higashiyama, 2011). This suggests that the responses seen in *in vitro* grown pollen tubes are likely to be different to those seen in *in vivo* grown pollen tubes due to the physical contact with the style tissue. This form of cell-to-cell interaction could be critical for the pollen tube's overall health, growth behaviour and perhaps may be associated with the different outcomes observed in *N. alata* and *P. pyrifolia.*

Recent studies using *N. alata* as the model plant also observed F-actin depolymerisation, in addition, the breakdown of the vacuolar compartment during late stage pollen tubes rejection. The breakdown of the vacuolar compartment was first observed by Goldraji et al., (2006) who proposed the compartmentalisation model proposed to explain the mechanism of late stage pollen rejection in RNase-based SI.

The uptake of S-RNase by compatible and incompatible pollen tubes was first shown by Luu et al., (2000) using a specific antibody raised against the hypervariable region of S_{11}-RNase from *Solanum chacoense*. The location of S-RNase suggests it is present in the cytoplasm of the pollen tubes rather than in the vacuole compartment. Goldraji and colleagues (2006) also observed uptake of S-RNases in compatible and incompatibe pollen tubes but in contrast to Luu et al., (2000) found S-RNase trapped within vacuoles. Goldraji et al., (2006) made other key observations in the *N. alata* pollen tubes and proposed the compartmentalisation model to explain what happen during late stage pollen tube rejection. Firstly, they identified two style modifier proteins, 120 kDa protein and HT-B, which are located around the vacuole containing S-RNase but the absence of either protein did not affect uptake of S-RNase into the pollen tube. As already shown by previous gene silencing studies, both modifier proteins are required for *S* specific pollen rejection suggesting they are important for pollen rejection. Secondly, it was observed during late stage pollen tube rejection, there was a reduced level of 120 kDa protein and HT-B remains intact. In contrast, HT-B was found degraded in the compatible pollen tubes. Lastly, the extent of vacuole compartment degradation is greater in incompatible pollen tubes.

The compartmentalisation theory states that S-RNase is first take up into a compartment in the pollen tubes together with HT-B and 120 kDa protein but HT-B and 120 kDa protein later transition to the vacuole surface. The stability of HT-B determines if S-RNase is released into the cytoplasm of the pollen tubes. A small amount of S-RNase would escape compartmentalisation making its way through the retrograde transport system to end up in the cytoplasm where SLF (S locus F-box; pollen S) is and interaction occurs. In an incompatible pollination, the complex form between self S-RNase and self SLF inhibits an unknown protein which in turn stabilises HT-B. Stabilised HT-B triggers the degradation of the vacuole and S-RNase is released into the cytoplasm for complete pollen rejection.

The interaction between self S-RNase and self SLF is a self-reinforcing mechanism causing more S-RNase to be released for complete rejection of the pollen tubes. In compatible pollination, HT-B is degraded as the complex formed between non-self SLF and self S-RNase could not inhibit the unknown protein which in turn caused the degradation of HT-B. The compartment remains intact and S-RNase remains trapped within the vacuole allowing pollen tube growth to continue. Although the precise role of 120 kDa protein remains unknown, the breakdown of vacuole which occurs much later after pollination coincides with the late stage absence of 120 kDa and presence of HT-B suggesting they are involved in the final stage release of S-RNase into the cytoplasm to completely inhibit the growth of incompatible pollen tubes (Goldraij et al., 2006).

Interestingly, a recent report also using *in vivo N. alata* pollen tubes showed pollen rejection is not only associated with the integrity of vacuoles but also with F-actin. Progressive dissociation of F-actin was observed in incompatible pollen tubes over a period of eight days, but the level of F-actin dissociation in compatible pollen tubes was observed to be consistent until 2.5 days post-pollination and pollen tubes reached the ovary within 72 hours, suggesting that fertilisation occurs three days after pollination and that F-actin remains intact close to fertilisation of ovule. This suggests F-actin depolymerisation plays a role in pollen tubes rejection. Interestingly, the level of F-actin disorganisation is significantly increased in incompatible pollen tubes between day one to day three post-pollination and thereafter only a very modest increase was detected up to eight days post-pollination. In contrast, the level of vacuole breakdown increases significantly three days after pollination. This suggest that F-actin disorganisation occurs prior to vacuole breakdown and could act as an upstream signalling event which leads to the release of S-RNase from vacuoles during late stage pollen tubes rejection (Roldan et al., 2012).

In summary, using *N. alata* a model organism, Lush and Clarke (1997) showed that *semi in vivo* incompatible pollen tubes are viable post-pollination and that incompatibility arises as a consequence of a slowed growth rate. Goldraji et al., (2006) showed that *in vivo* pollen tube rejection likely required the release of large amounts of S-RNase from the vacuole during late stages of pollination rejection. Roldan et al., (2012) also observed breakdown of vacuoles in late stage pollen tubes but suggested that F-actin depolymerisation occurred prior to vacuole breakdown. From the *N. alata* studies, pollen tubes can be all considered alive as there is no evidence to suggest they have undergone PCD. In contrast, *P. pyrifolia* studies identified molecular makers for PCD suggesting that PCD is triggered *in vivo* within hours of exposure of pollen tubes to non-self S-RNases. Despite the RNase-based SI molecules involved in both families being similar, a critical observed difference is whether the incompatible pollen tube remains viable.

1.3. Characterisation of the *S* locus genes, *S-RNase* and *S locus F-box*.

Style S and *pollen S* must possess certain characteristics to be an *S* locus gene. Firstly, *style S* must only be expressed in the style and similarly, *pollen S* is only expressed in the pollen. Secondly, there would be many different *S* alleles present in a population of plants since fertilisation is only possible when *style S* and *pollen S* are from non-matching *S* alleles. Hence, both genes from all *S* alleles would show a similar degree of sequence variation (*S* specificity). Lastly, both genes must be tightly linked to the *S* locus and inherited as a unit as recombination within the two genes would cause RNase-based SI to dysfunction (McClure, 2004). A *N. alata* style-specific protein linked to the *N. alata* S_2 allele was first identified in 1981 (Bredemeijer and Blaas, 1981) and subsequently cloned in 1986 (Anderson et al., 1986). Aligning sequences of this gene from different *S* alleles identified two hypervariable regions that were allele-specific, thus fulfilling all the expectations of an *S* locus gene (Ioerger et al., 1991). *Style S* was found to encode a protein with sequence similarity with yeast T2-RNase including the two conserved histidine residues required for RNase activity (McClure et al., 1989); the same group soon after named style S which degrades ribonucleic acid *S-RNase* (McClure et al., 1990). Orthologues of *S-RNases* are later on identified from the Plantaginaceae and Rosaceae families and found to have similar sequence features as *S-RNase* from the Solanaceae family (Sassa et al., 1992; Sassa et al., 1993; Sassa et al., 1996; Tao et al., 1997; Ushijima et al., 1998; Wiersma et al., 2001; Xue et al., 1996).

Since the discovery of S-RNase, many studies have provided functional evidence that S-RNase and its RNase activity are important for pollen tube rejection but only representative studies will be summarised here. Loss and gain-of-function studies by Lee et al., (1994) showed that S-RNase is necessary and sufficient for style function. In a gain-of-function study, transgenic S_1S_2 *Petunia inflata* plants gained the ability to reject S_3 pollen when an *S_3-RNase* transgene was present, suggesting that S-RNase is necessary for style function in RNase-based SI. Similarly, a loss-of-function study in which an antisense *S_3-RNase* construct was expressed in a S_2S_3 plant resulted in an inability to reject S_3 pollen. The discovery of naturally occurring self-compatible accession of the otherwise self-incompatible species wild tomato (*Lycopersicon peruvianum*; Kowyama et al., 1994) and Japanese pear (*Pyrus serotina*; Sassa et al., 1997) found that high levels of expression of a functional S-RNase were necessary for *style S* function. Matton et al., (1997) showed the hypervariable regions of S-RNase control *S* specificity. *S. chacoense* S_{11}-RNase and S_{13}-RNase protein sequence differ by only 10 amino acids, of which four lie within one or other of the two hypervariable regions. Using a 'domain swap' approach, in which short stretches of sequence from one S-RNase gene are replaced with the corresponding region from another gene, Matton et al., (1997) showed that transgenic $S_{12}S_{14}$ *S. chacoense* plants expressing an S_{11}-RNase with the hypervariable regions of the S_{13}-RNase, accepted

S_{11} pollen but rejected S_{13} pollen, demonstrating that the hypervariable regions can convey *S* specificity. However, not all domain swap experiments involving the transfer of hypervariable domains between two S-RNases have been successful, as most experiments have resulted in the production of correctly folded S-RNases expressed in styles that are unable to reject pollen carrying either of the S alleles used as a source of DNA (Zurek et al., 1997; Verica et al., 1998). Lastly, Huang et al., (1994) showed that one of the conserved histidine residues is required for S_3-RNase activity as S_1S_2 plant expressing mutant S_3-RNase was unable to reject S_3 pollen. S_3-RNase activity is lost by replacing the histidine residue with another structurally similar amino acid, asparagine.

Since *pollen S* must co-segregate with *S-RNase*, the two genes should be physically close to each other at the *S* locus. The first *pollen S* candidate was isolated by sequencing the genomic region surrounding the S_2*-RNase* locus of *Antirrhinum hispanicum* from the snapdragon or Plantaginaceae family. The closest gene, only 9 kb away, encoded an F-box protein specifically expressed in pollen (Lai et al., 2002). Although additional pollen-expressed *F-box* genes were found near the *Antirrhinum S* locus, only one displayed the *S* haplotype specificity expected of *pollen S*. This gene was named *S locus F-box* (*SLF*) and the other related but non-*S* haplotype specific *F-box* genes, thought not to have a role in RNase-based SI, were named *SLF-likes* (Lai et al., 2002). Three other alleles of *SLF* were soon identified but unexpectedly all *SLFs* were highly similar, with more than 97% sequence identity (Zhou et al., 2003). Progressively, *F-box* candidate genes were isolated from the *S* loci of the Solanaceae and Rosaceae families (Table 1.1). Similarly, multiple *F-box* genes were isolated and the one that showed *S* haplotype specificity was considered to be *pollen S*. This gene was named *SLF* in the Solanaceae and *SFB* in the Rosaceae, with other *F-box* genes named *SLF-likes* or *SFB-likes* although the term S locus F-Box Brothers (SFBBs) has also been used (Sassa et al., 2007). However, around 2007, studies (Table 1.2) identified multiple *S* haplotype-specific, highly conserved pollen-expressed *F-box* genes linked to the *S* locus of *Malus*, *Pyrus* and *Prunus* (Rosaceae), *Petunia*, *Solanum* and *Nicotiana* (Solanaceae), in contrast to the earlier studies (Table 1.1). All *F-box* identified from respective species represent *pollen S* as it is not possible to determine which one is *pollen S* based on linkage analysis and sequence polymorphism displayed by the *F-box*. In addition, recent linkage studies revealling recombination of *SFBBs*/*SLFs* at the *S* locus suggesting that not all *SFBBs*/*SLFs* are involved in RNase-based SI (De Franceschi et al., 2011b, Kakui et al., 2011, Li and Chetelat., 2015).

SLFs from *Petunia* and *Nicotiana* show a very low degree of sequence variation. The *SFBs* show higher sequence variation than *SLFs* but the level of variation is not comparable with the *S-RNases* (Ushijima et al., 2003; Zhou et al., 2003; Wang et al., 2003; Yamane et al., 2003a; Yamane et al., 2003b; Ikeda et al., 2004; Cheng et al., 2006 ; Wheeler and Newbigin 2007; Minamikawa et al., 2010;

Okada et al., 2011). The evolutionary force acting on the pair of tightly linked *S* genes dictates that both genes are subjected to the same evolutionary pressure and hence should have the same degree of sequence variation (Newbigin et al., 2008). In summary, as explained by Newbigin et al., (2008), pollen from a SI plant carrying a rare *S* allele is at a selective advantage over others as it would be compatible with most existing styles. Hence, instead of rare *S* alleles being lost over time, it is more likely that the frequency of these rare *S* alleles will increase and spread across the population of plants over a short period of time until its frequency is in equilibrium with the other alleles. This balancing selection action on all *S* alleles also allows the preservation of all *S* alleles over a long period of time enable positive selection to act on the *S* alleles to generate new *S* alleles by random mutation. Over time, this contributes to specific regions of high variability within the *S* alleles seen within a species and instead the *S* alleles sequences among different species are more similar. The *S-RNases* displayed the highly variable characteristic and are shown to be ancient genes but the *SLFs* are much less variable and younger making them not the best candidate of *pollen S* (Newbigin et al., 2008). A recent development in RNase-based SI study may explain why *SLF* does not necessarily co-evolve with *S-RNase* (see section 1.5). For more details on the evolution of *S* locus genes, refer to Newbigin et al., (2008); Vieira et al., (2009), De Franceschi et al., (2012) and Kubo et al., (2015).

1.4. *SLF* as the *pollen S* determinant

Polyploidy plants derived from self-incompatible diploid species are generally self-compatible because of an effect called competititve interaction (Golz et al., 2000). Although pollen is haploid, being produced through reductive meiotic division, the pollen of a tetraploid relative of a self-incompatible diploid will carry two *S* alleles. For example, a $S_1S_1S_2S_2$ tetraploid plant will produce pollen carrying either two copies of the S_1 allele (i.e., S_1S_1), two copies of the S_2 allele (S_2S_2), or one copy of each allele (S_1S_2). In a self-pollination (i.e., pollination of an $S_1S_1S_2S_2$ style) the S_1S_1 and S_2S_2 pollen is rejected normally by the style but the S_1S_2 pollen is accepted. This effect was first observed in tetraploids of various species in the Rosaceae (Crane and Lawrence, 1931; Crane and Thomas, 1939) but most comprehensively described in a tetraploid pear variety called 'Improved Fertility' (Crane and Lewis, 1942; Lewis and Modlibowska, 1942). Lewis (1943) suggested that competitive interaction occurs because the two pollen S proteins compete to convert a limiting substrate into an allele-specific S-antigen recognised by the style's S-antibody. Competitive interactions between non-identical S alleles only occur in pollen and do not affect the SI phenotype of the style (Golz et al., 2000).

The way competitive interaction came to be viewed is through the inhibitor model, in which the S-RNase is a cytotoxin and pollen S is an inhibitor able to inactivate all S-RNase cytotoxins except those encoded by its cognate or matching S allele (Thompson and Kirch, 1992; Golz et al., 2000). So in a pollen grain carrying both the S_1 and S_2 alleles, the pollen S determinant associated with the S_1 allele would neutralise the S_2-RNases present in style and the pollen S determinant associated with the S_2 allele rwould neutralise the S_1-RNases. Because the two pollen S determinants have complementary inhibitory effects in combination they can inactivate all S-RNases, meaning that the pollen grain will be compatible on all styles (Golz et al., 2000). Based on competitive interactions from tetraploid plants, a transgenic experiment was performed in diploid SI plant to investigate if a second different *SLF* transgene expressed in pollen (heteroallelic pollen) can cause a competitive interaction with the endogenous SLF to alter SI response. Such an *in vivo* experiment was performed by Sijacic et al., (2004) where S_1S_1 *P. inflata* pollen was genetically manipulated to express the *S_2-SLF* which resulted in a self-compatible plant. Pollen viability was unaffected as the average number of seed set was comparable to wild type compatible pollination (Sijacic et al., 2004). Most recently, using the tomato genome release, Li and Chetelat (2015) identified 23 *SLFs* in *S. pennellii* (SI) genome (locus ui1.1) using *Petunia SLF* sequence as reference sequence. Based on the genomic location from *S* locus, as many as eight are ruled out as candidates and four are not included for *in vivo* study due to mutations present within each sequence. Among the 11 *SLF* genes selected for *in vivo* analysis, only *SLF23* is able to alter SI phenotype to self-compatible phenotype. This provides further evidence that *SLF* is *pollen S*.

Since *SLF* is shown to be *pollen S*, then what about the *SLF-likes*? A separate functional study by Hua et al., (2007) investigates if the *Petunia inflata SLF-likes* (*PiSLFLs*) can also function as *pollen S*. Three *SLF-like* constructs *PiSLFLb-S_2*, *PiSLFLc-S_1* and *PiSLFLd-S_2* were translationally fused to the green fluorescent protein reporter (GFP) and separately introduced into *Petunia* S_2S_3 plants. The pollen from each transgenic line was viable and possessed normal pollen function, suggesting that the introduced SLF-likes cannot competitively interact with the endogenous pollen S (Hua et al., 2007). However, with the rise of the collaborative non-self recognition theory, *PiSLFLb-S_2*, *PiSLFLc-S_1* and *PiSLFLd-S_2* are now classified as *SLFs* (see next section).

1.5. Early and current biochemical models of RNase-based SI

Two early RNase-based SI models, the receptor model and the inhibitor model, are illustrated in Figure 1.5. The receptor model predicts that SLF is a receptor on the pollen tube wall which interacts with S-RNases, allowing only S-RNase from identical *S* halpotype (self S-RNase) to enter a pollen tube. The inhibitor model predicts that any S-RNase can enter a pollen tube and interact with SLF in

a *S* haplotype specific manner. The *S* specific binding between an S-RNase and SLF determines whether S-RNase remains in the pollen tube or gets neutralised (Thompson and Kirch, 1992). The pollen part mutant analysis described by Golz et al., (1999) and S-RNase immunolocalisation described by by Luu et al., (2000) provide evidence supporting the inhibitor model which was the widely accepted model at the time this thesis began.

The reasons for so many *F-box* protein genes at or near the *S* locus remained a mystery until a recent functional study performed by Kubo et al., (2010) in *Petunia* showed that all function as *pollen S*. The new model was named the collaborative non-self recognition system and is illustrated in Figure 1.6. The 30 *F-box* isolated from *Petunia* (new and previously identified) are grouped into six types (1 to 6) based on phylogenetic relationship, named *SLFx-*S_n where x refers to type and S_n refers to the haplotype. For example, *SLF1-*S_1 means it is a type 1 *SLF* from S_1 allele. The key findings from their study includes; 1) that all *SLF-likes* identified from *Petunia* are all re-considered as *SLF*; 2) that *SLFs* can be classified into different types based on their phylogenetic relationship; 3) that each type of SLF will only elicit competitive interaction in a specific subset of *S* genotypes and hence will only interact and neutralise a specific subset of S-RNases; 4) that sequence variation is higher between the different types than within each type of SLF (Kubo et al., 2010). The new model states that each type of SLF is responsible for neutralising a specific subset of S-RNase(s) present within a population of plants. Together, all SLF types function together to neutralise all non-self S-RNases except the cognate S-RNase because the type of SLF that neutralizes the cognate S-RNase will not be present on that *S* haplotype. In other words, each SLF type will elicit competitive interaction in a specific subset of *S* genotype plants depending on which S-RNases it neutralises. They showed *S_7-SLF2* elicits SI breakdown in plants carrying the S_9, S_{11} or S_{19} but not in plants carrying the S_5 or S_{17} alleles. In addition, *semi in vivo* interaction assays complemented these *in vivo* observations (Kubo et al., 2010). Since this model predicts that each SLF recognises and neutralises only some of the S-RNases, SLF sequences between S haplotypes need not be as divergent as for the S-RNase. However, biochemically, exactly how each SLF type specifically recognises and interacts only with some S-RNases and not others remain unknown.

A separate study also showed that the *Malus* and *Pyrus SFBBs* can be grouped into different types based on their phylogenetic relationship. 25 *SFBBs* isolated from S_1 to S_6 genotypes of Japanese pear and together with others already isolated, *SFBBs* were grouped into eight different types and all *SFBBs* were renamed as *SFBBx-*S_n, the same nomenclature used in *Petunia*. Another similarity between *SLFs* and *SFBBs* is that within each type of *SLF*/*SFBB*, each is more similar to one another whereas sequence diversity is greater between the different types (Kakui et al., 2011). In addition,

William et al., (2014a) isolated eight and nine *SLFs* from the *P. inflata* S_2 and S_3 pollen transcriptome respectively. Interestingly, based on their phylogenetic relationship, they determined only seven out of the nine types of SLF present in S_3 pollen are *SLFs* and the remaining two candidates as *SLFLikes*. The function of *SLFLike1* and *SLFLike2* remains unknown.

Functional evidence that supports the collaborative non-self recognition system came from work done on a self-compatible cultivar of Japanese pear (*Pyrus pyrifolia*), Osa-Nijisseiki ($S_2S_4^{SM}$). Osa-Nijisseiki is a style-part mutant (S_4^{SM}) that lacks *S_4-RNase* and is also missing *S_4-SFBB1* (also known as *S_4F-box(0)*) due to a 236 kb deletion in the S_4^{SM} allele. *S_4-SFBB1*, which codes for an F-box protein, is suspected to be *SLF*, anassumption that has been questioned since *S_4-SFBB1* is within the deletion (Okada et al., 2008) and the collaborative non-self recognition system had not been proposed. Interestingly, S_4^{SM} pollen from this style part mutant plant was also rejected by S_1 style in addition to S_4 style (Okada et al., 2008). The same observation was made by Saito et al., (2012) through a series of crosses performed in *P. pyrifolia* that are summarised in Table 1.3. This suggests that type 1 *SFBB* most likely codes for an F-box protein that neutralises S_1-RNase. Coincidently, *S_5-SFBB1* is found to encode for a truncated S_5-SFBB1 but S_5 pollen was accepted by an S_1 style. If type 1 SFBB is the only type of SFBB that can neutralise S_1-RNase, then S_5 pollen, which lacks type 1 SFBB, should be rejected by the S_1 style. However, if there are multiple types of *SFBB* present at the *S* locus and there is another type of SFBB that can also neutralize S_1-RNase, then S_5 pollen would be accepted by an S_1 style. The acceptance of S_5 pollen by the S_1 style suggests that there is at least one other type of SFBB present on S_5 haplotype that can neutralise S_1-RNase. Taken together, type 1 SFBB is not the only type of SFBB that can neutralise S_1-RNase, which supports the idea that there are multiple types of *SFBB* functioning at the *S* locus (Kakui et al., 2011).

In addition to the three types of SLF already shown to interact with specific S-RNases *in vivo* by Kubo et al., (2010), Williams et al., (2014b) showed that another four types of SLF, types 4, 5 6 and 8, do the same. The four SLF types each elicit SI breakdown in one of the *S* backgrounds tested, unlike that observed in Kubo et al., (2010).

Kubo et al., (2015) provide further evidence with a more thorough survey of the *S* locus using next generation sequencing. They isolated 168 new SLFs from the male reproductive organs of *Petunia* and based on phylogenetic relationships categorized these into 18 distinct types of SLF. The alleles within an SLF type are highly similar to each other with variation observed between the different types. The absence of an SLF type from an S haplotype is an indication that the product of that gene interacts with the S-RNase of the same S haplotype, afinding that agrees with the collaborative non-self recognition model (Kubo et al., 2010). Interestingly, the S_5-3 SLF interacts with S_7-RNase but not

S_7-SLF3 and the latter is a more divergent sequence than the rest of the alleles present in type 3 SLFs. In summary, predictions of interactions between cognate partners are possible based on *SLF* divergence and deletion (absence) of an *SLF* from a *S* haplotype. To summarise the new theory, more than one study has found similarlties to Kubo et al (2010) and functional evidence supports the colloaborative non-self recognition theory. It is expected that the genomic patterning on each *S* haplotype, the SLF type that neutralise cognate S-RNase must be absent. As redundancy is present between *SLFs*, on some *S* haplotype more than one *SLF* would be missing. Next, *in vitro* or *semi in vitro* interaction experiment should agree with *in vivo* transgenic experiment as transgenic experiment is the conclusive test. Lastly, SLF type is determined based on phylogenetic analysis and sequence variation is detected between different types of SLF.

1.6. Different components of the biochemical complex

SLF/SFB was proposed to be part of an E3 ligase complex because F-box protein is well characterised as a component of an E3 ligase. There are different types of E3 ligase and the one that involves an F-box is SCF E3 ligase (Vierstra, 2003). A complete SCF E3 ligase consists of a Skp1, Cullin1, F-box and Rbx1. Figure 1.7 illustrates the predicted outcome of S-RNase based on the inhibitor model when SLF is part of an E3 ligase. Based on the known SCF E3 ligase complex, the NH_3 terminus of an F-box will bind to Skp which in turn binds to Cullin1 and Rbx1. The role of SCF E3 ligase is to transfer ubiquitin molecules from an E2 ligase onto a lysine on the target protein. Addition of further ubiquitins to a lysine on the previously added ubiquitin forms a polyubiquitin chain. Polyubiquitinated proteins are recognised by 26S proteasome which degrades the ubiquitinated protein (Vierstra, 2009). In RNase-based SI, a SCF E3 complex consisting of SLF (SCF^{SLF}) acts to transfer ubiquitin to a lysine on S-RNase which binds to the domain at the SLF carboxy terminal end. S-RNase is removed from the pollen tube through protein degradation (Hua and Kao 2006).

Compounding the complexity of RNase-based SI are the different forms of the SCF^{SLF} ligase complex isolated from the Solanaceae, Plantaginaceae and Rosaceae families. Huang et al., (2006), Zhao et al., (2010), Entani et al., (2014), Matsumoto et al., (2012) and Yuan et al., (2014) all identified SSK1 and Cullin1 as part of the SCF^{SLF} from plant species including *A. hispanicum, P. hybrida*, *P. avium* and *M. domestica*. SSK1 has high sequence similarity to Skp1 apart from an additional 7-9 amino acids found on the carboxy terminus (Zhao et al., 2010). In contrast, Hua and Kao (2006) identified a complex from *P. inflata* which contained SBP1. SBP1 is a RING finger protein (and hence presumably can function as a ubiquitin ligase in its own right) that is proposed to replace Cullin1 and Skp1. O'Brien et al., (2004) also discovered that SBP1 from *S. chacoense* interacts with S-RNase implicating it as being important in RNase-based SI responses. All of the complexes mentioned so far are

identified using either yeast two hybrid assays or *in vitro* interaction assays. Taken together, SSK1 and SBP1 likely play a role in RNase-based SI response.

In vivo functional evidence that Cullin1 and SSK1 are required for incompatible pollen rejection suggest that the SCF^{SLF} complex is required for RNase-based SI response (Li and Chetelat 2010; Li and Chetelat 2014; Zhao et al., 2010). *In vivo* functional studies aim to understand unilateral incompatibility (UI) and revealed that Cullin1 plays a role in both SI and UI system. UI is a reproductive barrier between two plant species and occurs when pollinations between the species are successful in one direction but unsuccessful in the other direction (Lewis and Crowe, 1958). UI often obeys the SI×SC rule, which means that pollen from the self-compatible (SC) parent is rejected by styles of the SI parent but the reciprocal cross is compatible. Li and Chetelat (2010) isolated two loci in *Solanum*, ui1.1 (*SLFs* are located in this locus, section 1.4) and ui6.1, both are essential for UI. The ui6.1 locus encodes Cullin1 based on high sequence similarity to the Cullin1 involved in RNase-based SI. In the red-fruited SC species *Solanum lycopersicum* the ui6.1 locus codes for a truncated Cullin1 but in the green-fruited SI species *Solanum pennellii* ui6.1 codes for a functional Cullin1. Interestingly, other red/orange fruited SC tomatoes have non-functional Cullin1 genes while most green fruited tomatoes have a functional copy of *Cullin1*. To understand Cullin1 importance for UI response *S. lycopersicum* plants with or without a transgene copy of *S. pennellii Cullin1* (spCul1) are used as pollen donors on SI tester plants. Pollen carrying spCul1 is compatible with the tester plant but pollen without spCul1 is incompatible with tester plant, showing that a functional Cullin1 is essential for UI in the presence of ui1.1. This is evidence that the product of ui6.1 (*Cullin1*) is able to interact with the product of ui1.1 genetically providing support that the molecules involved in SI and UI are similar (Li and Chetelat 2010).

If Cullin1 and SSK1 are essential components of the SCF^{SLF} ligase complex for compatible pollination, a decrease in its transcript level of either in pollen should significantly affect pollen viability. This can be achieved by T-DNA insertions carrying RNA interference which knock down gene expression of either Cullin1 or SSK1. If the transmission of T-DNA insertion is distorted shows that SSK1 and Cullin1 are functionally important for compatible pollination.

Solanum arcanum (formerly *Lycopersicon peruvianum*) line LA2163 is SI while line LA2157 is SC due to non-functional S-RNase and the two lines are cross-compatible (Kowyama et al., 1994). The expression of Cullin1 in LA2163 is reduced by RNA interference through T-DNA insertion driven by tomato pollen-specific promoter producing a mixed population of transgenic pollen carrying the T-DNA insertion or not (Li et al., 2014). Since LA2157 has a non-functional S-RNase, the absence of Cullin1 should not affect transgenic pollen viability and transmission of the T-DNA insert. Indeed, all

pollen are viable on a LA2157 tester plant and progeny have the expected 1:1 segregation ratio for single copy T-DNA insert. In contrast, not all transgenic pollen are viable on LA2163 tester plants because transgenic pollen carrying the T-DNA insertion will not be able to convey pollen resistance due to the reduced level of Cullin1. Inheritance of the T-DNA insert in progeny plants is about 10%, much less than the expected 50%. Altogether, the distorted segregation ratio for the T-DNA insert show that reduced levels of Cullin1 affect transmission of T-DNA inserts and supports the idea that Cullin1 is essential for pollen function in SI (Li et al., 2014). Another similar study performed in *P. hybrida* showed that knock down of SSK1 is associated with reduced cross-compatibility and a distorted segregation ratio of T-DNA inserts, suggesting that SSK1 is also essential for pollen function in SI (Zhao et al., 2010). In conclusion, functional analysis shows that the SCF^{SLF} complex is essential for RNase-based SI.

1.7. Ubiquitination of S-RNase

The first *in vitro* S-RNase ubiquitination degradation assay performed revealed that bacterially produced non-glycosylated S-RNases are degraded in a non-*S* specific manner (Hua and Kao 2006). Glutathione *S*-transferases (GST) tagged S_1- and S_2-RNases are degraded when incubated with S_2 pollen tube extract in the absence of the 26S proteasome inhibitor, MG132. When MG132 is added, intact recombinant S_1- and S_2-RNase are detected. An identical experiment was performed using native glycosylated S_3-RNase incubated with S_1, S_2 and S_3 pollen tube extract with or without MG132. The outcome was native S_3-RNase remains intact, unaffected by MG132. Negative controls used in the study include GST tag alone and GST tagged RNase X2, a RNase that is not involved in RNase-based SI. Only the latter is degraded in S_2 pollen tube extracts without MG132 (Hua and Kao 2006). Altogether, the evidence suggests degradation of S-RNase by pollen tube extracts is not *S* specific, nor is it specific to S-RNases, and non-glycosylated S-RNases are more sensitive to degradation *in vitro*. An *in vitro* S-RNase ubiquitination assay showed that self, non-self recombinant non-glycosylated S-RNases and RNase X2 were ubiquitinated *in vitro* but the GST tag was not. Taken together, the ubiquitination and degradation of S-RNase is not *S* specific and is also not specific to S-RNase (Hua and Kao, 2006).

More recently, Entani et al., (2014) and Yuan et al., (2014) isolated a SCF^{SLF} complex from *Petunia hydrida* and *Malus domestica* respectively. Yuan et al., (2014) shows different SCF^{SLF} type complexes selectively interact with a specific subset of S-RNases and SBP1 could not replace the role of SSK1 to cause ubiquitination of S-RNase (Yuan et al., 2014). Entani et al., (2014) shows S_7-SLF2 selectively ubiquitinates S_9- and S_{11}-RNase but not S_5- or S_7-RNase. This finding is in agreement with the *in vivo* studies performed by Kubo et al., (2012). In addition, ubiquitinated S_9-RNase was found to

accumulate in the presence of MG132 suggesting the 26S proteasome is responsible for the degradation of ubiquitinated S_9-RNase.

1.8. Summary

Based on the genetics of RNase-based SI, it is expected that there is a single *pollen S* at the *S* locus that controls pollen function. Hence, a pollen-specific *F-box* located at the *S* locus which displays *S* haplotype specificity is an ideal candidate for pollen S. The definitive evidence *SLF* is *pollen S* is *SLF* causes competitive interaction *in vivo*. However, subsequent studies identify multiple *SLFs/SFBs*. The rise of the collaborative non-self recognition theory sought to explain how multiple *SLFs* function together to neutralise all S-RNases with experimental data currently supporting the new theory. An important characteristic of *pollen S* is its tight linkage to the *S* locus; however, some *F-box* genes undergo recombination with the *S* locus and should be considered as SLFLs.

Since F-box protein is known for its role as part of a larger complex involved in 26S proteasome pathway, it is proposed SLF may function similarly. Interaction assays have isolated different forms of the SCF^{SLF} complex which has a specialized form of Skp1 called SSK1 and is able to ubiquitinate S-RNases for likely degradation by the 26S proteasome.

Recently, it is revealed that F-actin integrity, vacuolar compartments disorganisation and PCD are related to RNase-based SI in *P. pyrifolia* but earlier work in *N. alata* has shown that incompatible pollen tubes remains viable. In addition, F-actin integrity, vacuolar compartment disorganization are also observed in rejection of pollen tubes in *N. alata* but there is no evidence to suggest it undergoes PCD. To better understand this complex system will perhaps require knowledge associated with other aspect of pollen tubes such as pollen tube elongation and programmed cell death in plants.

1.9. T2-RNases and its new role in plant biology

Recent studies on T2-RNases have revealed that transfer RNAs (tRNAs) as a likely endogenous substrate in many eukaryotic cells although as enzymes they can cleave single- and double-stranded RNA and DNA–RNA hybrid substrates as well as having biological functions unrelated to its RNase activity such as cell survival and regulation of translation (Lee and Collins, 2005; Haiser et al., 2008; Thompson and Parker, 2009; Yamasaki et al., 2009; Zhang et al., 2009). T2-RNases are transferase type RNases that cleave RNA endonucleolytically liberating oligonucleotides and/or mononucleotides with terminal 3' phosphate via 2',3' cyclic phosphate intermediate (Deshpande and Shankar, 2002). T2-RNases are found in a wide range of organisms including bacteria, plants and animals such as humans and are also found in some viruses including some which are highly contagious like classic swine fever virus. They are usually active under acidic condition and located in

organelles such as vacuoles or lysosomes. Unlike other families of RNases such as RNase A or T1-RNase, which cleave RNA at a specific base, T2-RNase cleaves at all four bases (Luhtala and Parker, 2010).

A T2-RNase from yeast (*Saccharomyces cerevisiae*), Rny1p, cleaves not only rRNA but also tRNA (Thompson and Parker, 2009). In addition, it is involved in regulating cell survival which is not associated with its ribonuclease activity. Rny1p is present within the vacuole of yeast cells and evidence suggests that when the cell is exposed to oxidative stress, Rny1p is released into the cytosol where it has access to rRNA and tRNA and cleavage of RNAs occurs. The function of cleaved RNAs remains unknown. The expression of a catalytic inactive Rny1p and functional Rny1p in yeast cells reveals both forms cause cells to be hypersensitive to oxidative stress and decrease cell viability, suggesting its role in cell death does not require ribonuclease activity. It is proposed that due to cellular damages or upon the cell entering the stationary phase, Rny1p released from the vacuole into the cytosol triggered two separate downstream responses. One is the cleavage of tRNA and the other not requiring its ribonuclease activity is activating cell death. The latter could also be used as a marker for assessing cellular damage. Interestingly, experimental data shows that Rny1p function can be complemented by its human orthologue *Rnaset2,* as well as other structurally unrelated secretory RNases (MacIntosh et al., 2001; Thompson and Parker, 2009). This suggests the functional role of T2-RNase is conserved between humans and yeast and possibly other organisms that have T2-RNase (Thompson and Parker, 2009).

A separate study found Rnaset2 in zebrafish is required for normal rRNA turnover and also contributes to neurodegenerative disease unrelated to its function as ribonuclease (Haud et al., 2011). Rnaset2 is predominantly localized in the lysosome of brain tissue, with some amount detected in endoplasmic reticulum and Golgi apparatus. Zebrafish expressing a mutant form of *Rnaset2* (*A0127*) which lacks RNase activity accumulated rRNA within the lysosomes in the brain contributing to leukoencephalopathy, a form of lysosomal storage disorder. Another characteristic display by A0127 is the deposition of white matter throughout the brain and using an antibody against amyloid precursor protein (APP) detected APP in area where the white matter is located. Hence, Rnaset2 is not only essential for ribosome recycling in brain, but also contributes to neuron degenerative disease in humans (Haud et al., 2011).

Human angiogenin, an angiogenic factor in a tumour cell line, is a secreted RNase which is more closely related to RNase A than T2-RNase (Yamasaki et al., 2009). Angiogenin is responsible for the accumulation of cleaved tRNA in mammalian cell line exposed to oxidative stress. The cleavage of mature tRNA occurs around the anti-codon loop give two unequal halves; the 3’ end of tRNA is slight

larger than the 5' end tRNA. The tRNA 5' end is able to inhibit protein synthesis suggesting that the 5' tRNA end may function as small RNA-protein complex involved in repressing translation. It is possible angiogenin responds to external stress because as a secreted protein, it can easily pass on the survival signal to adjacent cells as well as cells at a distance away to switch to cell survival mode by repressing translation (Yamasaki et al., 2009).

Apart from animals, cleaved tRNA fragments are also isolated from plant tissues where they have a role in inhibiting protein translation (Zhang et al., 2009). The tRNA fragments containing the CCA 3' modified end suggest they are derived from mature tRNA. Endogenous tRNA fragments isolated from the phloem sap of pumpkin are able to inhibit protein translation suggesting that these fragments may be part of the RNA-protein complex that interferes with ribosomal activity. Interestingly, tRNA ribonuclease activity is detected in nearby tissue such as stem and leaf but not in phloem sap indicating that tRNA is likely cleaved in other tissue and delivered into the sieve element for it to be sent to other parts of a plant. Therefore, tRNA fragments also act as signalling molecules in plants apart from repressing translation (Zhang et al., 2009).

Another plant T2-RNase, RSN2 (*Arabidopsis*) is slightly different from other T2-RNase as its optimum activity is at neutral pH rather than acidic pH. RSN2 is located in multiple locations including the vacuole, endoplasmic reticulum (ER) and ER-derived bodies. A transgenic line expressing a non-active RSN2 or with reduced level of RSN2 has an increased in rRNA half life as compared to wild type plant. In addition, under sucrose starvation, both mutant and knock down RSN2 transgenic line showed constitutive autophagy not seen in the wild type plant. This may be because under nutrient starvation, ribosomes may be selectively degraded to provide nutrient in the form of nucleotides. This shows that RSN2 is required for normal rRNA turnover and cell survival (Hillwig et al., 2011).

In summary, studies of T2-RNases reveal they have broader ribonuclease specificity, are compartmentalised within the cell in vacuoles and have other important biological functions that are not associated with its ribonuclease activity. These findings open up new possibilities for studies of S-RNase function. As shown by Goldraij et al., (2006), S-RNases compartmentalised into the vacuole once inside the pollen tube and can also degrade RNA (McClure et al., 1990). A possible prediction would be S-RNases may also cleaves other forms of RNA such as tRNAs, or could have another important biological function that is not associated with its ribonuclease activity.

1.10. Aims of this thesis

The first aim of this thesis was to investigate the RNase-based SI mechanism at transcript level, using *N. alata* as the model organism. The approach was to perform next generation sequencing and de

novo assembly on RNA extracted from *N. alata* pollen grains to isolate previously unidentified *DDs* and other RNase-based SI related transcripts reported by other studies. From this investigation, it may be possible to speculate on the form of the SCF^{SLF} complex present in *N. alata* pollen. The results are presented and discussed in chapter 2 of this thesis.

The second aim focussed on studying the protein-protein interaction between the DDs and *Petunia* SLFs and the S-RNases using either pull down or co-immunoprecipitation assays. At the start of this thesis it was assumed that one of the ten *DDs* was *pollen S* and only this protein will interact with S-RNases. However, based on the collaborative non-self recognition theory, all of the DDs could potentially be considered as pollen S and interact with one or more different S-RNases. The approach taken was to express DDs/SLF and S-RNases in *E. coli* as recombinant proteins with tags to facilitate later purification steps. Interactions between SLFs and S-RNases would be studied using purified recombinant proteins and pull down or co-immunoprecipitation assays. The results of this study are presented and discussed in chapters 3 and 4 of this thesis.

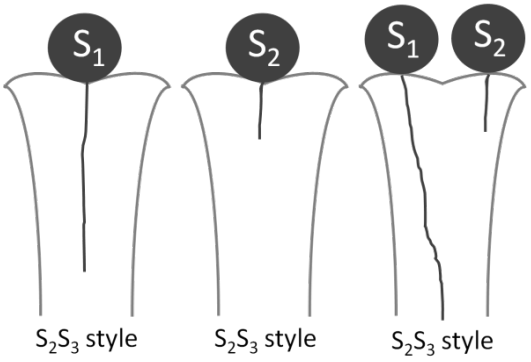


Figure 1.1: Gametophytic self-incompatibility system.

Style is diploid and in this case is S_2S_3. Pollen is haploid and in this case there are S_1, S_2 and S_3 pollen. Cross-compatibility is only possible when the *S* alleles expressed by pollen and style are different. On the left is a compatible pollination event, a S_1 pollen grain is compatible with S_2S_3 style and is accepted. In the middle is a self-pollination (incompatible pollination) event, a S_2 pollen grain is incompatible with S_2S_3 style and hence is rejected. On the right is a semi-compatible pollination event, S_1 pollen grain is accepted as it is compatible with S_2S_3 style but S_2 pollen grain is incompatible with S_2S_3 style and hence is rejected. In an event of self-pollination, growing pollen tube is significantly reduced and usually would not grow pass the top half section of the style and hence fertilisation of ovule cannot take place. In an event of cross-pollination, pollen tube would continue to grow through the style reaching the ovule resulting in fertilisation of ovule. In some GSI systems (e.g., those in the Papaveraceae family), pollen tube growth is arrested on the stigma surface, not within the style as shown here (modified from Newbigin et al., 1993).

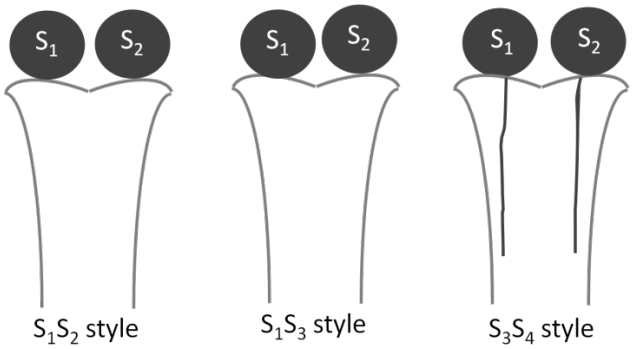


Figure 1.2: Sporophytic self-incompatibility system.

Genotype of the pollen parent (sporophyte) is S_1S_2. When an allele in the pollen parent matches that of the pistil (e.g., S_1S_2 or S_1S_3), pollen germination is arrested at the stigma surface. Where there is no match (S_3S_4), the pollen may germinate and grow through the style to the embryo sac. The central panel applies only if the S_1 allele is dominant to or codominant with S_2 in the pollen and if S_1 is dominant to or codominant with S_3 in the style. If S_3 is dominant to S_1 in the style, or if S_2 is dominant to S_1 in the pollen, pollen from the S_1S_2 parent will be compatible (modified from Newbigin et al., 1993).

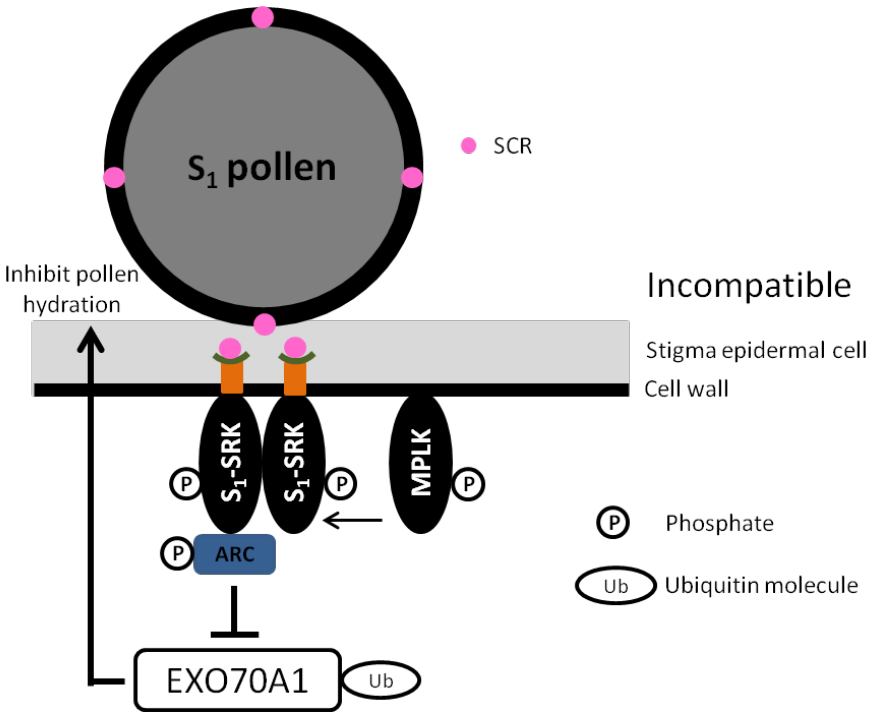


Figure 1.3: SI signalling in the *Brassicaceae*. The pollen grain (S_1) encodes for a pollen coat protein, SCR interacts with the female determinant protein, S-receptor kinase (SRK) in a allelic manner. In a self-pollination event, autophosphorylation and dimerisation of SRK occurs. Phosphorylated SRK interacts with phosphorylated M-locus protein kinase (MPLK). Together, they cause the phosphorylation an armadillo repeat containing (ARC1) protein. ARC is U-box E3 ligase and negatively regulates the ubiquitination of EXO70A1 which in turn affects pollen tube hydration to prevent self-fertilisation (modified from Tantikanjana et al., 2010).

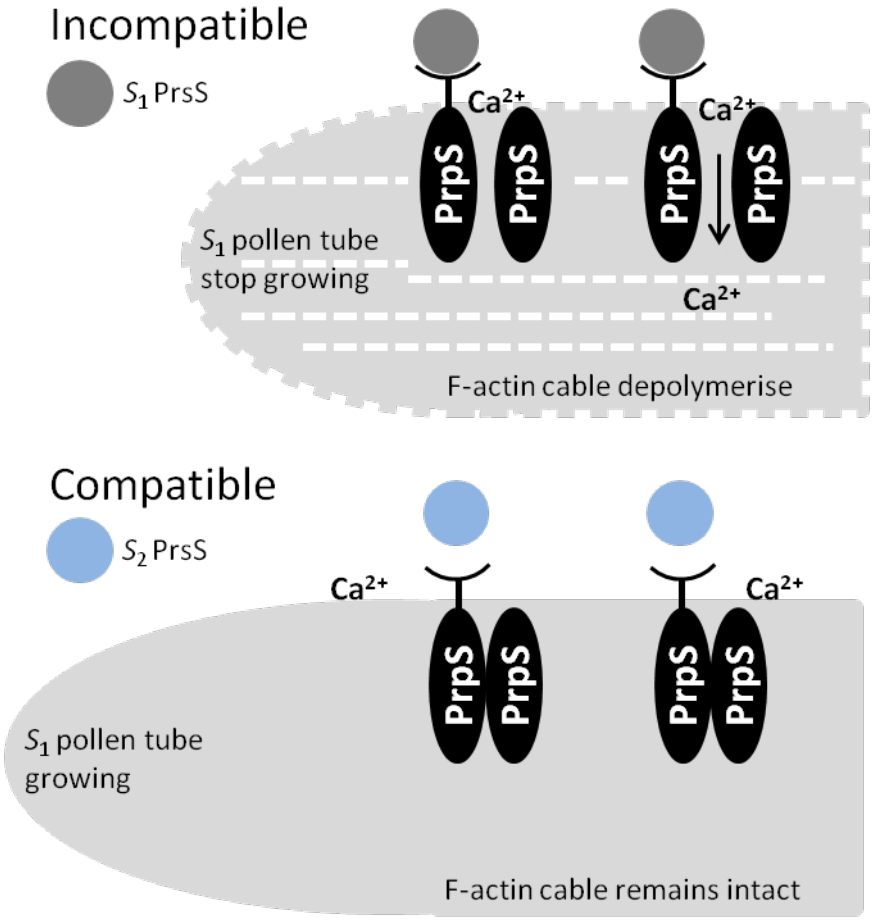


Figure 1.4: SI signalling in *Papaver rhoeas*. *P. rhoeas* pollen S (PrpS) is predicted to be a transmembrane protein which acts a receptor protein for *P. rhoeas* stigma S (PrsS). They interact in a *S* allelic manner and in an event of self-pollination triggers the influx of calcium ion (Ca^{2+}) ion into incompatible pollen tube. This affects F-actin integrity and the final outcome of pollen tube inhibition is the activation of programme cell death. In a compatible pollination, interaction between S_2 PrsS and S_1 PrpS prevents Ca^{2+} influx into pollen tube maintains the F-actin cable integrity which is essential for continuous pollen tube growth for fertilisation of ovule (modified from Wheeler et al., 2010).

Table 1.1: A list of studies which identified single *pollen S*.

Family	Species	pollen S name	Method	Reference
Solanaceae	*Petunia axillaris*	*PaSLF17, PaSLF19*	PCR	Tsukamoto et al., (2005)
Plantaginaceae	*Antirrhinium hispanicum*	*AhSLF1, SLF2, SLF4, SLF5*	Sequencing *S* locus	Lai *et al*., (2002)
Rosaceae				
Prunoideae	*Prunus mume*	*PmSLF1, SLF7*	Sequencing *S* locus	Entani *et al*., (2003)
	Prunus dulcis	*PdSFBa, SFBb, SFBc, SFBd, PdSLFc and PdSLFd are SLF-likes equvialent.*	Sequencing *S* locus	Ushijima *et al*., (2003)
	Prunus avium	*PaSFB3, 6*		Yamane *et al*., (2003)
	Prunus avium	*PaSFB1, 2, 4, 5*	PCR	Ikeda *et al*., (2004)
Maloideae	*Malus* x *domestic*	*SLF1, SLF2*	PCR	Cheng *et al*., (2006)

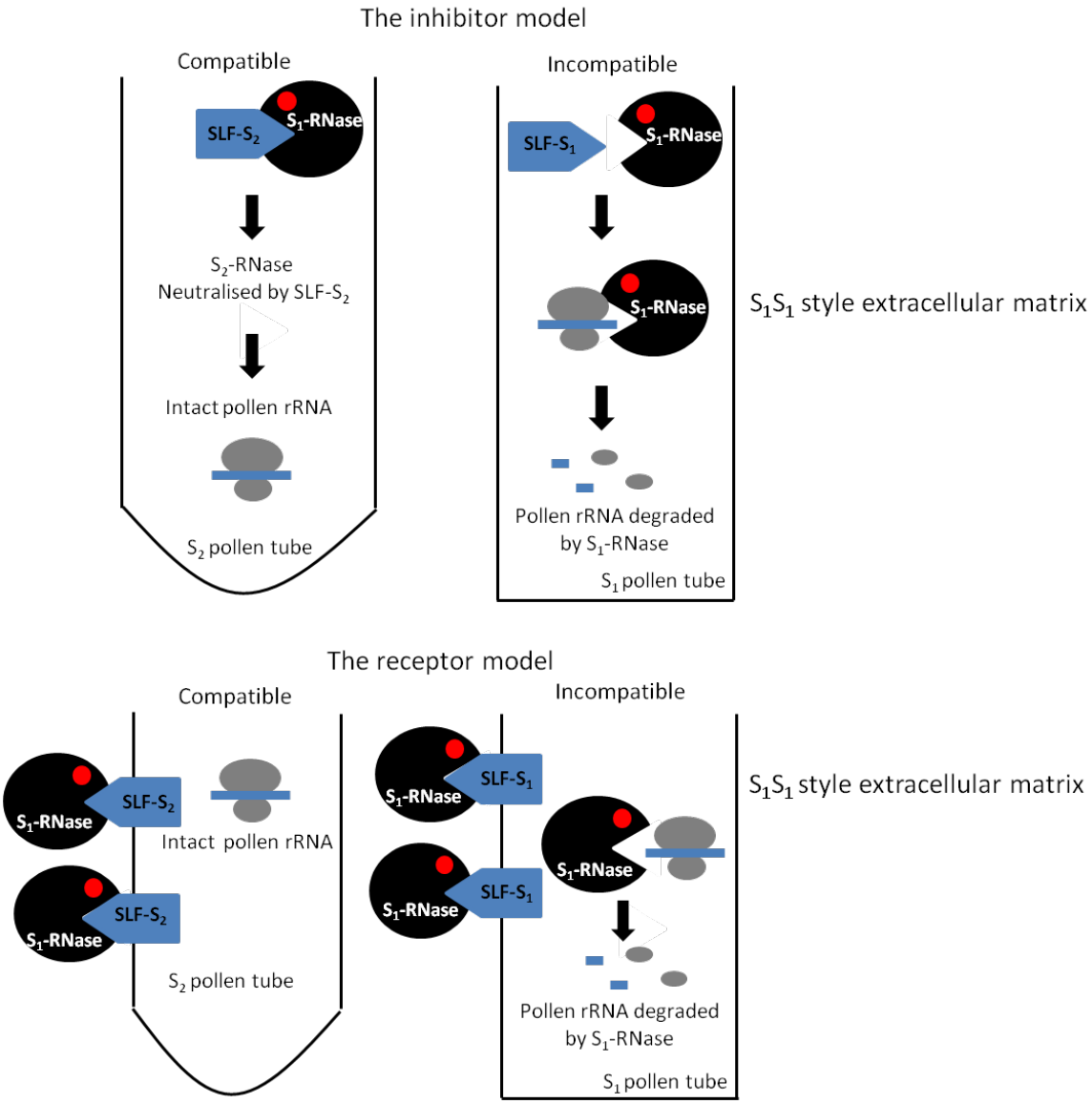
Table 1.2: A list of studies which identified multiple *pollen S*.

Family	Species	pollen S name		Reference
Solanaceae	*Nicotiana alata*	*DD1-DD10*	PCR	Wheeler *et al*., (2007)
	Solanum pennellii	*SpSLF 1-23*	Sequencing S locus	Li and Chetelat (2015)
	Petunia inflata	*PiSLF1-3*	Sequencing *S* locus	Wang *et al*., (2003), Wang *et al*., (2004), Sijacic et al., (2004)
	Petunia inflata	*SLF type 1 to 6*	PCR	
	Petunia inflata	*SLF type 4 to 6*	PCR	
	Petunia inflata	*SLF type 11 to 17*	Next generation sequencing	Kubo et al., (2010)
	Petunia inflata	*SLF type 1 to18*	Next generation sequencing, PCR	Williams et al., (2014a) Williams et al., (2014b) Kubo et al., (2015)
Rosaceae	*Malus* x *domestica*	*MdSFBB9a, MdSFBB9b, MdSFBB3a, MdSFBB3b*	Sequencing *S* locus	
	Pyrus pyrifolia	*PpSFBB4a, PpSFBB4b, PpSFBB4g, PpSFBB5a, PpSFBB5b, PpSFBB5g*	PCR	Sassa *et al*., (2007)
	Pyrus pyrifolia	*PpSFBB1g, PpSFBB2g, PpSFBB3g, PpSFBB6g, PpSFBB7g, PpSFBB8g, PpSFBB9g*	PCR	Kakui *et al*., (2007)
	Pyrus pyrifolia	*PpSFBB4-u1, PpSFBB4-u2, PpSFBB4-u3, PpSFBB4-u4, PpSFBB4-d1, PpSFBB4-d2. PpSFBB2-u1, PpSFBB2-u2, PpSFBB2-u3, PpSFBB2-u4, PpSFBB2-u5, PpSFBB2-d1, PpSFBB2-d2, PpSFBB2-d3, PpSFBB2-d4, PpSFBB2-d5.*	Sequencing *S* locus	Okada *et al*., (2011)
	Malus x *domestica*	*MdFBX1-20, they are MdSFBBs equvialent but named differently.*	PCR	Minamikawa *et al*., (2010)
	Pyrus pyrifolia	*Type 1 to 8 SFBBs* (25 new *SFBBs*)		Kakui *et al*., (2011)
	Pyrus communis L.	16 *SFBBs* isolate from 4 S-haloptype grouped into five types: *SFBBα, SFBBγ, SFBBβ, SFBBε, SFBBδ*	PCR	De Franceschi *et al*., (2011a)
	Pyrus communis L, *Malus* x *domestica*	*SFBBα, SFBBγ, SFBBβ, SFBBε, SFBBδ* (isolated additional 67 sequences)	PCR	De Franceschi *et al*., (2011b)

Figure 1.5: Early models of RNase-based SI.

Top: The inhibitor model. S_1-RNase enters the S_2 pollen tube from the style extracellular matrix. *S* specific interaction between S-RNase and SLF in compatible and incompatible pollen tube will determine the fate of S-RNase. In a compatible pollination by non-self S_2 pollen (left), S_2-SLF will interact with S_1-RNase and neutralise it. Hence, pollen ribosomal RNA remains intact, the pollen tube continues to grow resulting in fertilisation of ovule. In an incompatible pollination by self S_1 pollen, S_1-SLF is unable to neutralise S_1-RNase, S_1-RNase remains in pollen tube degrades pollen ribosomal RNA which prevent self-fertilisation of ovules due to the reduced pollen tube growth rate (modified from Golz et al., 2000).

Bottom: The receptor model. SLF are receptors on the wall of the pollen tube. *S* specific interaction between SLF and extracellular S-RNase will only allow self S-RNase (S-RNase is of the same *S* haplotype as SLF) to enter the pollen tube after an incompatible pollination. Non-self S-RNase (S-RNase is of a different *S* haplotype as SLF) entry into the compatible pollen tube is always prevented. In a compatible pollen tube (left), the S_2-SLF receptor prevents S_1-RNases from entering the compatible pollen tube. Pollen ribosomal RNA remains intact, the pollen tube continues to grow resulting in fertilisation of ovule. In an incompatible pollen tube, S_1-SLF receptor would allow S_1-RNase to enter pollen tube. S_1-RNase enters the pollen tube, degrades pollen ribosomal RNA which prevent self-fertilisation of ovules due to the reduced pollen tube growth rate (modified from Golz et al., 2000).


The inhibitor model
Compatible
Incompatible
SLF-S2
S1-RNase
S2-RNase
Neutralised by SLF-S2
Intact pollen rRNA
S2 pollen tube
SLF-S1
S1-RNase
Pollen rRNA degraded
by S1-RNase
S1 pollen tube
S1S1 style extracellular matrix
The receptor model
Compatible
Incompatible
S1-RNase
SLF-S2
Intact pollen rRNA
S2 pollen tube
SLF-S1
Pollen rRNA degraded
by S1-RNase
S1 pollen tube
S1S1 style extracellular matrix

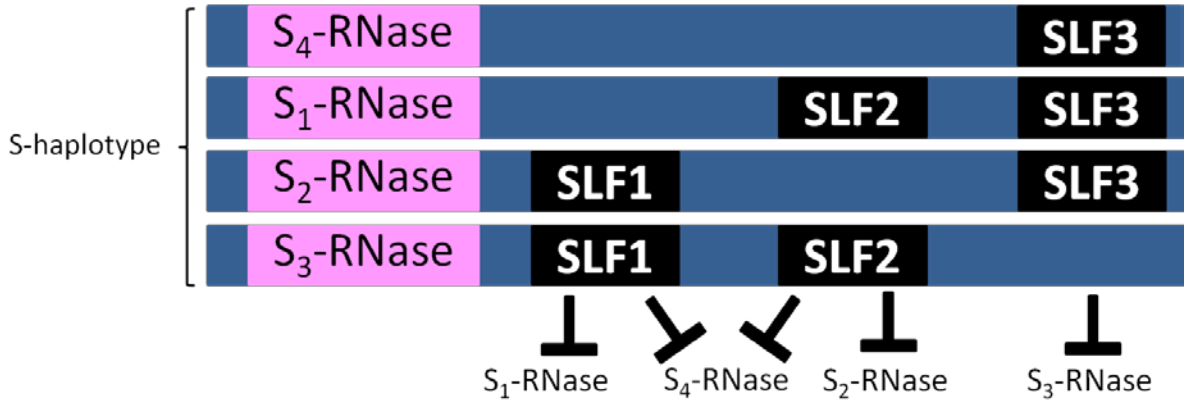


Figure 1.6: The collaborative non-self recognition system.

Multiple SLFs are present at the *S* locus and these SLFs are grouped into different types. Each type of SLF will only specifically recognise a subset of S-RNase and neutralise them. Together, all types of SLF present on a *S* haplotype will be able to neutralize all non-self S-RNases. S-RNase neutralisation redundancy is expected among the different types of SLF. For example, both SLF1 and SLF2 can neutralise S_4-RNase. This system predicts that the type of SLF that neutralise cognate S-RNase will not be present as the expression of this SLF will result in self-fertilisation (modified from Kubo et al., 2010).

Table 1.3: S_4^{SM} pollen rejection by S_1 style.

Crosses (pollen x style)	Expected F1 segregation ratio and phenotype	Results from F1 self crossing	Comment
$S_4^{SM}S_4^{SM}$ x S_5S_7	$S_4^{SM}S_5$: $S_4^{SM}S_7$ 1 : 1 SC : SC	All F1 progenies are SC	F1 phenotype are as expected. Plant that carries S_4^{SM} allele will produce functional S_4 pollen but non-functional S_4-RNase.
$S_4^{SM}S_4^{SM}$ x S_4S_5	$S_4^{SM}S_4$: $S_4^{SM}S_5$ 1 : 1 SI : SC	50% of F1 progenies are SC, 50% are SI	
$S_4^{SM}S_4^{SM}$ x S_1S_2	$S_4^{SM}S_1$: $S_4^{SM}S_2$ 1:1 SC : SC	Equal ratio of SI and SC F1 progenies	Pollen carrying S_4^{SM} allele is rejected by either S_1 or S_2 allele present in style

Crosses (pollen x style)	Seed set	Result of crossing	Comment
S_2S_4 x S_1S_2	Yes	compatible	S_4 pollen is accepted by S_1S_2 style but not S_4^{SM} pollen.
$S_2S_4^{SM}$ x S_1S_2	No	incompatible	
$S_4^{SM}S_4^{SM}$ x S_1S_6	No	incompatible	S_4^{SM} pollen is rejected by S_1 style.
$S_4^{SM}S_4^{SM}$ x S_1S_7	No	incompatible	

SC: Self-compatible phenotype, SI : Self-incompatible phenotype. F1 : first generation progeny.

S_4^{SM} allele carries a non-functional S-RNase but pollen function remains normal. Hence any style carrying S_4^{SM} allele will accept S_4 pollen and plant is thus self-compatible. S4 pollen is accepted by S_1S_2 but not S_4^{SM} pollen, in addition, SI phenotype follows S_1 allele shows S_4^{SM} pollen is also rejected by style with S_1 allele (Saito et al., 2012).

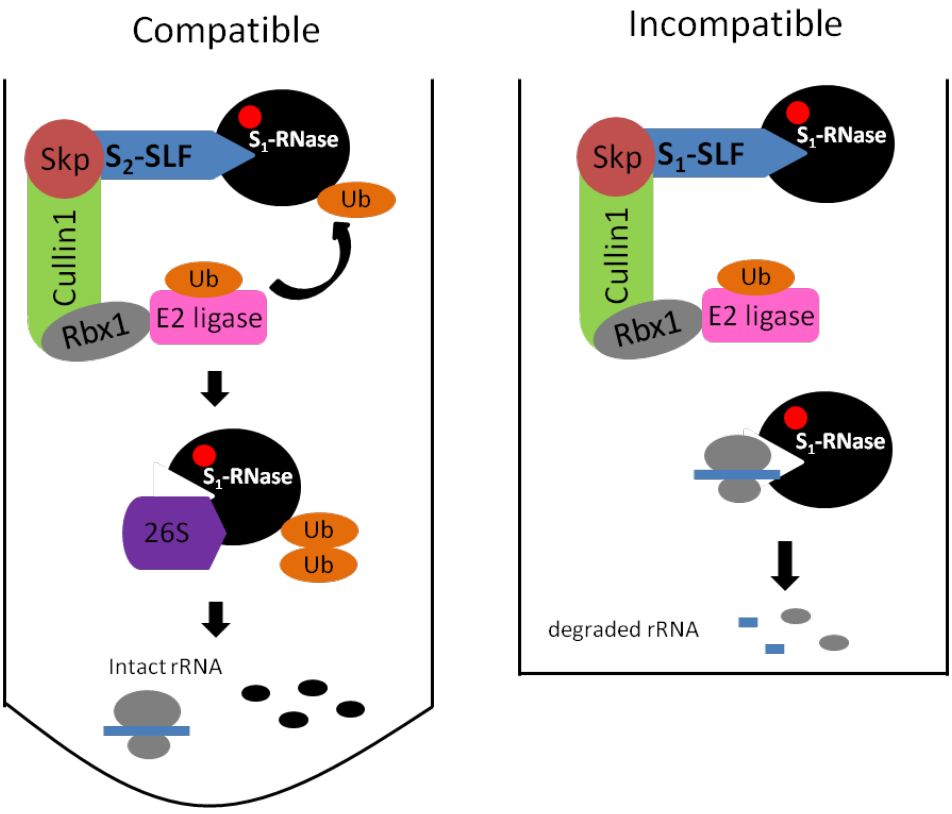


Figure 1.7: The SCF^{SLF} ligase complex.

The proposed SCF^{SLF} ligase model involved in RNase-based SI response consist of Skp1, Cullin1, Rbx1 and SLF. This complex forms in all pollen tube and is ready to "capture" any S-RNase that enters the pollen tube cytoplasm (modified from Viestra 2003; Qiao et al., 2004a).

Left: Compatible pollination. The SCF^{SLF} ligase complex form consists of a S_2-SLF and captured a S_1-RNase. In this case, the SCF^{SLF} ligase adds ubiquitin molecules to the lysine unit of S_1-RNase. Poly-ubiquitinated S_1-RNase is recognised and degraded by 26S proteasome. S_1-RNase is removed from the compatible S_2 pollen tube, pollen RNA remains intact and continuous pollen tube growth allows fertilisation of ovary.

Right: Incompatible pollination. The SCF^{SLF} ligase complex forms which consist of a S_1-SLF capturing an S_1-RNase. The SCF^{SLF} ligase will not be able to ubiquitinate S_1-RNase. S_1-RNase remains within the pollen tube, degrades pollen RNA causing a slow down in pollen tube growth. In this case, fertilisation of ovary will not occur because of the reduced pollen tube growth.

2.1: Introduction

The history of using *Nicotiana* to study GSI (gametophytic self-incompatibility) is indeed a long one, dating back nearly 100 years to the original description of genetic basis of self-sterility in *Nicotiana sanderae* by East and Mangelsdorf (1925). From the original elegant simplicity of a single *S* locus with multiple *S* alleles, a more complex picture of the *S* gene has emerged. As a result of mutational studies, Lewis (1960) proposed a tripartite structure for the *S* gene composed of three linked segment, each one controlling a separate aspect of the SI response: one segment for the style response, another for the pollen response and a third segment that controlled overall allelic specificity. Although subsequent molecular studies have failed to confirm this tripartite structure, the trend from simple towards more complex models of the *S* locus has continued. This trend is particularly evident in discussions of the nature of the segment controlling the pollen response (*pollen S*).

Throughout the 1990s, it was thought there was a single *pollen S* gene encoding the male counterpart of the female S-RNase encoded by the *style S* gene (e.g., McCubbin et al., 1997). Eventually, searches for *pollen S* resulted in the identification of an *F-box protein* gene (called *S locus F-box* or *SLF*) as encoding the male determinant (Lai et al., 2002; Sijacic et al., 2004). Questions about the relationship between *SLF* and *pollen S* persisted however (Newbigin et al., 2008), culminating in the discovery that *pollen S* wasn't encoded by a single gene but by multiple genes and the development of a "collaborative non-self recognition" model for S-RNase-based SI (Kubo et al., 2010). In this model each type of SLF interacts with a subset of non-self S-RNases so that collectively all SLFs recognise and detoxify all the S-RNases in a species. Although it is currently unknown how many different *SLF* genes make up *pollen S* in species in which the collaborative non-self recognition model appears to be operating, the number suggested for *Petunia inflata* is 17 (see Williams et al., 2014a; Williams et al.,2014b).

The first attempted to identify *pollen S*/*SLF* in *N. alata* was by Wheeler and Newbigin (2007), who used PCR and degenerate primers to identify a family of ten pollen-expressed *SLF-related* genes that they termed the *DD* genes. As all ten *DD* genes are located at or near the *S* locus, collectively they potentially comprise the *pollen S* genes of this species. There are, however, possibly other *DDs* in *N. alata* that were not amplified with the degenerate primers used by Wheeler and Newbigin (2007). Additionally, the pollen-expressed genes encoding other components of the SCF^{SLF} ligase complex of which SLFs are a part are unknown in *N. alata* although various components of this complex in other Solanaceae species have been suggested (Hua and Kao, 2006; Huang et al., 2006; Li et al., 2014; Entani et al., 2014).

This chapter uses RNA-Seq, an approach to transcript profiling (Wang et al., 2009), as a way of identifying as many of the transcripts expressed in *N. alata* pollen grains as possible. As the *N. alata* genome has not been sequenced, the RNA-Seq data was first *de novo* assembled into a transcriptome and the accuracy of the assembly confirmed before the transcriptome could be search for sequences related to the *DD* genes and to other genes suggested to be part of the SCF^{SLF} complex. Some of the research in this chapter has been published in the paper by Lampugnani et al., (2013). A similar transcriptomics-based approach to *SLF* discovery by Williams et al (2014a) appeared in the final stages of writing this thesis. Results from that paper have not been incorporated into this chapter although some of the points it raises are mentioned in the discussion.

2.2: Materials and Methods

2.2.1: RNA-Seq library preparations and sequence generation

Nicotiana alata plants (self-incompatibility genotype S_2S_3) were grown in soil in a glasshouse as previously described (Anderson et al., 1986). Pollen grains were collected and stored at -80°C until used. Total pollen grain RNA was extracted with an RNA extraction kit (Qiagen) according to manufacturer's protocol and 10 µg was sent to Australian Genome Research Facility (Brisbane, Australia) for mRNA-Seq library preparation and sequencing. Single-end sequencing (75 base pair reads) was performed on a GA II analyzer (Illumina).

2.2.2: Sequence pre-processing, *de novo* transcriptome assembly and contig annotation

Figure 2.1 shows the pipeline used to assemble the *N. alata* pollen grain transcriptome. Raw sequence reads were trimmed of adapter sequences and further processed to remove low quality reads. Preliminary assembly of the filtered reads was done using Velvet (version 1.0.12; Zerbino and Birney, 2008) and the data passed to Oases (version 0.1.15; Schulz et al., 2012) to produce an assembly. Assemblies were made using a range of kmer lengths. To avoid the dependence of the assembly on the k-mer length parameter, a superassembly was prepared by merging the assemblies produced from the k-mers 17, 25, 31 and 47 (Schulz et al., 2012). Redundancy in the superassembly was removed using CAP3 (Huang and Madan, 1999).

Homology searches of the assembled pollen grain transcriptome were performed against protein sequences of *Arabidopsis* obtained from UniProtKB as at June 2012 and used to construct a BLAST dataset. Functional annotation (GO terms) was performed using BLAST2GO using the default annotation parameters (Conesa et al., 2005). Transcript abundance (in reads per kb of exon per million reads mapped, RPKM) for individual contig in the superassembly was determined using RSEM

(Li and Dewey, 2011). Assembly and automated annotation of the pollen grain transcriptome was performed by Andrew Cassin (Australian Centre for Plant Functional Genomics, Melbourne Australia).

2.2.3: Molecular biology

Total RNA was extracted from the indicated tissues of *N. alata*, treated with DNase I (Invitrogen) and reverse transcriptase followed by PCR (RT-PCR) performed using a Superscript III kit (Invitrogen) according to manufacturer's recommended protocol. Primers used for amplification of individual genes were designed using Geneious (Biomatters) and are listed in Appendix I. Amplified sequences were cloned into pGEM T-easy vector (Promega) and transformed into TOP10 *E. coli* cells (Invitrogen) by electroporation. Recombinant plasmids were identified, purified and sequenced by Australia Genome Research Facility (Melbourne). All recombinant DNA techniques were done as described by Sambrook et al., (2001).

2.2.4. Bioinformatic analysis

Geneious version 5.5.6 (Biomatters) was used to perform BLAST searches and align DNA and amino acid sequences. The default settings were used unless otherwise indicated. Intron boundaries were predicted with NetGene2 using the *Arabidopsis* settings (Hebsgaard et al., 1996) and phylogenetic trees were built using MEGA5.0 (Tamura et al., 2011). All DNA and protein sequences used in this study were obtained from Genbank except for selected *N. benthamiana* sequences which were obtained from the Sol genomics network (http://solgenomics.net). The 75 bp reads were mapped to selected *Nicotiana* sequences from Genbank using Bowtie version 2.0.6 (Langmead et al., 2009) and viewed using Tablet version 1.12.08.29 (Milne et al., 2010).

2.3: Results

2.3.1: Pollen transcriptome assembly and annotation

Figure 2.1 is an overview of the pipeline used to assemble, annotate and validate the pollen transcriptome. In total, 7,698,092 75-bp long reads were obtained from a single lane of sequencing. After trimming and filtering to remove low quality sequences, around 4.78 million reads were left, of which about 3.1 million were unique. *De novo* assembly was performed using Velvet and Oases and different assemblies were built by varying the K-mer length across a range from 15-57. Four of the assemblies (17-mer, 25-mer, 31-mer and 47-mer) were merged to form a superassembly and CAP3 was used to reduce redundancy. K-mer values define a trade-off between sensitivity and accuracy in *de novo* assemblies, with smaller K-mer assemblies providing a better representation of low abundance transcripts at the expense of a higher misassembly rate, and longer K-mer assemblies reconstructing transcripts more accurately at the expense of a poorer representation of all the

transcripts present (Schulz et al., 2012). Contigs less than 200 bp long were then discarded to produce a final pollen grain transcriptome of 6,800 contigs that were numbered sequentially starting at 0. Table 2.1 shows some summary statistics for the superassembly. The average contig length was 725 bp and the total contig length was ~4.9 million bp. The L50 of the superassembly was 1,052 bp.

Figure 2.2 shows the distribution of contig lengths in the pollen grain superassembly and its relationship to transcript abundance as measured by RPKM (Reads Per Kilobase of transcript per Million mapped reads), a measure of relative expression level between contigs. RPKM values for the 6,800 contigs ranged from less than 10^{-2} to 10^{5}: 240 contigs had an RPKM value below 10^{-2} (data not shown). Overall no discernible relationship between contig length and RPKM value was noted.

Contigs were annotated with gene ontology (GO) terms using the BLAST2GO suite. Contigs were first used to search the *Arabidopsis* genome using a BLASTx e-value threshold of $1e^{-5}$ resulting in 3,074 contigs (45.2% of the superassembly) returning a positive BLAST hit. GO terms for the top BLAST hits were retrieved and used as the annotation for the contig. Figure 2.3 shows the distribution of GO terms for three functional categories for the pollen grain transcriptome and *Arabidopsis*. The pollen transcriptome covered functional categories as broadly as all *Arabidopsis* genes, with plasma membrane proteins, proteins with protein-binding activities, and proteins involved in pollination all seeming to be over-represented in the pollen grain transcriptome compared to *Arabidopsis* although the validity of this conclusion was not tested statistically.

2.3.2: Validation of transcriptome assembly

Validation of the pollen grain transcriptome was done using bioinformatic and molecular approaches (Figure 2.1). In the bioinformatics approach, the sequences of 57 genes (17 from *N. alata* and 40 from *N. tabacum*) known to be expressed in *Nicotiana* pollen were compared using BLASTn to contigs in the assembled transcriptome (Table 2.2). Of the 57 genes, six (three *N. alata* and three *N. tabacum* genes) did not return a hit in the superassembly with an e-value less than $1e^{-1}$. Of the 51 genes that did return a hit, 39 had a >95% pairwise identity over most of their length to at least one contig in the superassembly. One gene, *NaGSL1*, which encodes a putative callose synthase and has a messenger RNA of approximately 6.2 kb, had perfect matches to three different non-overlapping contigs. The lowest pairwise identities were the approximately 70% matches for five of the ten *N. alata DD* genes (*DD5*, *6*, *8-10*). Four of these matches (*DD5*, *6*, *8* and *9*) were to contig 452. Notably no match was found for *DD7*. RPKM values for the contigs in Table 2.3, ranging from 5 to 25,457, represent expression values from over four orders of magnitude and, consistent with the conclusion from Figure 2.2, no obvious relationship existed between contig size and RPKM value. Taken

together, the data in Table 2.2 indicate that the transcriptome, although produced with only a single lane of sequence data, contained representatives of most (80% or more) of the transcripts in *N. alata* pollen grains.

To understand why some known genes did not match a contig in the pollen superassembly, Bowtie was used to map reads to the Genbank sequence. The Bowtie output was read using Tablet and the results summarised in Table 2.3.

No reads mapped to the cysteine protease gene *CysP*, the sucrose synthase gene *SuSy* or the RING domain protein gene *SBP1*; and 9, 35 and 19 reads respectively mapped to the cellulose synthase gene *CESA1*, the pyruvate decarboxylase gene *PDC2* and the F-box protein gene *DD7*. Reads in these latter cases were mostly scattered across the sequence of the gene, with the longest stretch of overlapping reads for *CESA1*, *PDC2* and *DD7* being 96 bp, 273 bp and 163 bp, respectively. Thus, a combination of an absence of reads and the removal of contigs less than 200 bp in length as part of the pipeline (Figure 2.1) accounted for the absence of five of the six known genes from the superassembly. Reasons for the lack of a *PDC2* contig in the superassembly were not further investigated.

The absence of a *SBP1* contig in the transcriptome was further investigated, as the encoded protein is implicated in GSI and expression of the *Nicotiana* ortholog *NaSBP1* in pollen grain has been reported previously (Hua and Kao, 2006; Lee et al., 2008). Primers designed based on the *NaSBP1* sequence (accession number EU591514) were used to amplify a product from pollen cDNA and a search of GenBank with the sequence of this product found a 96.8% match at the amino acid level with *N. alata SBP1*. Thus, while not in the transcriptome, consistent with previous reports *SBP1* transcripts were present in pollen grain RNA.

Molecular validation of the transcriptome was done by RT-PCR. Primers designed to 45 selected contigs were used to amplify products from pollen grain cDNA (Appendix I). Contigs ranged in size from 602 bp to 3,982 bp and had RPKM values between 0 and 6,769 (Table 2.4 and Appendix II). As expected, transcripts for all contigs were detected in pollen grain cDNA and there was a good but by no means perfect concordance between a qualitative assessment of RT-PCR band intensity and the contig's RPKM value. For instance, transcripts for contig 6173 were more abundant by both measures than those for contig 6440. There were, however, many exceptions to this (e.g., contig 5066 had a high RPKM value but low expression as measured by RT-PCR and the reverse was true for contig 599). This may reflect inadequate optimisation of the PCR, but the more likely reason is the method used to quantify contig abundance from mapped reads.

An example of this is contig 2401, which was validated by RT-PCR but had an RPKM value of 0 (Table 2.4 and Appendix II). However, using Bowtie, 622 reads for contig 2401 were found that covered its entire length (~1.7 kb; Table 2.3). RSEM, the program used to calculate RPKM, aligns reads back to transcripts with each read being used only once. This results in a skewed estimation of abundance towards a more highly supported contig when multiple isoforms of the same transcript exist in the superassembly. This can mean that very few or no reads (and hence a low RPKM value) can be mapped to other isoforms even though these isoforms are well represented in the transcriptome.

Sequencing of the pollen cDNA products confirmed that the target transcript had been amplified for all 45 contigs (Table 2.4 and Appendix II). Pairwise identities between contigs and PCR products were mostly (36 out of 45) above 99% and in only four cases was sequence identity <95%. In two of these cases (contigs 4861 and 4913) it was due the presence in the PCR product of an insert that wasn't in the contig (Table 2.4). Bowtie results indicate contigs 4861 and 4913 were well-supported by numerous reads (Figure 2.4).

The PCR product of contig 4861 (4861p) had a 140 bp insert that was not in the contig itself (Figure 2.4 and Table 2.4). The best GenBank match for 4861 and 4861p (covering over 95% of each sequence and with an E value of 0) is a genomic sequence (accession no. M80492) that contains the last 9 of the 21 exons of *PMA2*, a *N. plumbaginifolia* gene for a plasma-membrane H^+ ATPase (Perez et al., 1992). *PMA2* is widely expressed in *N. plumbaginifolia* tissues, including floral tissues (Arango et al., 2003), so the presence of its ortholog in the *N. alata* pollen grain transcriptome is not unexpected. Remarkably, however, the 4861 sequence matches the last 4 introns of *PMA2*, the entire 3' untranslated region, and extends beyond the point 261 bp downstream of the stop codon where the *PMA2* cDNA is polyadenylated (Figure 2.4; Boutry et al 1989). 4861p is nested within 4861, with the difference between the two being a deletion in 4861 of 140 bp in intron 17 (Figure 2.4). There are stop codons interrupting all six reading frames in both sequences and RT-PCR didn't identify a fully processed version of the *PMA2* transcript because the forward primer used to amplify 4861p was based on the sequence of intron 17 (Figure 2.4). 4861 and 4861p could result from either alternative processing of *PMA2* or be due to genomic DNA contamination in the RNA-Seq analysis and the cDNA used in RT-PCR. However, given the steps taken to remove genomic DNA in both these analyses, the former seems the more likely possibility. Consistent with this, no amplification was seen in RT-PCR samples in which the reverse transcriptase step had been omitted (data not shown).

Contig 4913 is a 921 bp chimeric sequence made up of parts from two separate transcripts (Figure 2.5A and Table 2.4): the first 375 bp are 98% identical to an *S-adenosyl-L-methionine synthetase* (*SAMS*) cDNA from *N. tabacum* and the remainder is 91% identical to a *Solanum lycopersicum* fruit-

derived cDNA (Aoki et al., 2010). The PCR product of contig 4913 (4913p) has a 46 bp insert relative to the contig and only contains sequences from the pollen-grain ortholog of the tomato cDNA (Figure 2.5A). A BLASTn search of the *N. benthamiana* genome identified a genomic region that aligns with contig 4913, contains the 46 bp insert found in 4913p, and has an open reading frame that matches the tomato cDNA (Figure 2.5B). The 46 bp insert in 4913p lies within what appears to be an intron, as the genomic sequence of this region is predicted to be flanked by donor and acceptor splice sites. As both 4913 and 4913p are derived from cDNA, alternative processing of an intron within the 5' untranslated leader of this gene is a plausible explanation of the sequence difference between them.

Expression analysis of the 45 contigs was done using leaf, style, petal and 7 day old seedling cDNA. In summary, 22 of the 45 contigs (49%) were also expressed in leaf, 15 (33%) were also expressed in petal, 37 (82%) were also expressed in style and 13 (29%) were also expressed in 7 day old seedling. Four of the contigs (8.9%) were expressed in all tissues and 5 (11%) were solely expressed in pollen (Table 2.4, Appendix II). Although generally the products amplified from the other cDNA sources were the same size as the pollen cDNA product, the sequences were not checked and the possibility exists that some of the products were derived from the transcripts of related genes.

2.3.3: Identification and characterisation of RNase-based SI related transcripts in the pollen transcriptome

To identify RNase-based SI related transcripts in the superassembly, a list of target genes was prepared based on the *Petunia* sequences that have been implicated in this process (Table 2.5).

The canonical SCF (Skp1-Cullin1-F-box) is a multi-subunit E3 ligase that has as its components an F-box protein, Skp1, Cullin1 and Rbx1, a RING-H2 protein (Petroski and Deshaies, 2005). However, to date studies of the SLF-containing complex in *Petunia* have pointed to a range of possible components. Hua and Kao (2006), for instance, described a novel E3 ligase complex composed SLF, and Cullin1, and another potential E3 ligase, the S-RNase binding protein1 (SBP1), that, because it has a RING domain, could play the combined roles of Skp1 and Rbx1. By contrast, Zhao et al., (2010) described a more conventional SCF E3 complex in *Petunia* composed of SLF and Cullin1, and a novel Skp1 like protein called SSK1 that connects SLF to Cullin1. These authors did not identify an Rbx1 component for the complex. SSK1 has the same overall sequence structure as Skp1 and differs from it by the presence of an extra 7-9 amino acids at the COOH terminal end (Huang et al., 2006).

Contigs encoding *Skp1* and *Rbx1* were identified in the *N. alata* pollen transcriptome but there were no contigs for *SSK1* or, as previously noted, *SBP1* (Table 2.5). The bridging protein Skp1 links the F-

box protein to Cullin1 and a single contig in the transcriptome (6186) encoded a full-length Skp1 protein; another contig (3463) encoded part of a different Skp1 protein. Amino acid identity between the contig 6186 protein and *P. inflata* Skps 1, 2 and 3 was 83.4-88%, with the best match being to Skp1. As Skps and SSK1s are highly similar and are distinguished by their COOH terminal ends an alignment was made of the contig 6186 and 3463 proteins and representative *Arabidopsis* and *P. inflata* Skps and SSK1s from *P. hybrida* and *Antirrhinum hispanicum* (Figure 2.6). The alignment clearly shows that the 3463 and 6186 proteins have the COOH terminal ends typical of Skp proteins. Searching the *N. benthamiana* genome with *P. hybrida* SSK1 identified a probable SSK1 ortholog in this self-compatible species (Figure 2.6), a finding that suggests the likely presence of SSK1 orthologs in *N. alata* as well.

Figure 2.7 shows an alignment of the protein encoded by contig 6029 and *P. inflata* Rbx1. The contig 6029 sequence covers the COOH-terminal portion of the protein and includes the residues proposed to bind to Cullin1 as well as the putative Zn binding site (Wei and Sun, 2010). The sequence of the NH_3-terminal end of the protein is missing. The region of contig 6029 encoding Rbx1 comprises only about 40% of the total length (867 bp), suggesting that this contig is a chimera.

Two *Cullin1* contigs (3497 and 4884) were identified based on 91% pairwise nucleotide identity to *P. inflata Cullin1G* (Table 2.5). Pairwise nucleotide identity of both contigs to *N. tabacum Cullin1A* was even higher (98%). Contig 4884 appears to be a near full-length cDNA that matches *N. tabacum Cullin1A* for most of its length and encodes a protein that contains the residues implicated in Skp1 and Rbx1 binding (Zheng et al., 2002) (Figure 2.8). Contig 3497 is a partial *Cullin1A* cDNA that contains other sequences as well. The partial Cullin1 encoded by contig 3497 is identical to the 4884 protein (Figure 2.8).

RT-PCR performed using primers specific to the *Cullin1, Rbx1* and *Skp1* contigs and to *N. alata SBP1* found each gene was expressed in tissues other than pollen grains with all genes except *Skp1* expressed throughout the plant (Table 2.6). *Skp1* was not detectably expressed in seedlings. PCR products from pollen cDNA were sequenced and in each case the target transcript was the one that was amplified. Products from other tissues were not sequenced.

Kubo et al., (2010) describe *pollen S* in *Petunia* as comprised of members of at least six subgroups of *SLFs* that they named *SLF1-SLF6*. Sequence identities between alleles of an *SLF* subgroup range from 70% to 99% and identities between the different *SLF* subgroups are only about 50%. Representatives of the 18 known *Petunia SLF* subgroups were used to search for related sequences in the *Nicotiana*

pollen grain transcriptome and the results are shown in Table 2.5. Table 2.2 reports the results of searches with the *N. alata* DD sequences.

The best match identified in searches with *Petunia SLF1* and *SLF2* was to contig 3684, which is identical to *DD4* (Table 2.5). Similarly, the best match in searches with *SLF5-11*, and *SLF14-18* was contig 452 (Table 2.5), which previously had been identified in searches with *DD5*, *DD6*, *DD8* and *DD9* (Table 2.2). Contig 452 showed similar pairwise identities to each of these sequences, with *SLF6* being the lowest (~67%) and *DD6*, *DD8* and *DD9* the highest (~71%, Table 2.2). By contrast, searches with *SLF3/13* and *SLF4/12* identified contigs that had not previously been identified in searches with *DD* sequences: *SLF3/13* identified contig 2031 and *SLF4/12* contig 5258 (Table 2.5). Other contigs identified by searches with *DDs* were 5494 (99.5% identical to *DD1*), 1945 (97.4% identical to *DD2*), 1357 (98% identical to *DD3*) and 4791 (69.4% identical to *DD10,* Table 2.2). In summary, searches for *SLFs* in the transcriptome identified a total of eight contigs: 452, 1357, 1945, 2031, 3684, 4791, 5258 and 5494. Four of these were either identical or nearly identical to a known *DD* (*DD1* and 5494, *DD2* and 1945, *DD3* and 1357 and *DD4* and 3684). Contigs 1357, 1945 and 5494 presumably represent either the S_2 or S_3 allelic variant of the relevant gene, as the *DD1-DD3* sequences used in the search were from the S_1 allele (Table 2.2). Contig 3684 presumably represents the S_3 allelic variant as the *DD4* sequence used was from the S_2 allele. The remaining four contigs (452, 2031, 4791 and 5258) potentially represent novel *SLF* sequences.

Figure 2.9 shows an amino acid alignment of the ten DDs and four possibly novel SLFs. None of the contigs encodes a full length SLF: contig 452 encodes the longest sequence (314 amino acids) and contig 2031 the shortest (64 amino acids). Only contig 452 can be described as encoding an F-box protein, as only this sequence has the relevant motif at its NH_3-terminal end. The conceptual protein of contig 4791 includes part of the F-box motif and thus is probably one as well. However, extensive blocks of amino acid identity between the two remaining contigs and the DDs suggest that all four proteins are F-box proteins.

The conceptual proteins of contigs 452 and 4791 are 96% identical over 294 amino acids and so potentially represent the S_2 and S_3 allelic variants of the same SLF/DD protein. Although the region of overlap between the 2031 and 5258 conceptual proteins is small (approx. 40 amino acids), it is sufficiently long to suggest they are products of separate *SLF* genes. Thus the four novel *SLF* sequences appear to represent three distinct SLFs.

A distance tree was built using the proteins encoded by two longest contigs 452 and 4791, the *N. alata* DDs, representatives of each of the six classes of *Petunia* SLF described in Kubo et al., (2010)

and, as the outgroup, S_1-SLF from *Antirrhinum hispanicum* (Figure 2.10). The proteins encoded by contigs 2031 and 5258 were not included as their sequences were too short for phylogenetic analysis. All of the DDs except DD7 were contained in a large polytomy that also included the *Petunia* SLFs and the contig 452 and 4791 proteins. In similar trees, Wheeler and Newbigin (2007) and Newbigin et al., (2008) noted DD7 was at the base of a cluster of Solanaceae SLF sequences, suggesting it is the most divergent of these sequences. Within the polytomy, the *P. hybrida* SLF class 4, 5 and 6 representatives formed a well-supported clade with long terminal branches: DD5, 8, 9 and 10 were also in a well-supported clade as previously noted by Wheeler and Newbigin (2007) and Newbigin et al., (2008), with DD1, 3 and 4 clustering with *P. inflata* SLF1 S_1 as they do here. The contig 452 and 4791 proteins are in a separate cluster with short terminal branches, consistent with the suggestion these sequences represent allelic variants of the same SLF. All other SLFs are singletons attached to the polytomy by long terminal branches.

RT-PCR performed using primers specific for contigs 452, 1357, 1945, 3684, 4791 and 5494 found expression of all contigs except 1357 in pollen (Table 2.6). No expression of contig 1357 was detected in any tissue. Only expression of contig 1945 was restricted to pollen, as the other amplifiable contigs were expressed in a range of tissues other than pollen grains. For example, contig 3684, identical to *DD4*, was also expressed in style and petal as previously reported (Wheeler and Newbigin 2007). Contigs 452 and 4791 showed the same pattern of expression, consistent with their presumed allelism. No expression in seedlings was detected. PCR products from pollen cDNA were sequenced to confirm that the target transcript had been amplified.

2.4: Discussion

In this chapter RNA-Seq and *de novo* assembly were used to prepare a searchable transcriptome of *N. alata* pollen grains in which was found two novel *DD*/*SLF* transcripts as well as transcripts of other genes suggested to encode components of the SLF-containing E3 ligase complex. As none of these sequences is among the over 430,000 *Nicotiana* ESTs currently available (as of July 2014) in GenBank, or the over 43,000 unigenes present on the tobacco microarray (Edwards et al., 2010), even though assembled from a single lane of sequence data the transcriptome appears to contain many previously undescribed cDNAs. Figure 2.3 also indicated the transcriptome has been broadly sampled, as it contained genes for a range of different functions. However, it was equally evident that a number of known transcripts were missing from the transcriptome, for example no contigs were found for six of the 10 *DD* genes (Table 2.2). Reasons why some contigs may have been missing have already been mentioned and deeper sequencing would obviously expand the existing coverage. But an expanded coverage could potentially also be achieved by pooling the results of several

different approaches to *de novo* assembly in the bioinformatic pipeline. In this chapter the superassembly was produced using Velvet and Oases and *DDs 5-10* were missing. A separate transcriptome, produced from the same reads using the Trinity assembler (Grabherr et al., 2011), was also missing six *DDs* but in this case it was *DD1*, *DDs 5-8* and *DD10* that were missing (Lampugnani et al., 2013). Presumably sequence similarities between *DD* transcripts coupled with the low transcript abundances prevented correct assembly of some contigs in a method-dependent manner. Interestingly Williams et al., (2014a) also found a number of *SLF* contigs were missing from their *P. inflata* pollen grain transcriptomes, even though these were assembled using a much greater volume of sequence than was used here. Indeed, as 99% of the unigenes in each *P. inflata* assembly could be produced using 25% or less of the total reads, additional sequencing may only marginally improve coverage of low abundance transcripts. The use of paired-end reads and inclusion of sequences produced using different technologies (e.g., inclusion of some longer 454 sequences) should also result in a high quality assembly (Cahais et al., 2012).

Any *de novo* assembly of RNA-Seq data aims to produce truly trustable contigs. Although an ideal assembly would contain a single contig per expressed gene of the target genome, the absence of a reference genome makes it difficult to know the extent to which a transcriptome meets this ideal. *De novo* transcriptome assembly is known to produce a substantial fraction of erroneous predictions of various sorts including chimeras (a single contig for two or more genes) and fragmented transcripts (multiple contigs for each expressed gene). Table 2.4 shows that both chimeric and fragmented contigs were present in the superassembly. Chimeras are produced because assemblers such as Velvet use a de Bruijn graph approach to search for overlaps between reads (Zerbino and Birney, 2008). The presence of repeat sequences in the data means false overlaps are possible with a consequence that two unrelated sequences either side of the repeat can be joined together into a contig. Even though Table 2.4 suggests that only around 2% of contigs were chimeric (one chimeric contig among 45 sampled contigs) the actual percentage may have been higher as one of the eight contigs in Table 2.6 was a chimera. Although chimeric contigs are problematic for automated gene annotation approaches, the sequence they contain is still usable for other purposes.

If chimeras were one source of difference between assembled contigs and sequences in GenBank, then the presence of mRNA splicing variants was another. Previous RNA-Seq studies have shown the production of differentially spliced mRNAs to be highly prevalent in plants, seen in over 60% of intron-containing genes (Filichkin et al., 2010; Reddy et al., 2013). The predominant type of alternative splicing seen in *Arabidopsis* and other land plants is intron retention, where an intron is not spliced from the message (Filichkin et al., 2010; Marquez et al., 2012). In this chapter, contig

4861, an incomplete transcript encoding a plasma-membrane H^+ ATPase, retained four of the introns present in the *N. plumbaginifolia PMA2* gene (Figure 2.4). Although the presence of stop codons in the retained introns will affect translatability and make the 4861 transcript a candidate for degradation via the nonsense-mediated pathway (Kervestin and Jacobson, 2012), it is also possible that this is an example of a stored pre-mRNA synthesised during the later stages of male gametophyte development in the anther. Previous work on the so-called 'late' pollen transcripts has suggested that many are stored in mature pollen grains and only translated during germination and tube growth (e.g., Mascarenhas et al., 1984; Twell et al., 1989). Intron retention is one possible mechanism by which the translational repression of stored mRNAs in pollen grains is regulated. Consistent with this hypothesis, analysis of RNA-Seq data from the fern *Marsilea vestita* has revealed that many of intron-retaining transcripts in this species encode proteins that are translationally repressed during gamete development and are only translated following the regulated removal of the retained introns at specific times during development (Boothby et al., 2013). The contig 4861/*PMA2* transcript provides a hint that this may also be occurring during *N. alata* pollen development although demonstrating this experimentally lay outside the scope of the thesis.

Although a canonical E3 ubiquitin ligase consists of Skp1, Cullin1, Rbx1 and F-box proteins, various models of the complex associated with S-RNase based GSI have been proposed. Huang et al., (2006), for instance, identified a variant Skp1-like protein in *Antirrhinum* (Plantaginaceae) called SSK1 (SLF-interacting SKP1-like1) that could interact with both SLF and a Cullin1-like protein. Hua and Kao (2006) however reported that Skp1 was not a component of the SLF–containing complex in *P. inflata* and instead proposed an unorthodox E3-like complex composed of SLF, Cullin1 and the RING domain protein SBP1. Subsequently Zhao et al., (2010) identified the *Petunia* ortholog of *Antirrhinum* SSK1 and showed it was specifically expressed in pollen and acted as an adaptor protein in a modified Skp1-Cullin1-F-box complex. Moreover, reducing SSK1 expression in *Petunia* also reduced and even eliminated the ability of pollen grains to fertilise a compatibly pollinated *Petunia* flower. This result is consistent with the prevailing view that the SLF/pollen S complex acts as an S-RNase inhibitor (Golz et al., 2001), as pollen tubes unable to assemble this complex will not be able to inactivate the cytotoxic S-RNases and hence cannot grow through a compatible style expressing S-RNases.

Studies of a reproductive barrier called unilateral incompatibility (UI) have also implicated E3 ubiquitin ligases in pollen rejection. Unilateral incompatibility in the Solanaceae is closely associated with SI, the main difference being that UI is an inter-specific barrier that prevents related species from hybridising (most commonly when the style of a self-incompatible species non-specifically rejects pollen from a related self-compatible species (Lewis and Crowe, 1958) whereas SI is an intra-

specific barrier and pollen rejection is selective. Recently, Li and Chetelat (2010) showed that Cullin1 was an essential factor in the S-RNase-dependent UI system of *Solanum* and suggested that the intra- and inter-specific pollen rejection pathways share many components in common. This suggestion was subsequently confirmed when it was demonstrated that pollen from the self-incompatible wild tomato *Solanum arcanum*, modified so that it expressed less Cullin1, was rejected non-specifically by the compatible styles of self-incompatible sib plants, but accepted by the compatible styles of plants from a self-compatible accession of the same species that does not express an enzymatically active S-RNase (Li and Chetelat, 2014). As this chapter was being written, Li and Chetelat (2014) and Entani et al., (2014) showed using co-immunoprecipitation followed by proteomic analysis that in *Petunia*, SLF forms protein complex together with Cullin1 (specifically the protein ortholog of the *Solanum* Cullin1 associated with UI and SI), SSK1 and Rbx1. The RING protein SBP1 was not present at detectable levels in this complex.

Even though the *N. alata* transcriptome contained orthologs of *Cullin1* and *Rbx1*, no *SSK1* contig was identified: an *SSK1* genomic sequence (and corresponding cDNA) was, however, subsequently found in *N. benthamiana* (Figure 2.6 and data not shown). SSK1s are pollen-specific Skp1-like proteins that were isolated based on their ability to interact with Cullin1 and SLFs in species with S-RNase-based SI in the Plantaginaceae (*A. hispanicum*), Solanaceae (*Petunia*) and Rosaceae (*Prunus* and *Pyrus*)(Huang et al., 2006; Zhao et al., 2010; Matsumoto et al., 2012; Xu et al., 2013; Entani et al., 2014; Li et al., 2014; Yuan et al., 2014). In phylogenetic analyses of plant Skp1-like proteins, SSK1 proteins cluster together as a distinct lineage (Xu et al., 2013). Since SSK1s appear to be conserved pollen factors so far found only in species with S-RNase-based SI but absent from species lacking this reproductive barrier (such as *Arabidopsis* and rice), some authors have suggested that the SSK1s and the S-RNases must therefore share the same evolutionary origin (Xu et al., 2013). However, this apparent shared origin could also be a consequence of limited taxonomic sampling that is heavily biased towards species with S-RNase-based SI. The presence of an apparently functional *SSK1* gene and cDNA in the self-compatible plant *N. benthamiana* provides an example of SSK1 being expressed in a species that lacks S-RNases (Golz et al., 1998), pointing to functions for SSK1 outside of those associated with SI. Profiling *SSK1* expression in this and other self-compatible species from the Solanaceae would be one way of testing the possibility the encoded protein has a broader range of roles than have so far been allocated to it.

Six different *SLFs* were thought to encode pollen S determinants in *Petunia* (Kubo et al., 2010), a number that has recently risen to 18 (Williams et al., 2014a, b; Kubo et al., 2015). Equally large numbers of *SLF* genes (called *SFBBs*) have been detected in the Rosaceae: 20 at the *Malus S* locus

(Minamikawa et al., 2010) and 16 at the *Pyrus S* locus (De Franceschi et al., 2011a). Most of these genes are considered to encode pollen specificity determinants (Sassa et al., 2007). By contrast a single F-box protein gene (*SFB*) at the *Prunus S* locus appears to encode pollen S in this species (Ushijima et al., 2004; Sonneveld et al., 2005). Wheeler and Newbigin (2007) identified ten pollen-expressed *SLF-like* sequences (*DD1-10*) at the *N. alata S* locus but suggested there were likely to be more that had escaped detection. Here potentially three distinct SLFs have been identified, with contigs 452 and 4791 representing alleles of one novel SLF and contigs 2031 and 5258 potentially representing two additional SLFs.

In a phylogenetic analysis of available SLF sequences Williams et al., (2014a) defined a monophyletic 'Solanaceae SLF clade' comprised of 17 *Petunia* sequences and eight of nine *N. alata* DD sequences (DD10 was omitted from the analysis as its sequence is missing the initiator Met codon and 5' UTR). Not included in this clade was DD7, which Williams et al., (2014a) considered belonged to a separate clade of SLF-like sequences that was intermediate between the Solanaceae SLF clade and the *Antirrhinum* SLFs. Overall the topology of the phylogeny produced by Williams et al., (2014a) is concordant with the tree shown in Figure 2.10, suggesting that the SLF represented by contigs 452/4791 will also be part of the Solanaceae SLF clade. As several of the *Petunia* members of this clade alter the pollination phenotype when expressed in a transgenic plant; and as the other *Nicotiana* and all the listed *Petunia* genes are linked to the *S* locus in their respective species, Williams et al., (2014a) concluded that all members of the Solanaceae SLF clade likely encode pollen specificity determinants. While linkage of the 452/4791 gene to the *Nicotiana S* locus still needs to be determined, previous results suggest linkage is highly likely, which would imply that similar numbers of *SLF* genes at the *Nicotiana* and *Petunia S* loci collectively regulate pollen specificity as described by the collaborative non-self-recognition model (Kubo et al., 2010; Kubo et al., 2015).

By performing next-generation sequencing, this study revealed more *DDs* are present in *N. alata* and the number is likely the same other SI species such as *Petunia*. In addition, a list of SI related genes that comprises the E3 ligase complex is also present in *N. alata* suggest a smilar complex also function in SI in *N. alata*.

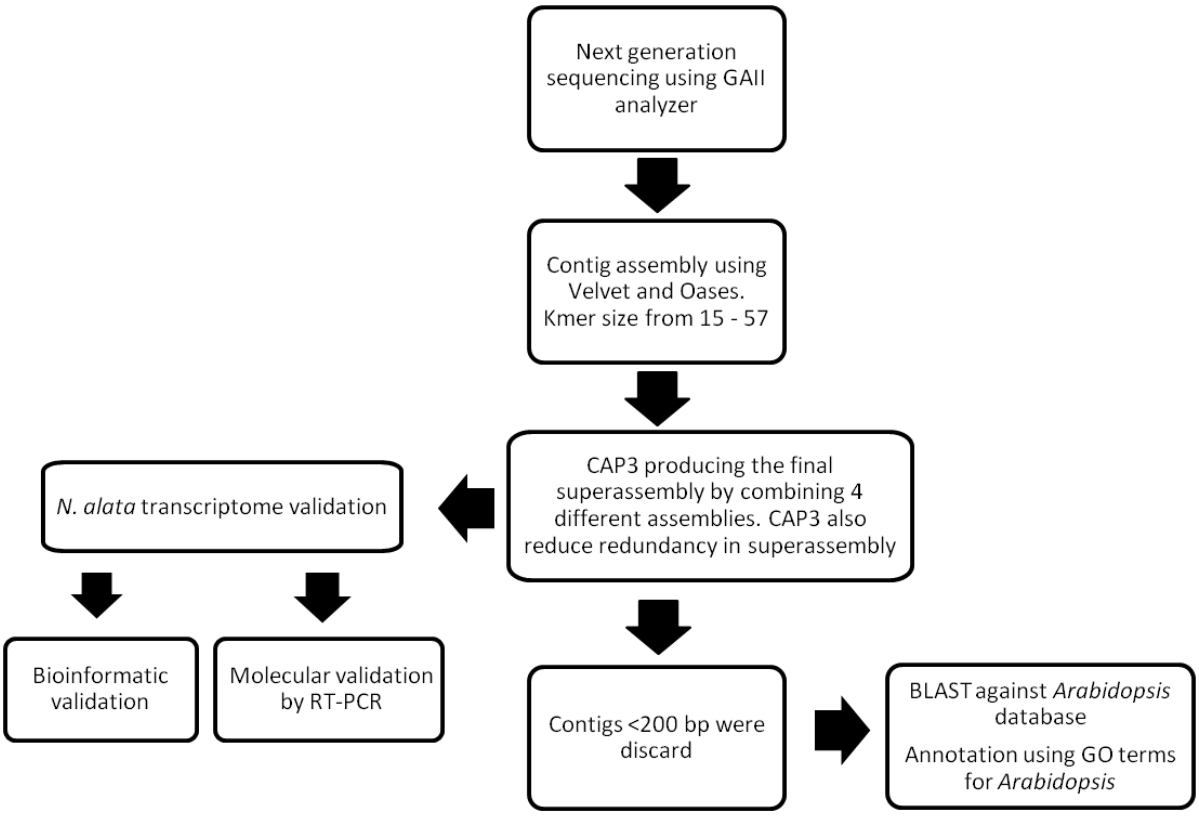


Figure 2.1: Pipeline showing the steps taken to assemble, annotate and validate the *N. alata* pollen grain transcriptome.

Table 2.1: Summary statistics of the *Nicotiana alata* pollen grain transcriptome.

Total number of 75 base pair reads	7,698,092
Number of reads after filtering	4,872,196
Total contig length	4,932,212 bp
Average contig length	725 bp
L50[1.]	1,052 bp
N50[2.]	1,468

[1.] L50 is the length of the contig separating the top half (N50) of the assembled transcriptome from the remainder of smaller contigs, if the sequences are ordered by size from shortest to longest.

[2.] N50 is the number of contigs representing the top half of the transcriptome, if the sequences are ordered by size from shortest to longest.

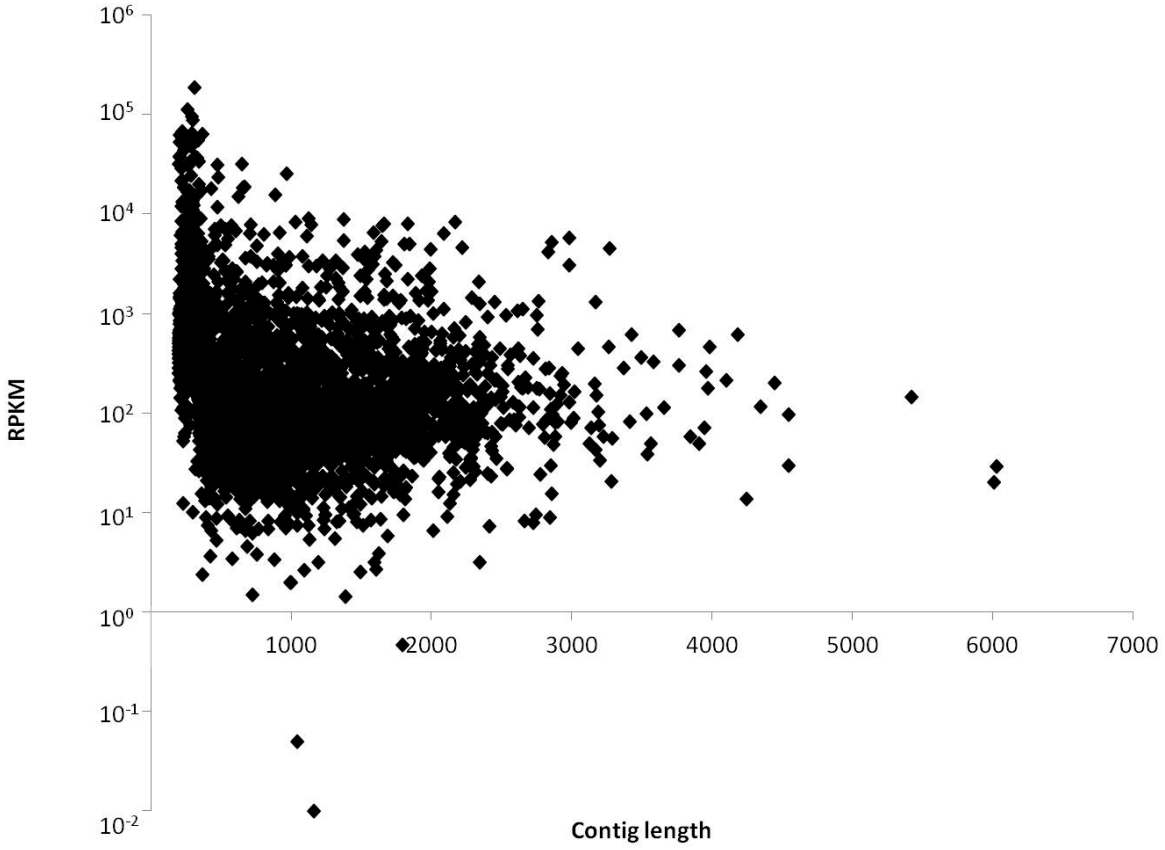


Figure 2.2: Distribution of the 6,800 contigs in the *N. alata* pollen grain transcriptome by contig length and RPKM.

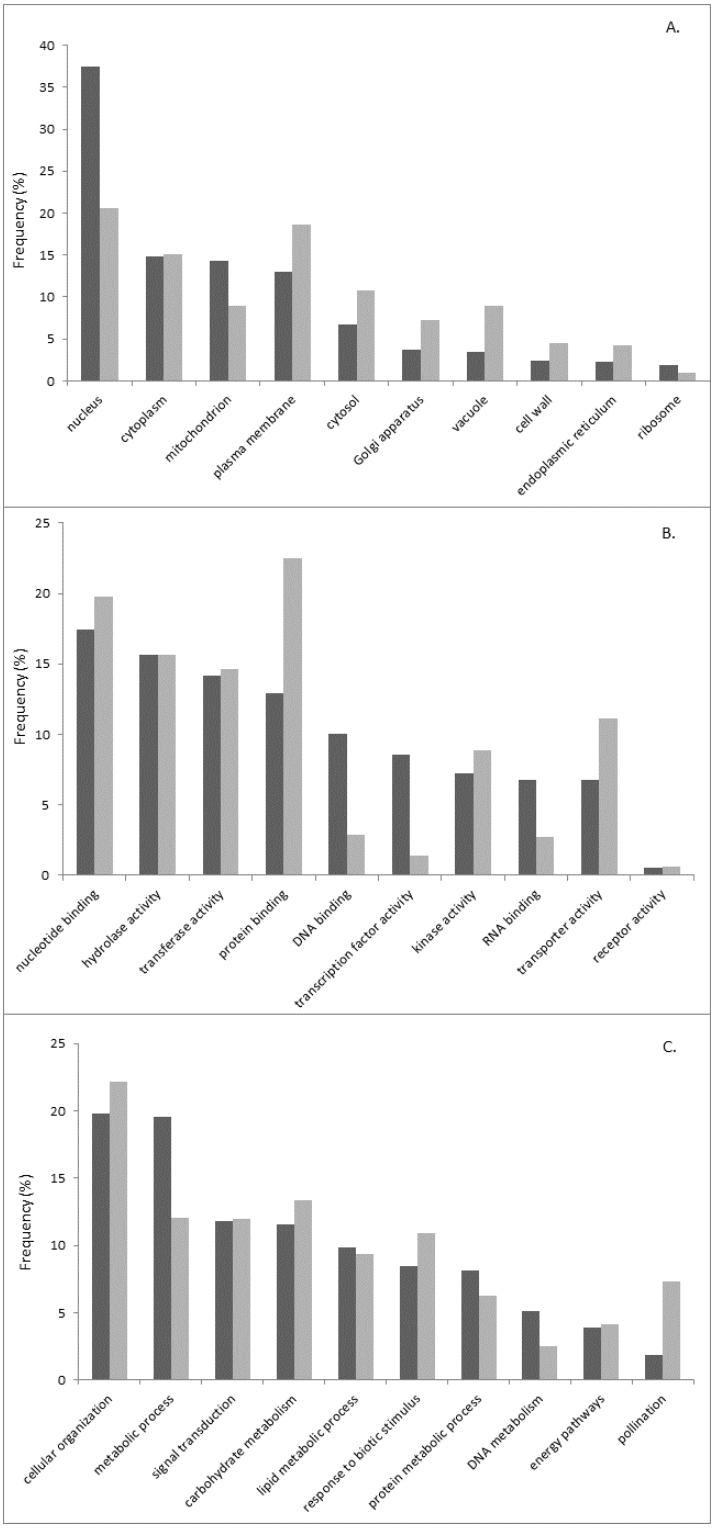


Figure 2.3: Profile of GO terms for the *N. alata* pollen grain transcriptome (grey bars) compared to GO terms for all *Arabidopsis* genes (black bars). *Arabidopsis* GO annotation terms were retrieved from TAIR. A. Cellular component GO terms; B. Molecular function GO terms; and C. Biological process GO terms.

Table 2.2: List of 57 known pollen-expressed genes from tobacco (*N. alata* and *N. tabacum*) used for validation of the *N. alata* pollen grain transcriptome. The indicated sequence was used to query the 6,800 contigs in the pollen grain transcriptome. Where relevant the sequence *S* haplotype is shown. The top BLAST hit (e-value cut off at $1e^{-1}$) for each gene is shown along with the contig's length and pairwise identity. A dash (-) indicates no hit was found in the assembly.

Gene name	Accession no.	Superassembly contig number	Contig size (bp)	Pairwise identity (%)	Contig RPKM	Comments
N. alata						
DD1 (S_1)	EF420251.1	5494	448	99.5	140	F-box protein
DD2 (S_1)	EF420252.1	1945	282	97.4	496	"
DD3 (S_1)	EF420253.1	1357	304	98	1,598	"
DD4 (S_2)	EF420254.1	3684	841	100	71	"
DD5 (S_2)	EF420255.1	452	1,066	69.3	33	"
DD6 (S_2)	EF420256.1	452	1,066	71.3	33	"
DD7 (S_2)	EF420257.1	-	-	-	-	"
DD8 (S_2)	EF420258.1	452	1,066	71.2	33	"
DD9 (S_2)	EF420259.1	452	1,066	70.8	33	"
DD10 (S_6)	EF420260.1	4791	893	69.4	86	"
GSL1	AF304372.2	671	4,183	100	606	Putative callose synthase
		4456	2,025	100	128	"
		4813	1,933	100	524	"
CSLD1	AF304375.1	2402	1,381	99.9	126	Cellulose synthase D-like
CESA1	AF304374.1	-	-	-	-	Cellulose synthase
P18	AJ004957.1	6078	958	99.8	3,038	Hypothetical protein
SBP1	EU591514.1	-	-	-	-	RING domain protein
MIP	U20490.1	6135	1,698	94.2	938	Probable aquaporin
PCCP	EU591515.1	5282	1,182	98.8	2,993	C2 domain containing protein
N. tabacum						
ADF1	AY081941.1	4527	713	97.6	1,467	Actin-depolymerizing factor
ADF2	AY081942.1	6331	943	97.3	766	"
RHOGD2	DQ416769.1	612	835	95.2	524	Rho GDP-dissociation inhibitor
Nict1	AB035706.1	6210	674	95.4	2,059	Calcium binding protein
Nict2	AB035705.1	5179	336	97.3	393	"
CysP	EU429306.1	-	-	-	-	Cysteine protease
NTP805	AY366400.1	6208	1,991	97.1	4,433	Pollen-specific protein
PNTP302	AY366399.1	5864	1,140	94.9	7,773	"
NTP303	X61146.1	6087	1,373	95.8	8,832	"
PLIM1	AF184885.1	6128	752	98.3	4,732	LIM domain-containing protein
PLIM2	AF116851.1	4618	3,423	96.9	614	"
AscOx	X96932.1	3211	514	73.3	3,406	Ascorbate oxidase
PLC3	EF043044.1	5034	1,943	97.3	268	Phospholipase C
SuSy	EU148354.1	-	-	-	-	Sucrose synthase
PRK1	AF246964.1	3014	2,313	97	253	Receptor-like protein kinase
PRK2	AF246967.1	4451	3,982	97.2	460	"
PRK4	AF252414.1	6187	2,637	94.9	180	"
GNL1	EF520731.1	4712	3,943	97.7	71	GNOM-like protein
eIF-4A	X79005.1	6594	1,861	80	51	Translation initiation factor
NPG1	X71020.1	5109	1,640	97.2	5,259	Polygalacturonase
PPME	AY772945.1	2881	965	96.2	25,457	Pectin esterase
NHA1	AY383599.2	6081	2,983	96.5	5,756	H^+ ATPase
AldH 2A	Y09876.1	2324	213	97.2	535	Aldehyde dehydrogenase
PDC2	X81855.1	-	-	-	-	Pyruvate decarboxylase
NTK-1	X77763.1	96	1,128	93.2	5	Shaggy-like kinase
PL	X67159.1	6146	1,593	96.6	3,725	Pectate lyase
ROP1	AJ222545.2	486	1,330	96.5	1,139	Rop subfamily GTPase
NSK 91	AJ224163.1	4455	1,263	96.7	222	Shaggy-like kinase
NSK 59	AJ002315.1	2531	2,072	98.1	223	"
NSK 111	AJ002314.1	4500	2,051	96.3	59	"
Rac1	AY029330.1	6140	1,199	97.5	93	Rac-like GTPase
PK2	AJ608157.1	2497	1,766	96.5	81	Ser/Thr protein kinase
PK1	AJ608156.1	4478	1,776	91.9	82	"
TP5	AJ250431.1	4563	2,853	97.1	5,212	Putative β-galactosidase
RAB2	AF397451.1	804	983	98.4	110	Rab2 GTPase
mybAS1	AF198499.1	3085	877	92.2	56	Myb-related protein
JD1	AF316320.1	2676	1,911	96.6	146	Putative Ca^{2+}-binding protein
SUT3	AF149981.1	376	1,577	96.7	99	Sucrose transporter protein
SKP1	AY702087	6186	674	97.1	155	SCF ubiquitin ligase component
CULLIN1	AJ344533	3497	2,422	95.4	46	SCF ubiquitin ligase component

Table 2.3: Summary of Bowtie results for the *Nicotiana* pollen grain transcriptome.

Gene/contig	Number of reads	Comments
Known pollen genes		
CysP	0	-
SuSy	0	-
SBP1	0	-
CESA1	9	No contig > 200 bp
PDC2	35	One contig > 200 bp
DD7	19	No contig > 200 bp
PiSBP1	0	-
PiHT-B	0	-
PhSSK1	0	-
SLF candidate		
1357	22	continuous from base 16-304
Molecular validation		
2401	622	continuous from base 2-1,750
4861	336	continuous from base 10-1,829
4913	537	continuous from base 1-913

Table 2.4: Molecular validation of *N. alata* pollen grain transcriptome. Transcript abundance in the indicated tissue is expressed qualitatively based on band intensity after a standard PCR of 30 cycles. (not detectable = -, detectable = +, abundant = ++, highly abundant = +++). An asterisk (*) indicates a PCR product that was not the expected size.

Contig no.	Contig size (bp)	RPKM	PCR product (bp)	Identity (%)	Expression					Comments
					Pollen	Petal	Style	Leaf	Seedling	
12	904	855	436	100	+++	-	+++	+	-	GTP binding
36	2,163	193	1,077	100	+	-	-	-	-	DNA binding
593	1,101	95	890	99.9	+++	++	+	+	+++	Transmembrane receptor
599	2,398	25	413	100	+++	+++	+	-	-	-
615	1,706	759	932	100	+++	+++	+	+	-	-
637	2,312	61	982	100	+++	-	+++	-	-	Steroid binding
700	907	55	750	99.9	++	-	-	-	-	-
887	1,800	88	982	99.9	++	*	++	+	++	Component of cell membrane
1026	1,404	110	857	100	+++	-	+	-	+++	SNAP receptor
1354	1,518	3,519	860	100	+++	-	+	-	-	Transmembrane transporter
2401	1,777	0[1.]	933	99.9	++	-	-	-	-	Ser/Thr kinase
2403	2,881	58	861	98.6	+++	-	+	++	-	-
2417	1,688	300	951	99.8	+++	+	+++	*	+	Ser/Thr kinase
2423	1,466	123	917	98.9	++	-	+	-	-	Trichome differentiation
2550	1,337	82	620	100	+++	+++	+++	-	+++	Ubiquitin-protein ligase
2637	890	51	1,059	100	+	+	+	+	+	-
2845	1,540	85	950	100	++	-	+++	-	+++	Ser/Thr kinase
2848	2,816	84	990	99.9	+++	-	+++	-	-	Potassium ion transport
2904	1,268	79	827	100	+	+++	+++	-	+++	Ubiquitin-dependent protein
3116	1,872	186	695	97.7	++	*	*	+++	*	Sulfate assimilation
4365	1,456	150	560	99.8	+++	-	+	+	-	SNAP receptor activity
4392	1,986	138	940	99.8	+++	-	+++	+	-	-
4394	920	117	405	82.7	+	+	+	+	-	Cellular component of cell wall
4398	1,649	837	884	99.7	++	++	++	+	-	Asp-type endopeptidase
4402	1,526	52	813	99.9	+++	*	+++	-	-	-
4422	1,141	85	782	100	+++	+	++	+	-	-
4451	3,982	460	869	100	++	-	++	-	-	Protein kinase activity
4518	1,114	1,401	636	99.6	+	-	++	-	-	-
4555	1,025	412	609	99.7	++	-	-	-	-	Protein binding
4768	1,604	89	685	99.8	+++	+++	+++	++	++	Ubiquitin-protein ligase
4783	2,178	19	876	99.9	++	-	++	+	-	-
4861	1,845	66	950	87.9[2.]	*	-	*	-	-	H^+ ATPase
4871	921	70	853	100	+++	+++	+++	+	++	SNAP receptor activity
4885	2,028	132	870	99.9	+	-	+++	+	-	GTPase activity
4913	921	17	643	92[3.]	++	-	-	+	-	Chimeric
4984	1,650	142	836	99.7	+	-	+	-	-	-
5011	1,689	110	837	99.9	+	-	-	+	-	Ethylene biosynthetic process
5066	659	1,631	288	100	+	-	-	-	-	-
5892	703	57	412	81.1	+++	+++	+++	+++	+++	Ubiquitin-protein ligase
6085	2,830	4,158	811	100	+++	-	+++	+	-	Potassium ion transport
6173	602	6,769	488	100	+++	-	+	+++	-	Anchored to membrane
6186	674	155	406	99.9	++	+++	+++	+	-	Ubiquitin-dependent protein
6203	2,375	475	638	95.1	+++	+++	+	-	+++	Phosphotransferase activity
6423	2,451	210	983	99.9	+	-	+	-	+	Ubiquitin-protein ligase
6440	2,358	66	609	96.4	+	-	+	-	-	Protein binding

[1.] RPKM values <1 are rounded to 0

[2.] 140 bp insertion

[3.] 46 bp insertion

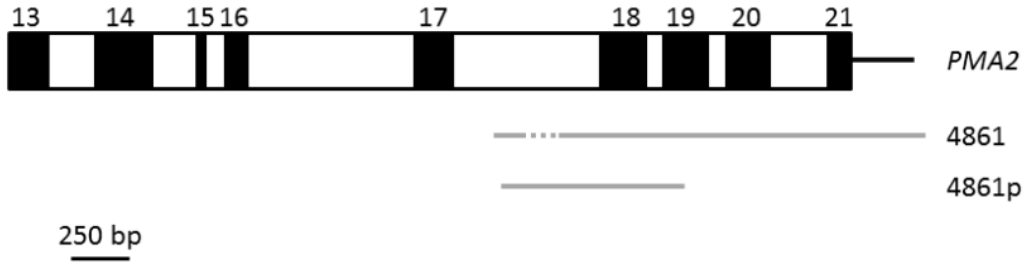


Figure 2.4: Diagrammatic representation of the alignment between GenBank accession M80492 containing exons 13-21 of *N. plumbaginifolia PMA2* (Perez et al., 1992), superassembly contig 4861 and its PCR product (4861p). Exons are shown as black boxes and are numbered. Introns are shown as open boxes and the 3' untranslated region of PMA2 is shown as a black line. The grey lines represent the 4861 and 4861p sequences and the dashed section in the 4861 sequence shows the 140 bp deletion relative to the *PMA2* and 4861p sequences.

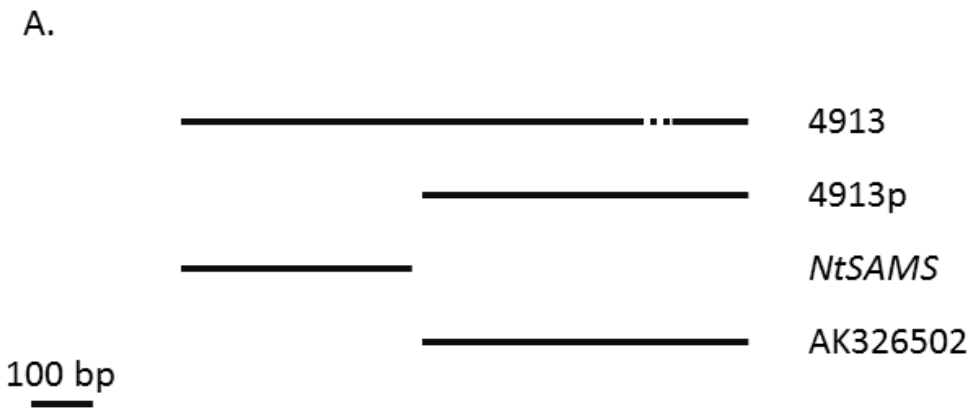
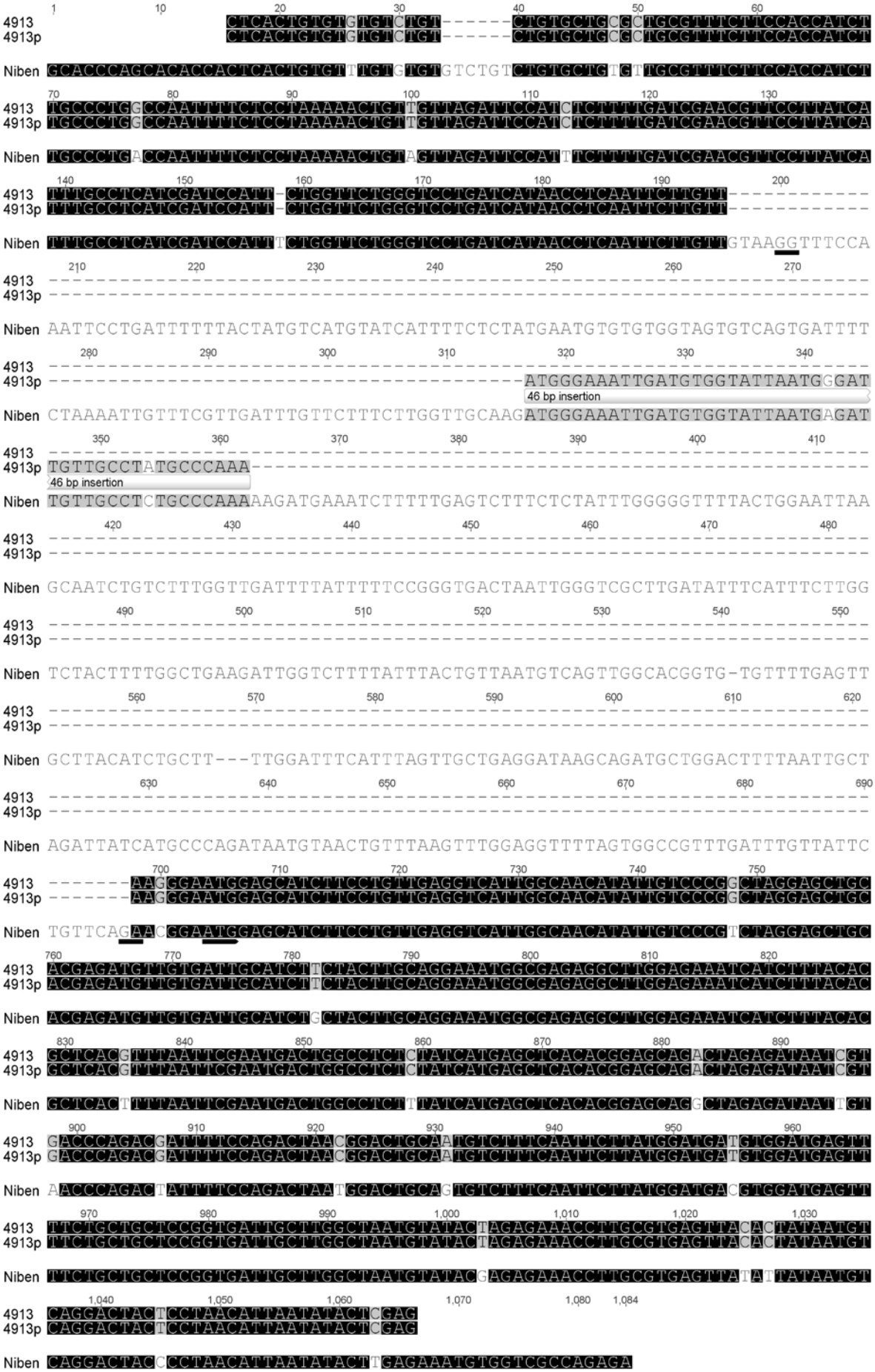


Figure 2.5A: Diagrammatic representation of the alignment of contig 4913, its PCR product 4913p, the *N. tabacum S-adenosyl-L-methionine synthetase* (*NtSAMS*) cDNA (accession no. AF127243) and a *S. lycopersicum* fruit-derived cDNA (accession no. AK326502). The dashed section in the 4913 sequence shows the 46 bp deletion relative to the 4913p and AK326502 sequences.

Figure 2.5B: Alignment of contig 4913, 4913p and part of scaffold 24821115 of the *N. benthamiana* (Niben) genome (Niben.v0.3). Dashes (-) indicate gaps introduced to maximize the alignment. Unshaded region indicates less than 60% sequence identity, light grey shading indicates 60 to 79% sequence identity, dark grey shading indicates 80 to 99% sequence identity and black shading indicates 100% sequence identity. A donor and acceptor site predicted by NetGene2 in the Niben sequence is each underlined by black bold lines and the ATG at the start of the open reading frame in the *N. benthamiana* sequence is indicated by black arrow. The 46 bp sequence present in 4913p but not in contig 4913 as indicated.


4913
4913p
Niben
46 bp insertion
46 bp insertion

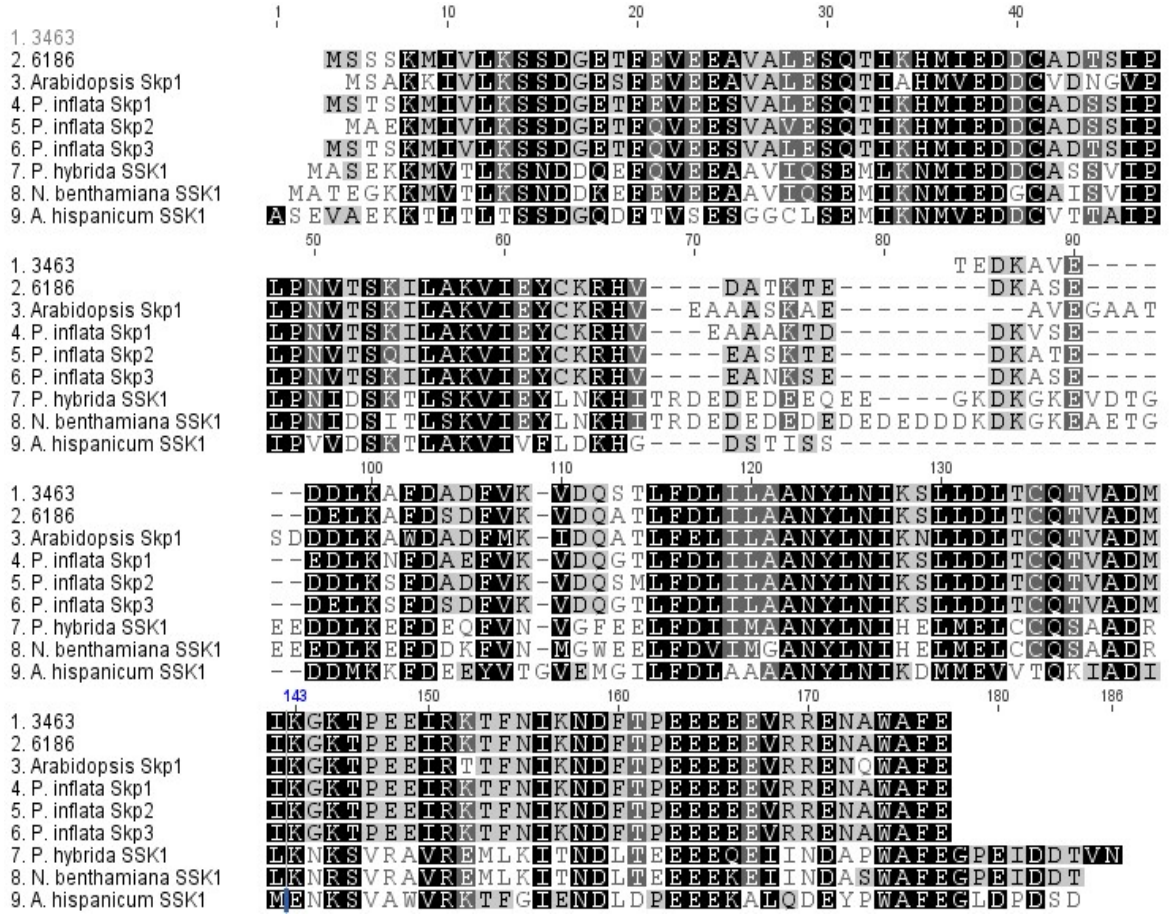


Figure 2.6: Sequence alignment of various plant Skp and SSK proteins and the proteins encoded by contigs 3463 and 6186 from the *N. alata* pollen transcriptome. Dashes (-) show gaps inserted in the alignment to maximize identity. Unshaded regions indicate <60% sequence identity, light grey shading 60-79% identity, dark grey shading 80-99% identity and black shading indicates 100% identity. Table 2.6 lists accession numbers for the *P. inflata* and *P. hybrida* sequences. Accession numbers for *Arabidopsis* Skp1 and *Antirrhinum hispanicum* SSK1 are NM_106245 and DQ355480, respectively. The putative *N. benthamiana* SSK1 ortholog was obtained from scaffold 25458 of Niben.v0.4.2 using the *P. hybrida* SSK1 amino acid sequence as a query.

Figure 2.7: Alignment of Rbx1 (*P. inflata* Rbx1) and the protein encoded by contig 6029. The Cullin1 binding site is indicated and black underlines indicate residues putatively involved in binding Zn.

Table 2.5: List of GSI-related transcripts from various *Petunia* species and matching contigs in the *N. alata* pollen grain transcriptome.

Gene name	Accession no.	Contig no.	Contig size (bp)	Pairwise identity (%)	RPKM	Reference
SLF1	AY500390 - AY500392	3684 (DD4)	841	73.8 – 74.7	71	Sijacic et al., (2004)
SLF2	AB568394 - AB568398	3684 (DD4)	841	75.1 – 73.8	71	Kubo et al., (2010)
SLF3	AB568399 - AB568404	2031	560	86.8 – 83	58	"
SLF4	AB568405 - AB568410	5258	592	81.1 – 79.3	84	"
SLF5	AB568411 - AB568416	452	1,066	69.1 – 68.7	33	"
SLF6	AB568417 - AB568422	452	1,066	69.6 – 67.8	33	"
SLF7	AB932987, AB933015, AB933078,	452	1,066	69-70	33	Kubo et al., (2015)
SLF8	AB933095	452	1,066	69-71.2	33	"
SLF9	AB932956, AB932977, AB932978, AB932988 AB933005, AB933016, AB933027, aB933043,	452	1066	67.7-68.4	33	"
SLF10	AB933044, AB933062, AB933079, AB033096, AB933112, AB933129, AB933130	452	1066	66.9-67.4	33	"
SLF11	AB932966, AB932979, AB932989, AB932990, AB933017, AB933029, AB933030, AB933063, AB933097, AB933098, AB933115, AB933131	452	1066	69-70	33	"
SLF12	AB932967, AB932981, AB932991, AB933006, AB933018, AB933031, AB933045, AB933046, AB933064, AB933080, AB933099, AB933116, AB933132	5258	592	77.9-78.6	84	"
SLF13	AB932968, AB932982, AB932992, AB933007, AB933019, AB933032, AB933047, AB933065, AB933081,	2031	560	82.6-84.4	58	"
SLF14	AB933100, AB933117, AB933133 AB932969, AB933008, AB933033, AB933034, AB933048, AB933066,	452	1066	69-70	33	"
SLF15	AB933082, AB933101, AB933102,	452	1066	69.8-70.2	33	"
SLF16	AB933118, AB933134 AB932970, AB932994, AB933009, AB933020, AB933036, AB933049, AB933067, AB933083, AB933103,	452	1066	68.9-70.3	33	"
SLF17	AB933119, AB933135 AB932983, AB932995, AB933010,	452	1066	68.4	33	"
SLF18	AB933021, AB933037, AB933050, AB933068, AB933084, AB933104, AB933129, AB933136 AB932996, AB933069, AB933105 AB932972, AB932984, AB932997, AB932998, AB933011, AB933023, AB933038, AB933052, AB933070, AB933086, AB933106, AB933121, AB933137 AB932973, AB932985, AB932999, AB933012, AB933039, AB933087, AB933138 AB933013, AB933071, AB933107, AB933139	452	1066	66.5	33	"
PiSKP1/2/3	DQ250013 - DQ250015	6186	674	83.4 - 88	155	Hua and Kao (2006)
PiCullin1G	DQ250017	3497	2,422	91	46	"
		4884	2,690	90	71	"
PiRBX1	DQ250021	6029	867	85.6	80	"
PiSBP1	DQ250022	no hits	-	-	-	"
PhSSK1	FJ490176	no hits	-	-	-	"

Figure 2.8: Alignment of human Cullin1 (UniProt accession Q13616), NtCul1A (*N. tabacum* Cullin1A; UniProt accession Q711G8) and the proteins encoded by contigs 3497 and 4884. Grey overlines indicate the Skp1 binding residues and black overlines the Rbx1 binding residues in human Cullin1.

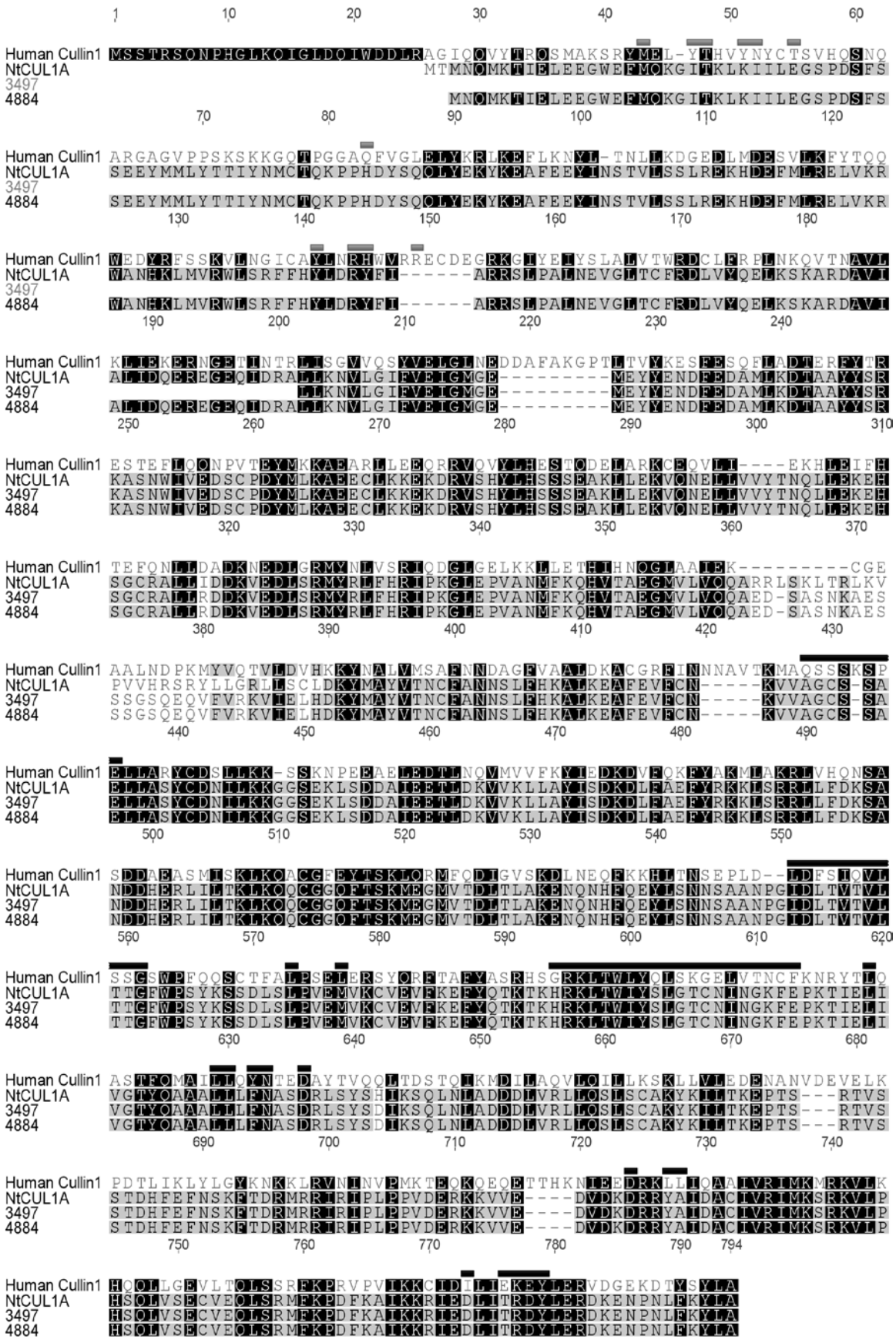

Human Cullin1
NtCUL1A
3497
4884

Table 2.6: Expression of SCF E3 ligase components and putative SLFs in various *N. alata* tissues. Transcript abundance in the indicated tissue is expressed qualitatively as described in the legend to Table 2.5.

Gene	Contig no.	Expression level				
		Leaf	Petal	Style	Seedling	Pollen
Skp1	6186	+	+++	+++	-	++
Cullin1	3497/4884	+++	+++	+++	+++	+++
SBP[1]	-	+++	+	+++	++	+++
Rbx1	6029	+++	+++	+++	+++	+++
SLF	452	++	+	++	-	+++
	1357 (DD3)	-	-	-	-	-
	1945 (DD2)	-	-	-	-	++
	3684 (DD4)	-	+	+++	-	+++
	4791	++	+	++	-	+++
	5494 (DD1)	+	+	+++	-	+++

[1.] primers designed to the *NaSBP1* sequence.

Figure 2.9: Amino acid alignment of the four putatively novel SLFs identified in the pollen transcriptome and the ten *N. alata* DDs. The F-box domain as indicated. Dashes (-) show gaps inserted in the alignment to maximize identity. Black shading indicates 50% or more amino acid identity.

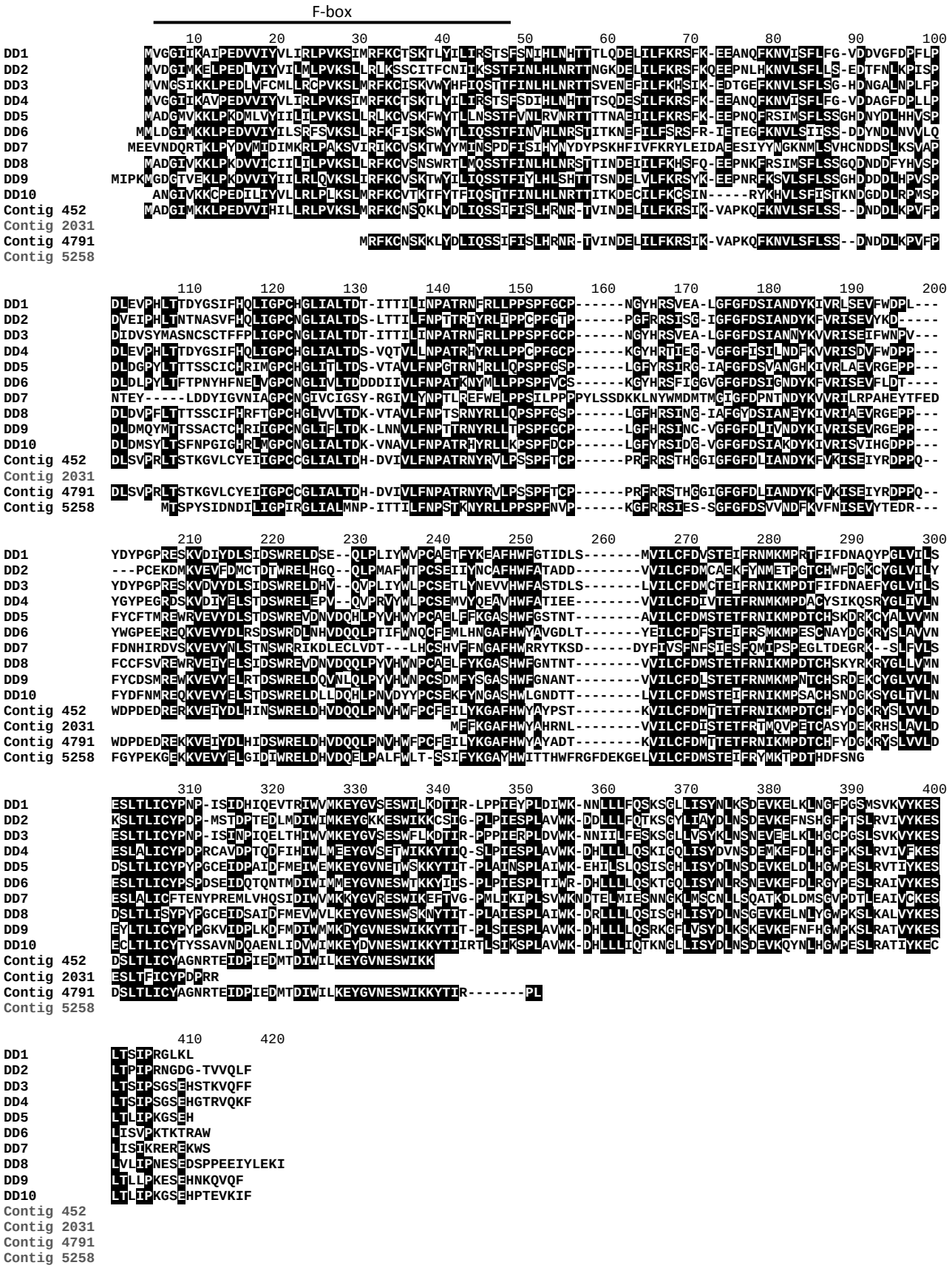
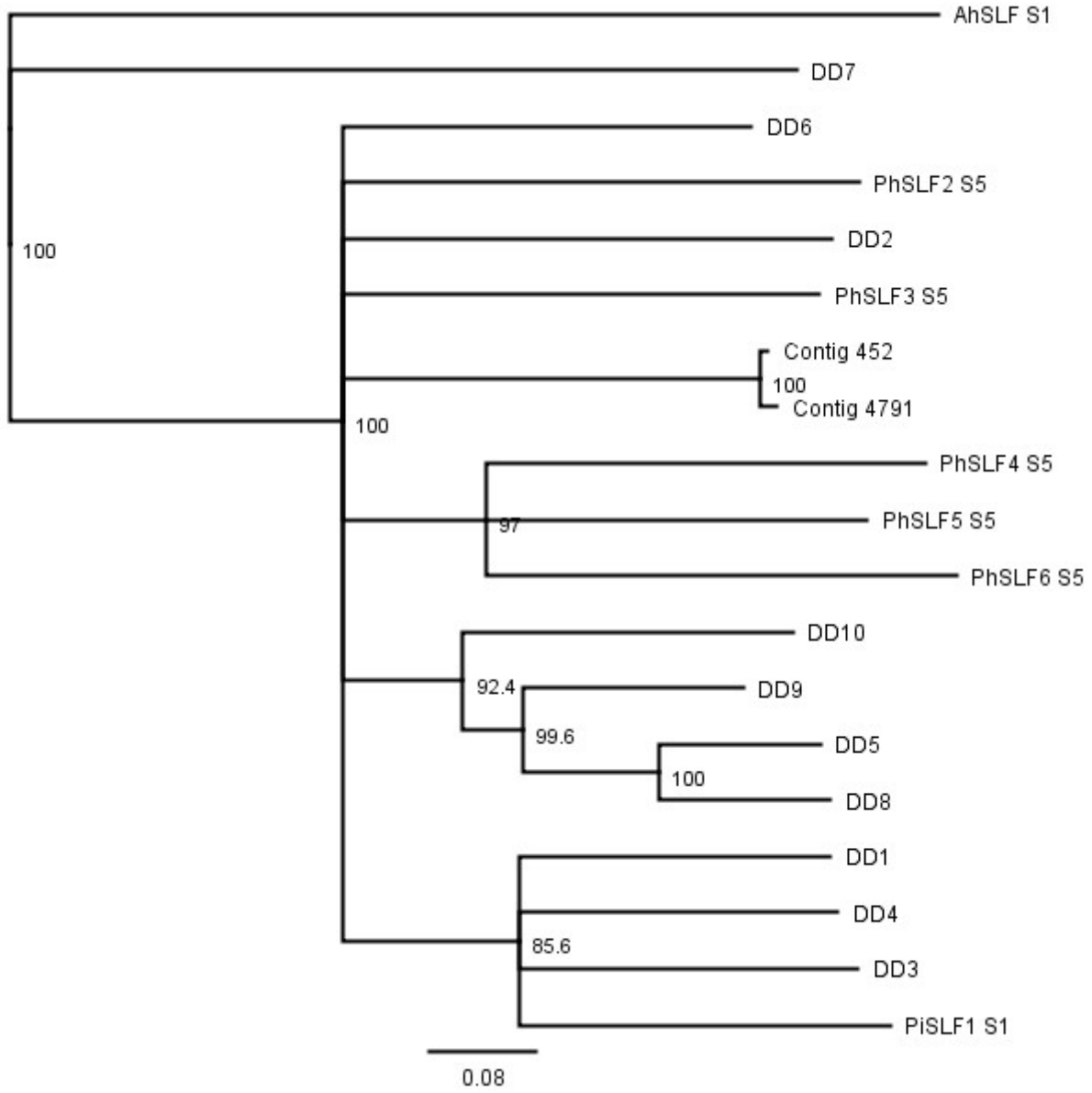

F-box
DD1
DD2
DD3
DD4
DD5
DD6
DD7
DD8
DD9
DD10
Contig 452
Contig 2031
Contig 4791
Contig 5258

Figure 2.10: Consensus distance tree produced from an amino acid alignment of the proteins encoded by contig 452 and 4791, the *N. alata* DDs, representatives of each of the six classes of *Petunia* SLFs as described in Kubo et al., (2010), and *Antirrhinum hispanicum* SLF S_1 (AhSLF S1). Numbers to the right of nodes show % bootstrap support (5,000 replicates) with any nodes receiving less than 75% support being collapsed. The alignment was produced using ClustalW and scored with a BLOSUM substitution matrix. The tree was produced in Geneious using the neighbor joining method with AhSLF S1 chosen as the outgroup. Accession numbers of sequences from species other than *N. alata* are AhSLF S1 (CAD56663), *P. inflata* SLF1 S_1 (PiSLF1 S1, AAS79484), *P. hybrida* SLF2 S_5 (PhSLF2 S5, BAJ24853), *P. hybrida* SLF3 S_5 (PhSLF3 S5, BAJ24858), *P. hybrida* SLF4 S_5 (PhSLF4 S5, BAJ2486), *P. hybrida* SLF5 S_5 (PhSLF5 S5, BAJ24870) and *P. hybrida* SLF6 S_5 (PhSLF6 S5, BAJ24876).


AhSLF S1
DD7
DD6
PhSLF2 S5
DD2
PhSLF3 S5
Contig 452
Contig 4791
PhSLF4 S5
PhSLF5 S5
PhSLF6 S5
DD10
DD9
DD5
DD8
DD1
DD4
DD3
PiSLF1 S1
100
100
100
97
92.4
99.6
100
85.6
0.08

3.1: Introduction

After the discovery of the Solanaceae *style S* gene by Anderson et al., (1986), the search for the *pollen S* counterpart of this complex locus was an ongoing research activity pursued by many groups for many years (McClure, 2004). The paper by Sijacic et al., (2004) appeared to bring this search to an end, as they reported that transgenic *P. inflata* plants expressing an extra copy of *SLF* in pollen were self-compatible. Self-compatibility in these plants arises because the extra SLF protein allowed pollen tubes to overcome the cytotoxic effect of all stylar S-RNases, as predicted by the then-current inhibitor model of pollen S action (Thompson and Kirch, 1992; Golz et al., 2001; see Chapter 1 for details). The SLFs are F-box proteins and presumably function as one component of a multi-subunit protein ubiquitin ligase, with their main role being to select target proteins for ubiquitylation (Vierstra, 2003, 2009). As the most common outcome of ubiquitylation is protein degradation by the 26S proteasome pathway, the findings of Sijacic et al., (2004) provided a mechanistic explanation for the inhibitor model.

To further this model, Hua and Kao (2006) studied the binding of gluthathione S-transferase (GST) tagged versions of *P. inflata* S-RNases and six histidine-tagged ($(His)_6$) versions of SLFs *in vitro*. A typical binding assay used a $(His)_6$-tagged version of $PiSLF_1$, the SLF from the S_1 allele of *P. inflata* (the protein was renamed *P. inflata* S_1-SLF1 by Kubo et al., (2010) and will simply be called PiSLF1 henceforth) and GST-tagged versions of *P. inflata* S_1-RNase and S_2-RNase, to show that PiSLF1 bound more strongly to the S_2-RNase than to the S_1-RNase. That is, PiSLF1 interacted differently with the S-RNases of other *S* alleles than with its cognate S-RNase, as suggested by the inhibitor model. Hua et al., (2007) then showed that $(His)_6$-tagged SLF-like proteins – SLF-like proteins are pollen-expressed and similar in sequence to SLFs but at the time were presumed not to control pollen function – either failed to bind S-RNases or could not compete for binding with PiSLF1 in an *in vitro* assay, because SLF-like proteins lacked certain SLF-specific domains that regulate interactions with self- and non-self S-RNases. Later, as part of their collaborative recognition model, Kubo et al., (2010) showed that at least some SLF-like proteins do control pollen function by interacting with a subset of non-self S-RNases, and accordingly classified the 30 known *Petunia* SLF and SLF-like sequences into six SLF subgroups. The number of *Petunia* SLF groups was recently increased to 17 (Williams et al., 2014a, b).

Wheeler and Newbigin (2007) isolated ten *F-box protein* cDNAs (the *DD*s) from *N. alata* pollen and showed that most of the *DD* genes were at or near the *S* locus. Thus, when this thesis started, any one of the *DD* genes could have been the *N. alata* homolog of *PiSLF1* (the others would have been *SLF-likes*). The intention was thus to use the *in vitro* binding assay approach of Hua and Kao (2006)

and Hua et al., (2007) to study the *N. alata* DD proteins and identify the likely *N. alata* SLF ortholog. Because the *in vitro* assay showed that PiSLF1 interacted with *P. inflata* S_2-RNase (Hua and Kao 2006), PiSLF1 and S_2-RNase (PiS_2-RNase) were chosen as positive controls for this work.

This chapter reports on expression studies in *E. coli* and attempts at purifying $(His)_6$-tagged PiSLF1 ($(His)_6$:PiSLF1) and various DDs, GST-tagged PiS_2-RNase (GST:PiS_2-RNase) and *N. alata* S_6-RNase (GST:NaS_6-RNase). A pull-down assay similar to that described by Hua and Kao (2006) but using crude cell lysates instead of purified proteins, was developed but proved unreliable and was later replaced with a co-immunoprecipitation assay that used $(His)_6$-tagged SLF proteins and native S-RNases from *N. alata* styles. Homology modeling of a DD protein was performed and reasons for the difficulties experienced in replicating earlier studies and some suggested solutions are discussed.

3.2: Materials and Methods

3.2.1: Recombinant protein expression in *Escherichia coli*

Figure 3.1 shows in diagrammatic form the recombinant proteins used in this chapter. The cDNAs for *DDs 2*, *5-8* were obtained from David Wheeler (School of Botany, University of Melbourne) and were cloned in-frame into the bacterial expression vector pET30a (Merck Millipore) using standard recombinant DNA techniques (Sambrook and Russell, 2001). DD2 was from S_1 allele and DD5, DD6, DD7 and DD8 were from S_2 allele. David Wheeler also provided plasmids containing *N. alata RNase NE* (accession no. NAU13256) from bases 107 to 718 (Dodds et al., 1996) and *N. alata* S_6*-RNase* (accession no. NAU8860) from bases 66 to 645 (Anderson et al., 1989). Neither plasmid contained the sequence encoding the signal peptide; the open reading frames were cloned in-frame into the bacterial expression vector pGEX 4T-1 (GE Healthcare). The cDNAs for PiSLF1 (Genebank accession AY500390, bases 107 to 1276) and PiS_2-RNase (Genebank accession number AF301533, bases 67 to 663) were chemically synthesis by GeneArt (www.lifetechnologies.com) and cloned into pET30a and pGEX4T-1, respectively. All cDNAs cloned into pET30a were expressed as proteins with a $(His)_6$-tag at their N-terminal and all cDNAs cloned into pGEX 4T-1 were expressed as proteins with a GST tag at their N-terminal.

For pGEXT4T-1 constructs expression in *E. coli* BL21 DE3 strain, recombinant plasmids were transformed into *E. coli* by the heat shock method and positive colonies selected after plating out on solid media using the appropriate antibiotic selection (Sambrook and Russell, 2001). A single transformed bacterial colony was inoculated into 5 mL of LB medium (LB medium: 10 g NaCl, 5 g yeast extract and 10 g tryptone per litre) supplemented with ampicillin (50 µg/mL) and grown with continual shaking at 37°C until the optical density at 600 nm (OD_{600}) was approximately 0.5. For

some experiments, glucose was added to the medium to a final concentration of 3% (w/v). Protein expression was induced by adding isopropylthio-β-D-galactoside (IPTG; Biovectra) to a final concentration of 1 mM, unless otherwise stated. For post-induction growth at temperatures other than 37°C, the culture was equilibrated to the required induction temperature for 30 min prior to addition of IPTG. Cultures induced at 25°C and 37°C were harvested 3 h post-induction and cultures induced at 16°C were harvested 16-20 h post-induction by centrifugation at 3,000 g for 30 min. Protein expression was carried out as recommended by the manufacturers of the expression plasmid (GE Healthcare).

For pET30a expression constructs, the following bacterial strains were used: BL21 DE3, BL21 DE3 codonplus RIL, ArticExpress (all from Stratagene) and BL21 DE3 star (Invitrogen). Recombinant plasmids were transformed into *E. coli* by the heat shock method and positive colonies selected after plating out on solid media using the appropriate antibiotic selection (Sambrook and Russell, 2001). For BL21 DE3 and BL21 star, the medium was supplemented with kanamycin (30 µg/mL), for BL21 DE3 codonplus RIL, the medium was supplemented with kanamycin (30 µg/mL) and chloramphenicol (50 µg/mL), and for ArticExpress the medium was supplemented with streptomycin (75 µg/mL), kanamycin (30 µg/mL) and gentamycin (20 µg/mL). Single colonies were grown with continual shaking and protein expression induced with IPTG as described above. Addition of 3% glucose and post-induction growth at lower temperatures was done as described. Bacterial pellets were harvested by centrifugation and frozen at -20°C if not used immediately.

3.2.2 Extraction and analysis of bacterial proteins

The bacterial pellets were lysed using Bugbuster master mix (Novagen) with or without added EDTA-free protease inhibitor cocktail (Roche) and soluble and insoluble protein fractions were prepared according to manufacturer's recommended protocol (Novagen). Briefly, the insoluble fraction (pellet) was separated from the soluble fraction (supernatant) after cell lysis by spinning at 16,000g for 30 min. The soluble fraction (80 µl) was mixed with 20 µl protein loading buffer (Invitrogen) to a final 1× concentration, heated at 70°C for 15 min and cooled to room temperature before being briefly spun to collect any condensate. The cell pellet was resuspended in the original amount of Bugbuster master mix solution used for cell lysis and 80 µl of the suspension was mixed with 20 µl protein loading buffer and treated as for the soluble fraction. Total lysate was obtained by resuspending a small quantity of pelleted cells in 5x protein loading buffer and heating at 70°C for 15 min.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed using the Xcell Surelock system (Invitrogen). Samples were run on either a 12% gel or a 4-12% gradient gel in 1× MOPS (50 mM MOPS, 50 mM Tris-Base, 0.1% SDS, 1 mM Ethylenediaminetetraacetic acid **(**EDTA), pH 7.7) or 1× MES running buffer (50 mM MES, 50 mM Tris-Base, 0.1% SDS, 1 mM EDTA) using the manufacturer's recommended conditions. After electrophoresis, the gel was either stained with Gelcode blue stain (Thermo Fisher Scientific) according to manufacturer's recommended protocol, or the proteins were transferred to a nitrocellulose membrane (Osmonics) for immunoblot analysis. Protein transfer was performed using a Xcell Surelock system transfer module in transfer buffer (25 mM bicine, 25 mM bis-Tris, 1 mM EDTA, pH 7.2) with 10% (v/v) ethanol per membrane at 40 volts for 90 min at 4°C. Membranes were blocked with 5% (w/v) milk powder (Diploma instant) in Tris-buffered saline (TBS; 20 mM Tris, pH 7.6, 100 mM NaCl) with 0.01% (v/v) Tween 20 (TBST) for 1 h at room temperature with gentle shaking. The blocking solution was discarded and membranes were probed with a selected primary antibody in TBST plus 5% milk overnight at 4°C. The following primary antibodies were used at the indicated dilution: for $(His)_6$-tagged proteins, a 1:3,000 dilution of a mouse monoclonal anti-$(His)_6$ antibody conjugated to peroxidase (Sigma Aldrich); for GST-tagged proteins, a 1:5,000 dilution of a goat anti-GST antibody (Abcam); for *N. alata* S_2-RNase, a 1:5,000 dilution of a polyclonal rabbit antibody (Anderson et al., 1989); and for *N. alata* S_7-RNase, a 1:3,000 dilution of a polyclonal rabbit antibody (prepared by A. Vissers and provided by E. Newbigin, School of Botany, University of Melbourne). The next day, membranes were washed several times in TBST (10 min per wash), then incubated with either a rabbit anti-goat antibody (for detection of GST-tagged proteins; Thermo Fischer Scientific) or a goat anti-rabbit antibody (for detection of S-RNases; Sigma Aldrich) at a dilution of 1:8000 and 1:50,000 respectively in TBS at room temperature for 60 min. Both secondary antibodies were conjugated to peroxidase and no secondary antibody was needed for detection of $(His)_6$-tagged proteins as a peroxidase-conjugated primary antibody was used. Membranes were washed as before, bathed in chemiluminescence solution (Thermo Fisher Scientific) and either exposed to X-ray film (GE healthcare) or digitally scanned using ChemiDoc imager (Biorad).

3.2.3: Pull-down assay

Soluble cell lysate (1 mL) from bacterial cells expressing GST:PiS_2-RNase was prepared as described in section 3.2.2 using Bugbuster master mix solution. Lysate was incubated with 30 µl of glutathione sepharose resin (GE Healthcare) at 4°C for 90 min with gentle agitation. The resin was allowed to settle by gravity, washed twice with a 1:10 dilution of Bugbuster master mix and incubated with 1 mL of $(His)_6$:fusion protein soluble cell lysate at 4°C for 3 hrs. The resin was washed three times as

before and bound proteins eluted by heating at 95°C for 5 min in 2× protein loading buffer. Samples were separated on 12% polyacrylamide gels, as described in section 3.2.2.

3.2.4: Extraction of *Nicotiana alata* styles

Nicotiana alata plants (*S* genotype S_7S_7) were grown in a pollinator-proof glasshouse, as described in Anderson et al., (1986). Styles (including stigma) were collected at anthesis and used either fresh or frozen at -80°C until needed. Styles were frozen in liquid nitrogen, placed in a 1.5 mL tube with a 3 mm ball bearing and homogenised to a fine powder for 1 min (25 vibrations/s) using a Retsch mixer miller MM400 (Qiagen). Ice-cold style extraction buffer (100 mM Tris pH 8.0, 50 mM EDTA and 14 mM β-mercaptoethanol) at a 1:4 fresh weight per volume ratio was added and the extract incubated on ice for 30-60 min. Insoluble material was removed by centrifugation at 16,000 g for 30 min and supernatant transferred to a new tube. Supernatant was filter through 0.22 µM filter (Millipore) to remove remaining debris. Total stylar extracts protein concentration was determined using the Bio-Rad protein assay reagent (Bio-Rad) with bovine serum albumin (BSA) as the standard. A 1:1 serial dilution of BSA standard (1mg/mL) was performed with water and Bradford assay reagent was diluted 5 times with water. 20 µL of each standard, 1:9 diluted and undiluted total style extract was incubated with 480 µL of Bradford assay reagent for 5 minutes before absorbance was taken at 595 nm. All samples were prepared in triplicates. Base on the average BSA standards reading, a curvilinear regression graph was plotted using Excel spreadsheet and a polynomial equation was derived from graph which was used to calculate protein concentration. Aliquots were kept at -80°C until needed.

3.2.5: Co-immunoprecipitation assay

Co-IP assays were performed using either anti-$(His)_6$ tag mouse monoclonal antibody (3 µl) or rabbit anti-S_7-RNase polyclonal antibody (10 µl; see section 3.3.2) in binding buffer (50 mM Tris pH 7.6, 100 mM NaCl, 5 mM $MgCl_2$ 0.01% Nonidet P40, 1 mM dithiothreitol (DTT); see Hua and Kao, 2006). The selected antibody was incubated with 30 µl of protein A agarose beads (50% suspension; GE healthcare) in a 1.5 mL tube at room temperature for 90 min with gentle rocking. The beads were allowed to settle by gravity, the supernatant was removed and the antibody-loaded beads were incubated with $(His)_6$:fusion protein soluble lysate or *N. alata* S_7S_7 style extract (depending on which antibody was bound to the agarose beads). The tube was left at room temperature for 90 min with gentle rocking and the beads allowed to settle by gravity. The supernatant was removed and the beads were washed 3 times with 1 mL of binding buffer. Bound proteins were eluted by heating at

95°C for 5 min in 2× protein loading buffer, separated on 12% polyacrylamide gel and immunoblotted as described in section 3.2.2.

3.2.6: Protein purification

For nickel-affinity chromatography of $(His)_6$:fusion proteins, bacterial cells were lysed as described in section 3.2.2 and affinity chromatography with nitroloacetic acid (NTA) beads was done according to the manufacturer's recommended protocol (Novagen). For small-scale purifications, a batch-wise protocol was used with up to 5 mL of bacterial lysate and 1 mL of a 50% suspension of NTA beads. For larger scale purifications, a column-based protocol was used with up to 40 mL of bacterial lysate and 10 mL of NTA beads. The last wash fraction (washes done with 300 mM NaCl, 50 mM Tris buffer, 20 mM imidazole, pH 8.0) prior to elution was retained for later analysis. Bound proteins were eluted with elution buffer (300 mM NaCl, 50 mM Tris buffer, 250 mM imidazole, pH 8.0) and the NTA beads were then incubated at 95°C for 5 min in 30 µL of 2× protein loading buffer to elute any remaining bound protein. Cell lysates and protein fractions (last wash, eluate and beads) were analysed on polyacrylamide gels as described in section 3.2.2

For glutathione-affinity chromatography of GST:fusion proteins, bacterial cells were lysed as described in section 3.2.2 and affinity chromatography with glutathione sepharose beads was done batch-wise according to the manufacturer's recommended protocol (Novagen). The last wash fraction (washes done with 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3) prior to elution was retained for later analysis. Bound proteins were eluted with elution buffer (50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0) and the glutathione sepharose beads beads were then incubated at 95°C for 5 min in 30 µL of 2× protein loading buffer to elute any remaining bound protein. Cell lysates and protein fractions were analysed as described above.

An eluate fraction containing soluble $(His)_6$:PiSLF1 obtained using the column-based protocol was concentrated using an Amicon spin column with a 10 kDa molecular weight cut-off (Millipore), filtered through a 0.22 µM filter (Millipore) and injected into a MonoQ anion exchange column (GE Healthcare) pre-equilibrated with running buffer (50 mM Tris pH 7.6, 100 mM NaCl, 5 mM $MgCl_2$ and 1 mM dithiothreitol). The column flow rate was 0.5 mL/min. Running buffer was applied for first 15 min of the run and proteins were eluted with a gradient of 0-1M NaCl in same buffer over 2.5 column volumes. Fractions (1 mL) were collected over the gradient and 20 µl of each fraction was analysed on a 12% polyacrylamide gel as described in section 3.2.2.

3.2.7: Protein structure modeling

A protein alignment of PiSLF1 and DDs 1-10 (accession numbers AAS79484 and ABR18781 to ABR18790), produced using CLUSTAL omega (Sievers et al., 2011) as implement by Geneious Pro version 5.5.6 (Biomatters, Auckland, New Zealand), was used as input for secondary structure prediction by Jpred3 (Cole et al., 2008) with the default settings. A REP search (Andrade et al., 2000) was performed using PiSLF1 and the DD1-10 protein sequences, without assigning a cut off value, to identify putative repeat sequences. Weak non-overlapping Kelch repeats were identified by REP in the DDs and PiSLF1 sequences and these were used as input into a PSI-BLAST search (Altschul et al., 1997) (e-values<0.005), of the UniProtKB/Swiss-Prot database to assess the similarity of these putative blocks of Kelch repeats and surrounding sequences (60 residues) to known Kelch repeat sequences. The resulting alignments to strong Kelch repeats were used to iteratively back-predict the position of additional Kelch repeat sequences in PiSLF1 and the 10 DD proteins.

The Kelch repeat protein Keap1 (PDB 1X2J; (Padmanabhan et al., 2006) was used for homology modeling of DD1. The Keap1 crystal structure was obtained from the Protein Data Bank (www.rcsb.org) and a homology model of DD1 was built using the Swiss-Model online tool (swissmodel.expasy.org; Arnold et al., 2006). SuperLooper (bioinf-applied.charite.de/superlooper; (Hildebrand et al., 2009) was used to model loop regions of DD1 where Keap1 did not provide a suitable template. The software SwissPDB Viewer 4.0.1 (Guex and Peitsch, 1997) was used to produce the final protein structure and the figures were created using PyMOL 1.3 (www.pymol.org).

3.3: Results

3.3.1: Production of soluble tagged versions of PiSLF1 and PiS_2-RNase in *E. coli*

GST:PiS_2-RNase and $(His)_6$:PiSLF1 were expressed in *E. coli* with the aim of purifying soluble forms of recombinant protein for use in pull-down assays as described by Hua and Kao (2006). Figure 3.1 shows in a diagrammatic form the overall structure of the recombinant proteins that were expressed in the BL21 (DE3) strain of *E. coli*, as used in Hua and Kao (2006).

Figure 3.2 shows the accumulation of soluble and insoluble products over time in cells expressing GST:PiS_2-RNase, as detected by immunoblots with an anti-GST antibody. The expected molecular weight of GST:PiS_2-RNase is about 52 kDa, and a protein of this size was detected among the insoluble products 2 hr after IPTG induction in cells grown at 37^{o}C (Figure 3.2A). A second band of about 40 kDa also present in the insoluble fraction presumably represents a loss of about 10 kDa from the COOH-terminal end of GST:PiS_2-RNase. Levels of the 50 kDa and 40 kDa proteins in the insoluble fraction declined over the 4 hr time course and neither protein accumulated in the soluble

fraction, where the only protein detected, of about 26 kDa, presumably represented the GST tag which is roughly this size (Figure 3.2A). Levels of the 26 kDa protein in the soluble fraction remained approximately the same over the time course.

To see if levels of intact, soluble GST:PiS_2-RNase could be improved, cells were grown at various temperatures after induction (Figures 3.2B and C). Results with cells grown at 25^{o}C were essentially no different to those with cells grown at 37^{o}C: the 50 kDa GST:PiS_2-RNase was only present in the insoluble fraction and levels of this protein declined with time. Only the 26 kDa protein accumulated in the soluble fraction (Figure 3.2B). However, cells grown at 16^{o}C did accumulate the 50 kDa protein in the soluble fraction, with optimal production seen 3 hr after induction in medium supplemented with 3% glucose (Figure 3.2C). The 40 kDa and 26 kDa proteins were also present in this fraction and more GST:PiS_2-RNase and 40 kDa protein accumulated in the insoluble fraction than in the soluble fraction.

A similar set of experiments was done with $(His)_6$:PiSLF1 expression. The expected molecular weight for $(His)_6$:PiSLF1 is about 50 kDa, and a protein close to this size was detected by an anti-$(His)_6$ antibody in lysates from cells grown at 25^{o}C and 16^{o}C (Figure 3.3A, B). Cells grown post-induction at 25°C contained more $(His)_6$:PiSLF1 in the soluble fraction than cells grown at 16°C; however, unlike the case with GST:PiS_2-RNase, no lower molecular weight forms were observed in the soluble fraction (Figure 3.3). The lower molecular weight (~39 kDa) protein detected in the insoluble fraction of cells grown at 25°C presumably represents some trimming at the COOH-terminal end of $(His)_6$:PiSLF1. When cells were grown at 16°C, considerably less soluble $(His)_6$:PiSLF1 accumulated, with most of the protein being in the insoluble fraction (Figure 3.3B). This could be because the lower temperature result in slower growth and less protein being made. Other conditions were tried (growth at different temperatures, varying the amount of IPTG, addition of glucose to the growth medium), but none significantly improved the yield of $(His)_6$:PiSLF1 above those seen in Figure 3.3A (data not shown).

Figure 3.4A shows an attempt at purifying soluble GST:PiS_2-RNase by affinity chromatography using glutathione resin. Although the soluble fraction contained GST:PiS_2-RNase and the 40 kDa and 26 kDa proteins, none of these proteins eluted from the resin and only the 26 kDa protein was found bound to the beads. Since it was possible GST:PiS_2-RNase was being cleaved during purification, protease inhibitors were added to the solution used to make the lysate but this did not alter the result (data not shown). As cleavage of the GST tag makes purification of intact GST:PiS_2-RNase impossible, no further attempts were made to optimise this method and all future experiments used the soluble cell lysate instead.

Figure 3.4B shows an attempted purification of soluble $(His)_6$:PiSLF1 by nickel (Ni)-affinity chromatography. After incubation with Ni-NTA beads, very little $(His)_6$:PiSLF1 remained in solution (FT fraction), indicating that almost all the protein bound to the resin. Binding, however, appeared irreversible as $(His)_6$:PiSLF1 was not eluted when the beads were incubated with 250 mM imidazole, suggesting a very strong, non-specific interaction.

3.3.2: Production of soluble *N. alata* proteins

Two extracellular ribonucleases from *N. alata* were selected for expression in *E. coli*: the self-incompatibility associated S_6-RNase (Anderson et al., 1989) and RNase NE, an extracellular ribonuclease from *N. alata* styles. RNase NE was intended for use as a negative control in the *in vitro* SLF/S-RNase interaction studies, because it is a member of the T2 RNase family like the S-RNases but it plays no role in self-incompatibility (Dodds et al., 1996). Both proteins were expressed with a GST tag and were accordingly named GST:NaS_6-RNase and GST:RNaseNE. Expected molecular weights of GST:NaS_6-RNase and GST:RNaseNE were about 49 kDa (Figure 3.1).

Figure 3.5A shows expression of GST:NaS_6-RNase and GST:RNaseNE constructs in *E. coli*. The GST:NaS_6-RNase observed in the soluble fraction of cells grown at 30^{o}C was sometimes larger than the expected 49 kDa although this result was not highly reproducible and more commonly no soluble protein was observed. Similarly no GST:RNaseNE was detectable in the soluble fraction. For both recombinant proteins, multiple bands were detected in the insoluble fraction indicating degradation of the protein was occurring at the COOH-terminal end. The largest proteins were about 42 kDa in size, which is smaller than the expected sizes of GST:NaS_6-RNase and GST:RNaseNE.

Production of soluble protein was possible in cells grown at 16^{o}C after induction. Figure 3.5B shows a typical time course for GST:NaS_6-RNase accumulation. Under this growth condition (with or without added glucose), the only recombinant protein to accumulate in the soluble fraction was a protein of approximately 30 kDa that presumable represented the GST tag. As it seemed unlikely GST:NaS_6-RNase could be expressed in a soluble intact form in *E. coli*, no further work was done with this protein.

Figure 3.5C shows the accumulation over time of the soluble form of GST:RNaseNE in *E. coli* cells grown at 16^{o}C before and post-induction. A protein of the expected size for GST:RNaseNE was detected 2 hr after induction and levels of this protein increased over time. By 4 hr post-induction, two smaller proteins of 32 and 27 kDa were observed in cells grown in medium containing added glucose. Less degradation occurred in cells grown without added glucose. In addition to increased

degradation, the inclusion of glucose in the medium also seemed to result in reduced production of soluble GST:RNaseNE.

Figure 3.6 shows expression of five $(His)_6$-tagged DD ($(His)_6$:DD2 (S_1 allele), 5, 6, 7 and 8 (S_2 allele)) constructs in *E. coli*. Expected sizes for the recombinant proteins are all about 50 kDa and bands of close to this size were detected although some (notably $(His)_6$:DD2, 5 and 8) migrated faster than expected. After induction and growth under optimal conditions for expression (16^{o}C with 3% glucose), soluble and insoluble forms of the recombinant DD proteins were detected, with majority of the protein being in the insoluble fraction. The apparent yields of soluble recombinant protein varied, with $(His)_6$:DD8 producing the least and $(His)_6$:DD6 the most. Trial purifications of one of the soluble DD proteins showed it behaved in a similar manner to $(His)_6$:PiSLF1 and remained tightly bound to the Ni-NTA beads after elution (data not shown).

In summary, the only recombinant proteins produced in reasonable amounts in an intact and soluble form were $(His)_6$:PiSLF1, some $(His)_6$:DDs and GST:RNaseNE. Soluble GST:PiS_2-RNase was also produced but appeared unstable and existed as a mixture of truncated and apparently full-length forms. A truncated form of GST:RNaseNE was also produced. None of these proteins could be enriched using appropriate affinity chromatography methods.

3.3.3: *In vitro* binding assays

Various *in vitro* binding assays were developed to examine the interaction between $(His)_6$:PiSLF1 and GST:PiS_2-RNase. Because purification of the recombinant proteins was not successful, these assays used the soluble bacterial cell lysates instead. Initial experiments were a variation of the pull-down assay described by Hua and Kao (2006) that used glutathione sepharose beads to which a GST:PiS_2-RNase from soluble lysate was bound. A $(His)_6$:PiSLF1-containing soluble lysate was then mixed with the loaded beads to test if $(His)_6$:PiSLF1 could be pulled down by GST:PiS_2-RNase: immunoblots probed with tag-specific antibodies were used to detect proteins that remained bound to the beads following extensive washing.

Figure 3.7 shows the results of two such pull-down experiments. In Figure 3.7A the anti-$(His)_6$ tag and anti-GST antibodies detected $(His)_6$:PiSLF1 and GST:PiS_2-RNase in the respective soluble cell lysates, although the 50 kDa GST:PiS_2-RNase band was the least abundant of the three recombinant protein forms. $(His)_6$:PiSLF1 stayed on the glutathione beads when GST:PiS_2-RNase was present and did not interact with the beads non-specifically: the 26 kDa GST tag bound to the beads as expected. This argues that GST:PiS_2-RNase can pull down $(His)_6$:PiSLF1 from an *E. coli* soluble cell lysate. However, reproducibility of this result was poor and seen in only three of 11 assays. Figure 3.7B

shows a pull-down experiment where the 50 kDa GST:PiS_2-RNase band was at least as abundant as the other two recombinant protein forms in the soluble cell lysate. In this experiment, when $(His)_6$:PiSLF1 soluble lysate was added only the 26 kDa GST:PiS_2-RNase band remained on the beads. In both pull-downs, the supernatant from the last washing step before elution was immunoblotted to detect residual recombinant proteins. As no proteins were detected, washing presumably had removed all unbound proteins from the beads (Figure 3.7). Proteins bound to the glutathione beads were therefore present because of interactions between components in the pull down and not incomplete washing. Because of poor reproducibility, the pull-down assay was abandoned.

As neither *N. alata* S_6-RNase nor *P. inflata* S_2-RNase was made in a mostly intact and soluble form by *E. coli*, a second *in vitro* binding assay was developed that used native S-RNases from *N. alata* style extracts instead. This assay was called the co-immunoprecipitation or Co-IP assay and was performed using either the anti-$(His)_6$ tag antibody or an anti-S-RNase antibody (antibodies for the *N. alata* S_2- and S_7-RNases were available). An extract containing the recombinant protein to be tested was mixed with a stylar extract and the proteins allowed to interact. Either anti-$(His)_6$ or anti-S-RNase antibody was incubated with protein A beads and the protein A antibody beads then incubated with the mixture of style and bacterial extracts. The beads were washed to remove unbound proteins and immunoblots used to detect proteins that remained on the beads.

Figure 3.8A shows a Co-IP assay performed using the anti-$(His)_6$ tag antibody and extracts from $(His)_6$:PiSLF1-expressing *E. coli* and *N. alata* S_7S_7 styles. $(His)_6$:PiSLF1 and S_7-RNase were detected in the relevant extracts and neither $(His)_6$:PiSLF1 nor S_7-RNase bound non-specifically to the beads if the anti-$(His)_6$ antibody was omitted. When both extracts and anti-$(His)_6$ antibody were present, $(His)_6$:PiSLF1 and S_7-RNase remained on the beads, suggesting an interaction between the two proteins. Figure 3.8B is a duplicate Coomassie stained gel showing that none of the proteins in the Co-IP remained on the protein A beads when the anti-$(His)_6$ antibody was omitted, suggesting that the washing steps removed all non-specifically bound proteins. Proteins that remained on the beads when anti-$(His)_6$ antibody was present were tentatively identified based on their sizes and included $(His)_6$:PiSLF1, S_7-RNase and the anti-$(His)_6$ antibody heavy (approximately 50 kDa) and light (approximately 25 kDa) chains. Additionally, a protein of approximately 60 kDa (indicated in Figure 3.8B) seen on the Coomassie-stained gel was of probable bacterial origin. This protein was only observed in Co-IPs when antibody was also present.

Figure 3.8C shows a similar Co-IP performed using the anti-S_7-RNase antibody. When the anti S_7-RNase antibody was omitted, neither $(His)_6$:PiSLF1 nor S_7-RNase remained bound to the protein A beads. The approximately 50 kDa protein in this lane weakly detected with the anti-$(His)_6$ antibody is

presumably either from *E. coli* or the stylar extract. When both extracts and the anti-S_7-RNase antibody were added, $(His)_6$:PiSLF1 and S_7-RNase remained on the beads, confirming the interaction seen in Figure 3.8A. In summary, Co-IPs detected an interaction between $(His)_6$:PiSLF1 and *N. alata* S_7-RNase when antibodies to either protein were loaded onto protein A sepharose beads.

To extend this finding, Co-IP assays were performed using *E. coli* extracts containing one of $(His)_6$:DD5, 6 or 7, an S_7S_7 stylar extract and the anti-$(His)_6$ antibody (Figure 3.9). $(His)_6$:DD5, $(His)_6$:DD6 and the S_7-RNase were detected as single bands of the expected size in the relevant extract, and $(His)_6$:DD7 was detected as two bands, one being full-length and the other slightly smaller. All $(His)_6$:DDs bound to the protein A beads when anti-$(His)_6$ tag antibody was present. When the S_7S_7 stylar extract was also present S_7-RNase remained bound to the beads, suggesting that S_7-RNase was able to bind to all of the $(His)_6$:DDs tested. As no proteins were detected in the last washes prior to elution (Figure 3.9, middle panels), the presence of these proteins was not due to incomplete washing of the beads. Within the limits of this experiment, there was no obvious difference in the binding abilities of any of the tested $(His)_6$:DDs for S_7-RNase, as the $(His)_6$:DD bands in each lane were roughly similar in intensity and roughly similar amounts of S_7-RNase remained on the beads. Another protein retained on the beads and faintly detected all lanes was the antibody heavy chain (marked by an arrow in Figure 3.9). An equivalent coomassie-stained gel shows the major protein bands correspond to the $(His)_6$:DDs, S_7-RNase and antibody heavy and light chains. The 60 kDa *E. coli* protein that remained on the beads when antibody was present, was also seen.

Figure 3.10 shows Co-IP assays performed with *E. coli* extracts containing one of $(His)_6$:DD2 (S_1 allele), 7 or 8 (S_2 allele), S_7S_7 or S_2S_3 stylar extracts and the anti-$(His)_6$ antibody. $(His)_6$:DD2, $(His)_6$:DD8, the S_2- and S_7-RNases were detected as single bands of the expected size in the relevant extracts and $(His)_6$:DD7 was detected as a doublet as in Figure 3.9. S_3-RNase is only weakly detected by the anti-S_2-RNase antibody (this is described further in the next chapter in Figure 4.7). Figure 3.10A shows that S_7-RNase was retained on the beads by $(His)_6$:DD2 and $(His)_6$:DD8 as well as $(His)_6$:DD7, suggesting that S_7-RNase was bound by all of the tested DDs (DDs 2, 5, 6, 7 and 8; DD1 from S_1 allele and the rest of the DDs from S_2 allele). As in Figure 3.9, no obvious difference was observed in the binding abilities of any of these $(His)_6$:DDs for S_7-RNase. Figure 3.10B shows that $(His)_6$:DDs 2 (S_1 allele), 7 and 8 (S_2 allele) were also able to bind S_2-RNase. These proteins could all be identified on an equivalent coomassie-stained gel along with other proteins (antibody heavy chain and the 60 kDa *E. coli* protein) that had previously been shown to remain on the beads. No proteins were detected in the final washes before proteins were eluted.

A series of Co-IPs was performed to test the binding of $(His)_6$:PiSLF1, $(His)_6$:DD2 or $(His)_6$:DD5 to the negative control GST:RNaseNE (Figure 3.11). $(His)_6$:PiSLF1, $(His)_6$:DD2 and $(His)_6$:DD5 were all detected as single bands of the expected size in the relevant extracts and the interaction between $(His)_6$:PiSLF1 and S_7-RNase seen earlier (Figure 3.8) was confirmed. Consistent with previous results (Figure 3.5), a range of full-length and shorter forms of GST:RNaseNE was detected in the relevant *E. coli* extract. When $(His)_6$:PiSLF1, $(His)_6$:DD2 or $(His)_6$:DD5 and anti-$(His)_6$ antibody were present in the Co-IP, GST:RNaseNE was retained on the protein A beads. However, GST:RNaseNE was also retained when only anti-$(His)_6$ antibody was present or when no antibody was present, indicating GST:RNaseNE interacted non-specifically with the protein A beads. Although binding appeared to be enhanced by the presence of $(His)_6$:PiSLF1, $(His)_6$:DD2 and $(His)_6$:DD5, the fact that GST:RNaseNE also bound non-specifically to the protein A column made it difficult to interpret the experiment unambiguously.

In summary, the *in vitro* binding assays showed interactions between $(His)_6$-tagged PiSLF1, all tested $(His)_6$-tagged DDs and native *N. alata* S_2- and S_7-RNases. $(His)_6$-tagged proteins also appeared to bind to GST-tagged RNaseNE, even though this protein has no known role in the self-incompatibility response. There were numerous problems, however, including non-specific interactions and low yields of soluble $(His)_6$-tagged proteins that could not be purified with NTA columns. For these reasons, further work was done to improve recombinant protein solubility and to develop an alternative to GST:RNaseNE as a negative control.

3.3.4. Modifying the PiSLF1 construct to improve protein solubility

As very little of the *E. coli*-expressed $(His)_6$:PiSLF1 or other $(His)_6$-tagged proteins was able to be purified in a soluble form, a range of approaches known to aid in the production of soluble proteins was attempted (Sorensen and Mortensen, 2005), some of which (manipulation of growth conditions) have already been described and others (use of different *E. coli* strains and co-expression with molecular chaperones) that have not. One reason recombinant proteins form insoluble aggregates is because kinetic barriers prevent the folding of sub-domains, resulting in an accumulation of partially folded species with exposed hydrophobic 'sticky' surfaces that promote self-association (Georgiou and Valax, 1996). Thus, one possible way of producing more soluble protein is to delete individual sub-domains that are disordered when expressed in *E. coli*. Before this could be done for PiSLF1 or the DDs, however, it was first necessary to understand the protein's likely overall structure.

F-box proteins have a bipartite structure with an amino terminal F-box motif that mediates binding to Skp1p and a carboxy-terminal protein-protein interaction domain that recruits the target of the

SCF complex. The second domain is often composed of a repeat motif (e.g., Gagne et al., 2002) and an initial search of the DDs and PiSLF1 indicated a region with weak homology to a partial Kelch repeat (Figure 3.12). Iterative searching with this sequence uncovered further copies of the repeat in the DDs and PiSLF1, with PSI-Blast searches using DD query proteins identifying similarities to Kelch-repeat proteins from *Arabidopsis*. Iterative Blast searches and manual sequence comparisons identified six repeats in the DDs and PiSLF1, although these had low similarity to a canonical four-bladed Kelch repeat such as that of the mouse Kelch-repeat protein Keap1 (Figure 3.12A). Importantly, signature residues of a Kelch repeat, such as the diglycine (GG) doublet and reasonably well−conserved Tyr (Y), Trp (W) and Arg (R) residues (Chen et al., 2011), were largely missing from the DD and SLF repeats. However, consistent with the presence of a Kelch-like repeat, secondary structure predictions of the DDs and PiSLF1 indicate that their COOH-terminal ends contain extended regions of β-strand, with each region corresponding to one of the four blades of the repeat structure (Figure 3.12B).

Figure 3.12C shows that PiSLF1 with the F-box domain removed can be folded into a protein with the same overall shape and structure as Keap1. The six-bladed propeller appears as disc when viewed from above, with the NH_3- and COOH-terminal ends closed to complete the β-propeller by a split Kelch repeat made up of two β-strands from each end that comes together like a tight clasp.

Given this possible structure, the only sub-domain that could be deleted from PiSLF1 and the DDs without destroying the β-propeller was the N-terminal F-box motif. In addition, some sequences introduced during the initial cloning of $(His)_6$:PiSLF1, such as those encoding protease cleavage sites for subtilisin and thrombin, could also be removed (Figure 3.1). The resulting constructs for PiSLF1 were $(His)_6$:PiSLF1 FL (with just the protease cleavage sites deleted) and $(His)_6$:PiSLF1 FBD (with the cleavage sites and F-box motif deleted).

Figure 3.13A shows $(His)_6$:PiSLF1 FL expression in the *E. coli* strain BL21 codon plus (RIL), a strain that contains extra copies of the *argU*, *ileY*, and *leuW* tRNA genes. $(His)_6$:PiSLF1 FL expression was observed but the majority of the approximately 45 kDa protein was in the insoluble fraction and only a small amount (in BL21 cells) was soluble. Attempts to purify $(His)_6$:PiSLF1 FL from the soluble fraction on NTA beads were unsuccessful, as most of the protein remained on the beads and the small amounts of soluble protein that did elute were not sufficient for further work (Figure 3.13B).

Figure 3.14 shows expression of the other truncated protein, $(His)_6$:PiSLF1 FBD, in *E. coli* BL21 star cells, a strain with enhanced mRNA stability due to a mutation in the RNaseE gene. $(His)_6$:PiSLF1 FBD was observed as a major protein of approximately 38 kDa in the insoluble fraction. A similar

construct expressing a truncated version of DD6 lacking the F-box domain coding region, named $(His)_6$:DD6 FBD, was also expressed in BL21 star (Figure 3.14A). Intact $(His)_6$:DD6 FBD was also detected as the major protein in the insoluble fraction. No soluble form of either protein was detected.

Figure 3.14B shows $(His)_6$:PiSLF1 FBD expressed in *E. coli* Arctic Express cells grown at 13°C post-induction. This strain overexpresses two cold-adapted chaperonins Cpn60 and Cpn10, and extra copies of several rare tRNA genes. Intact $(His)_6$:PiSLF1 FBD was detected in the soluble and insoluble fractions, but was not a major protein in the soluble fraction.

Figure 3.15 shows that some of the soluble $(His)_6$:PiSLF1 FBD from the Arctic Express cells could be eluted from the NTA beads although most of the protein remained bound. However, a comassie-stained gel of each fraction of the purification shows that the eluted protein fraction was heavily contaminated and $(His)_6$:PiSLF1 FBD was not the major protein in it. Based on band intensity, it is possible $(His)_6$:PiSLF1 FBD represented about 10% of the total protein in the fraction. Attempts at further purification of this protein were unsuccessful. Hence no further work was done with the soluble $(His)_6$:PiSLF1 FBD fraction and efforts at producing a soluble, pure form of $(His)_6$:PiSLF1 FBD were re-focussed, as discussed in Chapter 4.

3.4: Discussion

The work described in this chapter began before the publication of Kubo et al., (2010) and the advent of the collaborative recognition model. The original intention was to identify which of the *N. alata* DD proteins was the functional ortholog of *P. inflata* SLF using recombinant proteins and the *in vitro* binding assay described in Hua and Kao (2006) and Hua et al., (2007). Using $(His)_6$:PiSLF1 and GST:PiS_2-RNase constructs similar to those in these papers, a major problem was that the recombinant proteins were largely insoluble and often not full-length due to proteolytic degradation. For example, expressing GST:PiS_2-RNase in *E. coli* produced three different molecular weight products, most likely because of proteolysis at the COOH-terminal as these species were identified by immunoreactivity towards the N-terminal GST tag. The recombinant proteins could also not be enriched by affinity chromatography and were very 'sticky', presumably because of exposed hydrophobic regions. However, other aspects of Hua and Kao (2006) and Hua et al., (2007) could be reproduced, at least in part, such as the interaction between SLF and S-RNase.

Binding assays were performed using total *E. coli* lysates in a commercial bugbuster master mix, which contains non-ionic detergents and is designed for gentle lysis of bacteria cell for maximium protein solubility. As this buffer should keep recombinant soluble during the interaction assay,

dialysis into the buffer used in Hua and Kao (2006) was not carried out. It is possible that the detergents present in the bugbuster master mix interfered with interactions with S-RNase but it is the recommended binding buffer for NTA beads and is designed for protein work.It was not unexpected that expressing S-RNases in *E. coli* would be problematic, as the *E. coli* cytoplasm is a reducing environment in which the four pairs of disulphide bonds that stabilise the S-RNase's tertiary structure are unlikely to form (Ida et al., 2001). Work on expressing S-RNases from various species in *E. coli* has also consistently found the proteins made are inactive and incorrectly folded (Professor Hidenori Sassa, Chiba University, personal communication; Liu, 1993). Indeed, previous work on heterologous expression of *P. inflata* S-RNases noted that in order for these proteins to be made in an enzymatically active form, it was necessary for them to be expressed in a system able to correctly form the intramolecular disulphide bonds (Mu and Kao, 1992). Although the effect of removing the NH_3-terminal signal peptide from PiS_2-RNase on expression can be tested, the solubility and degradation problems experienced here with GST:PiS_2-RNase are consistent with previous work and suggest this was not the source of problems, especially as similar problems of degradation and solubility were experienced with *N. alata* S_6-RNase and RNase NE, neither of which had signal peptides (Figure 3.5). Interestingly, Hua and Kao (2008) reported that *P. inflata* S_3-RNase, made by *E. coli* was enzymatically active. As these authors used an in gel RNase assay to measure activity, it would be interesting to know whether the purified recombinant protein itself was active (e.g., using the spectrophotometric RNase assay described by McClure et al., (1989)), or whether the protein was inactive but able to refold into an active form during the washing steps that are part of the in gel assay.

Full-length, soluble $(His)_6$:PiSLF1 was only obtained in small amounts as the majority of the protein remained bound tightly to the NTA column, an indicator of protein 'stickiness'. Heterologous expression of a gene with codons rarely used by *E. coli* is likely to lead to translational problems and a reduction in either the quantity or quality of the protein being synthesized (Kane, 1995). The rare codon usage calculator (http://nihserver.mbi.ucla.edu/RACC) shows that $(His)_6$:PiSLF1 FL has 15 codons that are rarely used by *E. coli*, specifically the Pro codon CCC (2), the Ile codon AUA (6), the Leu CUA (1) and the Arg codons CGA, AGA and AGG (6 in total). The presence of numerous rare codons, especially Arg codons, presents an obvious translational problem for which one solution was to express $(His)_6$:PiSLF1 in *E. coli* strains that supply these codons, such as a RIL-containing strain. However, Figure 3.13A shows that expressing $(His)_6$:PiSLF1 FL in such strains did not bring about a major improvement in the yield of soluble protein. Testing a range of other *E. coli* strains, growth conditions and constructs was equally unsuccessful in improving the yield of soluble protein. Similarly, Qiao et al., (2004b) reported that, despite repeated efforts, they were unable to obtain

from *E. coli* a full-length version of an SLF from the plant *Antirrhinum hispanicum*. As with the experiments reported here, this protein was also made with an N-terminal $(His)_6$-tag. Thus, while optimising the *PiSLF1* sequence for rare *E. coli* codon usage could be done, it is unlikely this will result in an increased yield of soluble protein. Deleting particular domains within the protein appeared more likely to be successful.

Plant genomes encode an impressive variety of modular F-box proteins with an NH_3-terminal F-box motif and a diverse array of COOH-terminal interaction domains (Gagne et al., 2002). Among the most abundant classes of interaction domain are those that contain WD and Kelch repeats. Both repeat types form four-stranded, antiparallel β-sheets (Hudson and Cooley, 2008; Chen et al., 2011). When sufficient repeats ('blades') are present they can be arranged in a disc around a central axis to generate a β-propeller structure that is closed through interactions between the NH_3- and COOH-terminal ends of the repeat region. Despite being structurally similar, WD- and Kelch-repeat proteins are unlikely to have evolved from a common ancestor and are classified into separate protein families. Analysis of the SLFs shown in Figure 3.12 detected the presence of six copies of a Kelch-like repeat motif in the interaction domain, and this part of PiSLF1 can be folded into a six-bladed β-propeller structure using as a template the Kelch repeat protein Keap1 (Li et al., 2004). Chen et al., (2012) recently reached the same conclusion regarding a β-propeller structure in the SLF interaction domain, but proposed instead a structure based on six copies of the slightly shorter WD repeat. However, most WD repeat proteins have seven-bladed rather than six-bladed propellers (Hudson and Cooley, 2008; Chen et al., 2011). Indeed, the only six-bladed, WD repeat protein known is the *Saccharomyces cerevisiae* protein Sec13, which is an open, six-bladed β-propeller, the ends of which are closed by the insertion of a single β-blade from another protein (Brohawn et al., 2008). Although there is relatively little structural information available for Kelch proteins generally, the β-propellers of these proteins are generally six-bladed except for the case of galactose oxidase from the fungus *Dactylium dendroides*, which is seven-bladed (Ito et al., 1991; Chen et al., 2011). Thus a six-bladed β-propeller based on a Kelch-like repeat is a more likely structure for the SLFs than one based on six WD repeats, although both models are still highly speculative.

Although only weak similarity to the Kelch repeat was seen in PiSLF1 and the DDs, identifying Kelch repeats from primary sequence alone is problematic. The Kelch repeat is defined by a handful of conserved residues and substitutions are known to occur in these positions (Hudson and Cooley, 2008). For example, Kelch-related protein 1 Krp1 has only five Kelch repeats but still forms a six-bladed propeller using a non-Kelch amino acid sequence to make one of the blades (Gray et al., 2009). Of the Kelch repeats present in the other five blades of Krp1, variation is seen in the GG

dipeptide of B2 and the Y residue of B3. The Kelch repeats of the *Caenorhabditis elegans* protein SPE-26 are even more degenerate with variation seen in most of the signature residues that define the motif (Varkey et al., 1995). The variation seen in Kelch proteins suggests that a broad range of residues can produce the hydrophobic interactions needed to produce a blade.

The β-propeller structure is a stable scaffold that potentially allows interactions with proteins and other ligands to occur on various surfaces. Unfortunately, only a few of the structures of a Kelch protein interacting with peptides from binding partners have been solved. However, these structures show binding sites can exist across much of the structure. For example, Keap1 interacts with its binding partners through loops on the underside of the propeller (Lo et al., 2006; Padmanabhan et al., 2008) and Krp1 has one binding site preceding the first blade and another within the last blade of the β-propeller (Gray et al., 2009). Thus, while it may be possible to improve protein solubility by deleting various regions from the COOH-terminal end of PiSLF1, this could result in the production of a non-functional protein lacking the S-RNase binding site or with a non-native structure. The F-box deleted forms of PiSLF1 and the DDs therefore seem to represent the minimum length of protein that can be expressed and still retain binding activity.

The binding assay described by Hua and Kao (2006) could not be performed for the reasons described above. Hence, to study the interaction of SLFs and S-RNases, alternative assays were developed, with the Co-IP assay that used *E. coli* and *N. alata* extracts containing various recombinant SLFs and S-RNases being the most reproducible. Using this assay, an interaction was shown between a *Petunia* SLF and two *N. alata* S-RNases. Intuitively, there seems no reason why the SLFs of *Petunia* should recognize *Nicotiana* S-RNases, given the taxonomic distance separating these two solanaceous genera (Olmstead et al., 2008). However, a feature of solanaceous S-RNases is that an allele from one species is often more closely related in sequence to an allele from another species than to other alleles of its own species, a feature known as trans-specific polymorphism (Ioerger et al., 1990). Abundant trans-specific polymorphisms among Solanaceae S-RNases are evidence that the common ancestor of this family was self-incompatible and already possessed much of the extant allelic diversity that has been passed down to its descendants (Paape et al., 2008). Thus, the ability of *Petunia* SLFs to recognize *Nicotiana* S-RNases is entirely expected, given the great age of *S* alleles and the extreme sequence diversity that exists among the S-RNase within a species. An even more extreme example of this same principle is reported in Qiao et al., (2004b), who showed an interaction between a recombinant *A. hispanicum* SLF and an S-RNase from *P. hybrida* styles. *Antirrhinum* and *Petunia* are members of two separate plant families (Plantaginaceae and Solanaceae, respectively).

As well as the interaction between PiSLF1 and the *N. alata* S-RNases, interactions between *N. alata* S-RNases and DDs 2 (S_1 allele), 5, 6, 7 and 8 (S_2 allele) were also detected with the Co-IP assay. The promiscuous nature of these interactions raised questions as to the specificity of the assay, especially as the negative control protein, RNase NE, also appeared to bind to PiSLF1 and some DDs (Figure 3.11). Because of their low solubility none of the recombinant proteins could be purified and the assay was instead performed with crude *E. coli* lysates. A 'false positive' interaction (i.e., one observed in vitro that does not occur in vivo) can result if one of the interacting partners is misfolded (Mackay et al., 2007; Wissmueller et al., 2011). This could be what was being observed in the Co-IP assay, especially as other indications of protein misfolding, such as low solubility and non-specific binding to various chromatography resins, were also noted (Fletcher et al., 2003; Risk et al., 2009). The interaction observed here is and its implications with the non-self collaborative new model cannot be made. Thus, for this interaction to be studied further it was first necessary to use purified recombinant proteins rather than crude extracts. The production of soluble $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD in an enriched form, and the interaction of these proteins with various S-RNases, forms the subject of the next chapter.

Figure 3.1: Outline of the recombinant proteins used in this chapter.

The *P. inflata* S_2-RNase coding region (including signal peptide) was chemically synthesized and cloned into pGEX4T-1 to place a GST tag and thrombin cleavage site at the encoded protein's N-terminal end. The final protein was called GST:PiS_2-RNase and had an expected size of 52.3 kDa.

The coding region of *N. alata* S_6-RNase (without signal peptide) was cloned into pGEX4T-1, placing a GST tag and thrombin cleavage site at encoded protein's N-terminal end. The final protein was called GST:NaS_6-RNase and had an estimated size of 48.6 kDa.

The coding region of *N. alata* RNaseNE (without signal peptide) was cloned into pGEX4T-1, placing a GST tag and thrombin cleavage site at encoded protein's N-terminal end. The final protein was called GST:RNaseNE and had an estimated size of 48.8 kDa.

The *P. inflata* SLF1 (PiSLF1) coding region was chemically synthesised and cloned into pET 30a, which placed a six Histidine ($(His)_6$) tag at the encoded protein's N-terminal end. Cleavage sites for thrombin and subtilisin tag were also incorporated. The final protein was called $(His)_6$:PiSLF1 and had an estimated size of 49.8 kDa. The F-box motif is indicated. The coding regions of various DD proteins were cloned into the same vector to yield recombinant proteins with an identical overall structure to $(His)_6$:PiSLF1 (not shown). The expected sizes of these proteins ranged from 49.1 to 49.8 kDa.

$(His)_6$:PiSLF1 FL is identical to $(His)_6$:PiSLF1 except that the coding region encompassing the cleavage site for thrombin and subtilisin has been removed. Estimated molecular weight is 45 kDa.

$(His)_6$:PiSLF1 FBD protein is identical to $(His)_6$:PiSLF1 FL except that the N-terminal F-box motif has been removed. Estimated molecular weight is 39.5 kDa.

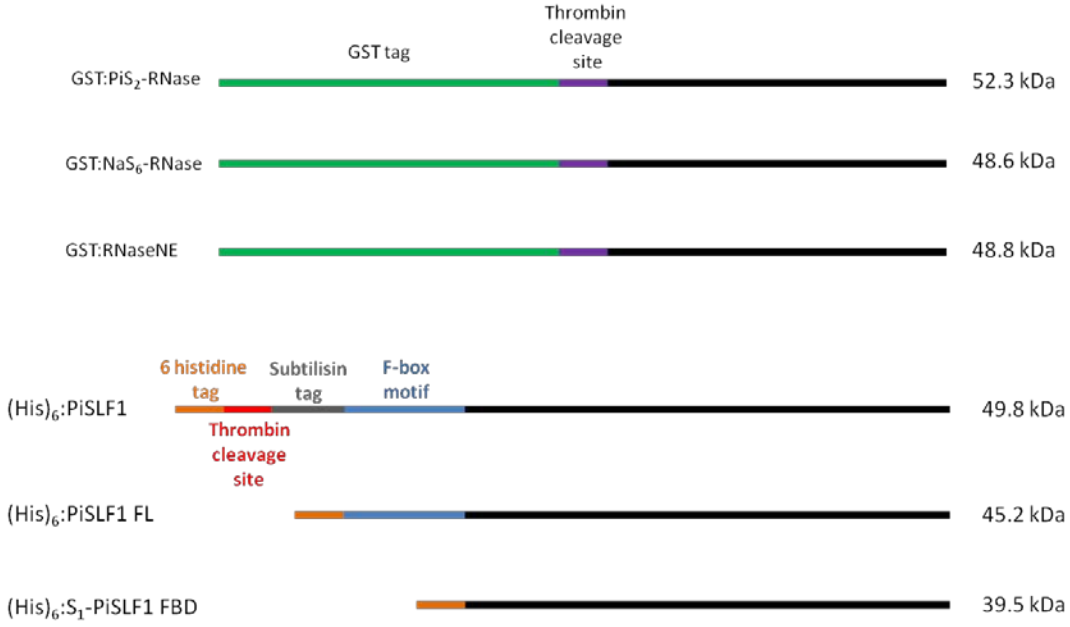
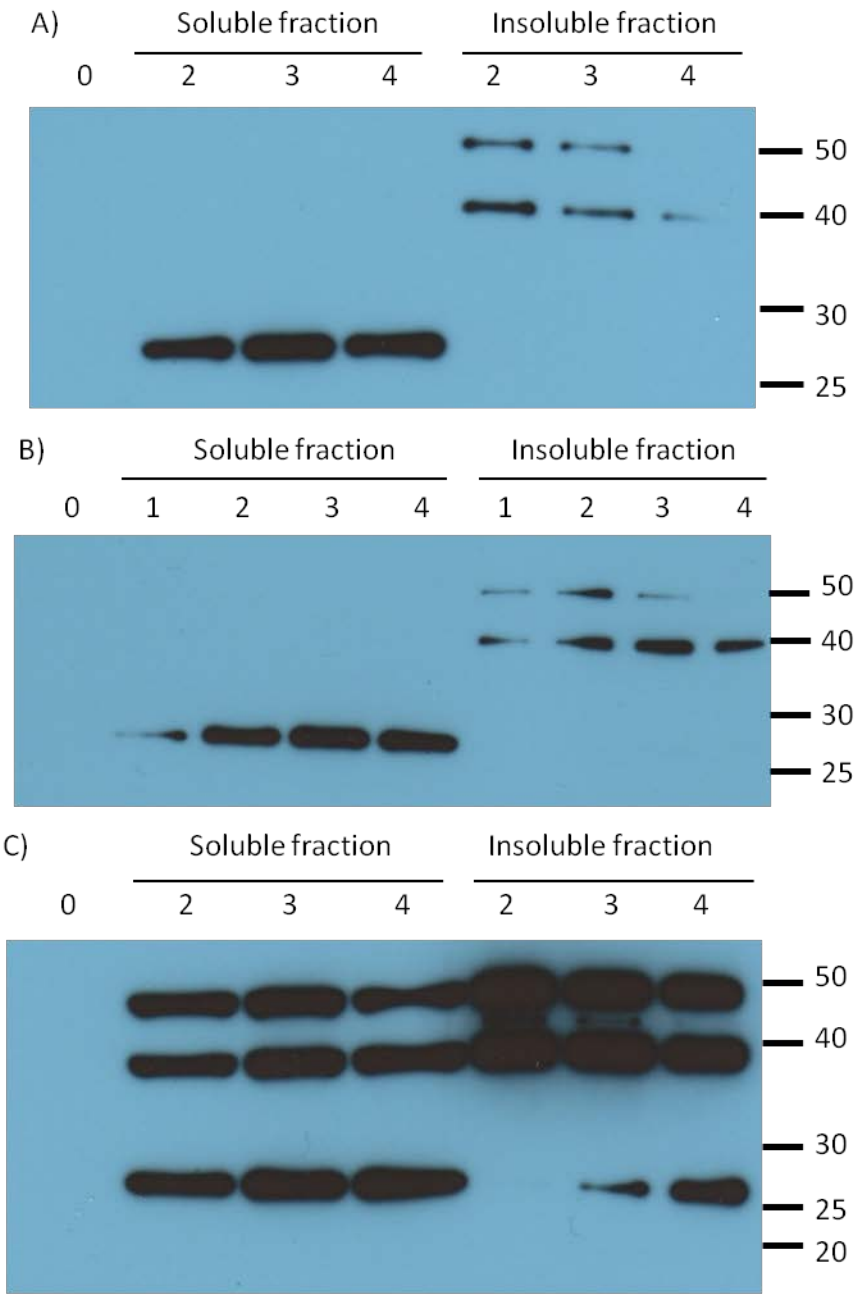

Thrombin cleavage site
GST tag
GST:PiS2-RNase
52.3 kDa
GST:NaS6-RNase
48.6 kDa
GST:RNaseNE
48.8 kDa
6 histidine tag
Subtilisin tag
F-box motif
(His)6:PiSLF1
Thrombin cleavage site
49.8 kDa
(His)6:PiSLF1 FL
45.2 kDa
(His)6:S1-PiSLF1 FBD
39.5 kDa

Figure 3.2: Optimising expression of GST:PiS_2-RNase in *E. coli*.

Detection of GST:PiS_2-RNase expression in *E. coli* by immunoblotting with an anti-GST antibody. Growth after IPTG induction was at 37°C (A), 25°C (B) or 16°C (C). Numbers above lanes represent hours post-induction; the 0 time sample was a total extract of cells prior to induction and extracts made at later time points were separated into soluble and insoluble fractions. The same amount of protein was loaded in each lane. Fractions were separated by SDS-PAGE and proteins detected using an anti-GST antibody. Numbers to the right of the figure are sizes in kDa.


A)
Soluble fraction
Insoluble fraction
0
2
3
4
2
3
4
50
40
30
25
B)
Soluble fraction
Insoluble fraction
0
1
2
3
4
1
2
3
4
50
40
30
25
C)
Soluble fraction
Insoluble fraction
0
2
3
4
2
3
4
50
40
30
25
20

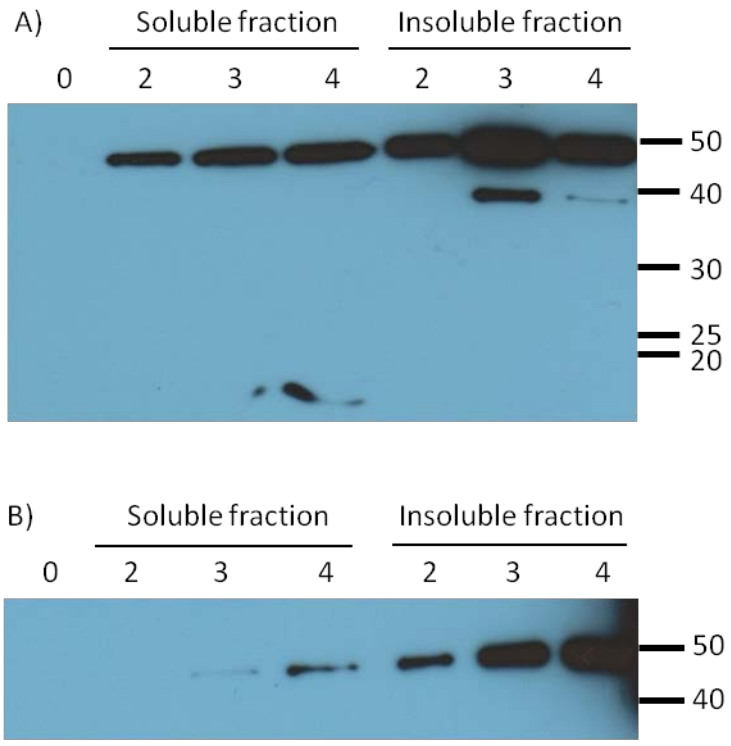


Figure 3.3: Optimising expression of $(His)_6$:PiSLF1 in *E. coli*.

Detection of recombinant $(His)_6$:PiSLF1 expression in *E. coli* by immunoblotting with an anti-$(His)_6$ antibody. Growth after induction was at 25°C (A) or 16°C (B). Other features are as described in the legend to Figure 3.2.

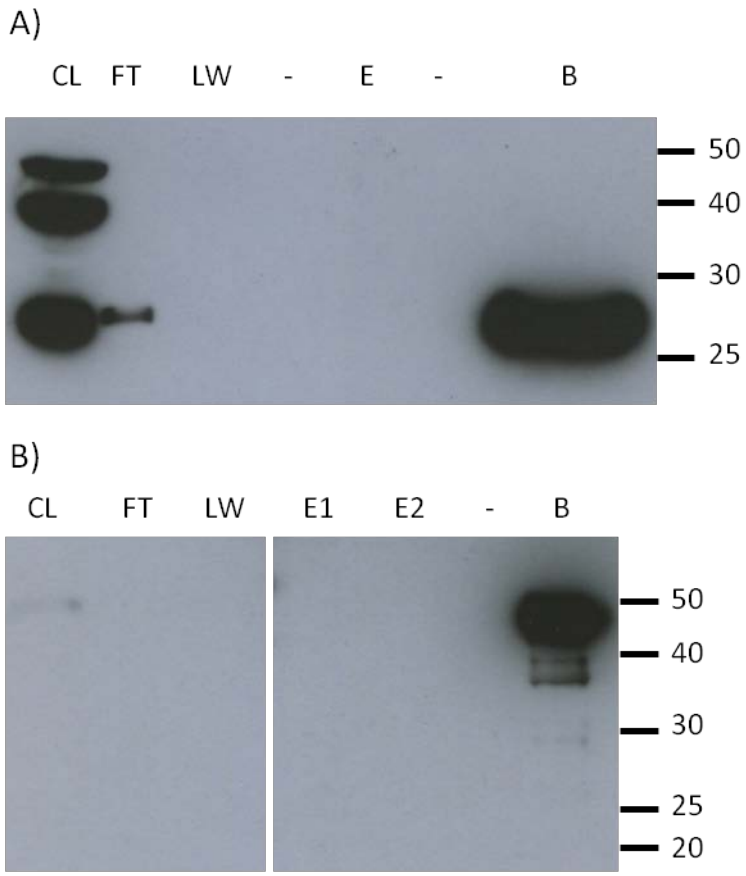


Figure 3.4: Recombinant GST:PiS$_2$-RNase and (His)$_6$:PiSLF1 purification.

A: Soluble GST:PiS$_2$-RNase was purified batch-wise using glutathione sepharose and the proteins present in the various fractions of the purification process were detected by immunoblotting using an anti-GST tag antibody. CL: soluble cell lysate; FT: flow-through; LW: last wash of resin before elution; E: elution fraction (20 mM reduced glutathione); B: protein remaining on the resin after elution; - empty lane.

B: Soluble (His)$_6$:PiSLF1 was purified batch-wise using NTA affinity beads and the proteins present in the various fractions were detected by immunoblot analysis using an anti-(His)$_6$ tag antibody. CL: total soluble cell lysate; FT: flow-through; LW: last wash of beads before elution; E1: first elution fraction (250 mM imidazole); E2: second elution fraction (250 mM imidazole); B: protein bound to the beads after elution; - empty lane. Numbers to the right of the figures indicate sizes (in kDa) of molecular weight markers.

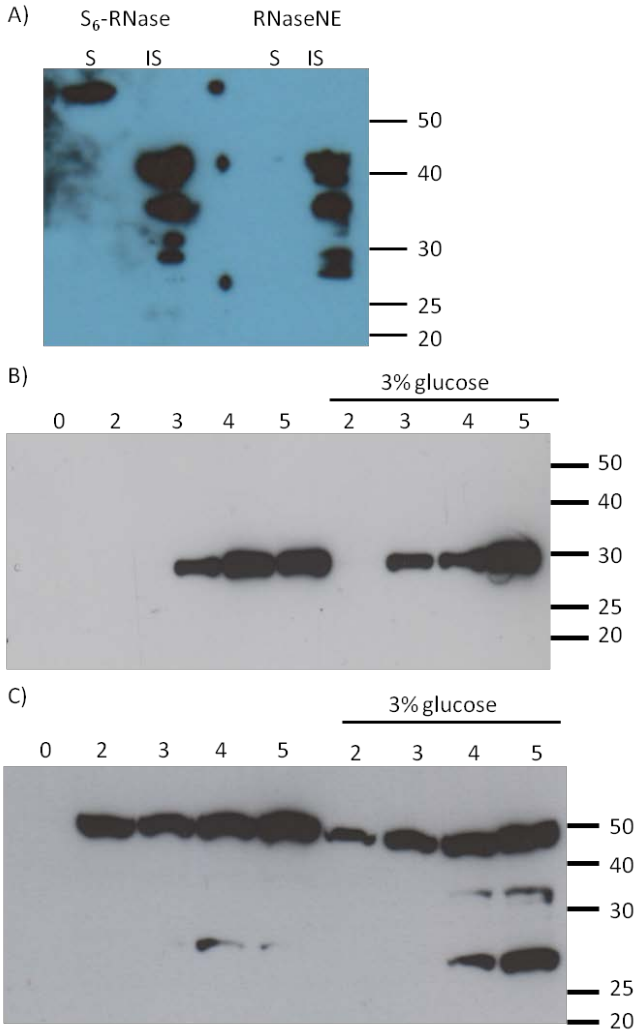


Figure 3.5: GST:RNaseNE and GST:NaS$_6$-RNase expression in *E. coli*.

A: Cells expressing GST:NaS$_6$-RNase and GST:RNaseNE were grown at 30^{o}C after induction. Cells were harvested 2 hr after induction and soluble and insoluble protein fractions were prepared. Proteins were detected by immunoblot analysis using an anti-GST tag antibody. S: soluble fraction; IS: insoluble fraction.

B: Cells expressing GST:NaS$_6$-RNase were grown with or without added glucose at 16°C after induction. Cells were harvested at the indicated time points after induction (hr) and soluble protein fractions were prepared. Proteins were detected by immunoblot analysis using an anti-GST tag antibody. Numbers to the right indicate the sizes (in kDa) of molecular weight markers.

C: Cells expressing GST:RNaseNE were grown harvested and fractionated as in B. Proteins were detected by immunoblot analysis using an anti-GST tag antibody. Numbers to the right indicate the sizes (in kDa) of molecular weight markers.

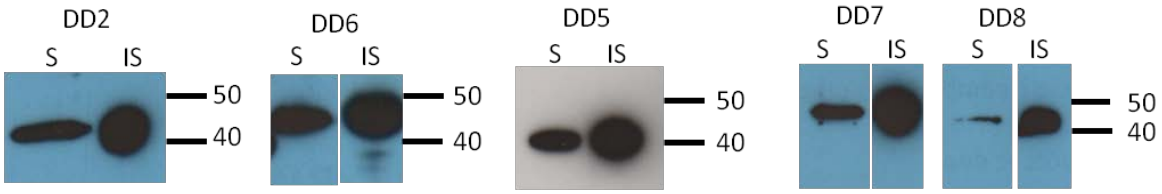


Figure 3.6: $(His)_6$:DD2, 5, 6, 7 and 8 expression in *E. coli*.

Expression of the indicated protein was induced with IPTG and cells subsequently grown at 16°C in medium supplemented with 3% glucose. Cells were harvested 16 hrs after induction and soluble (S) and insoluble (IS) fractions were prepared. Proteins were detected by immunoblotting with an anti-$(His)_6$ antibody. Numbers to the right of the figure indicate the sizes (in kDa) of molecular weight standards.

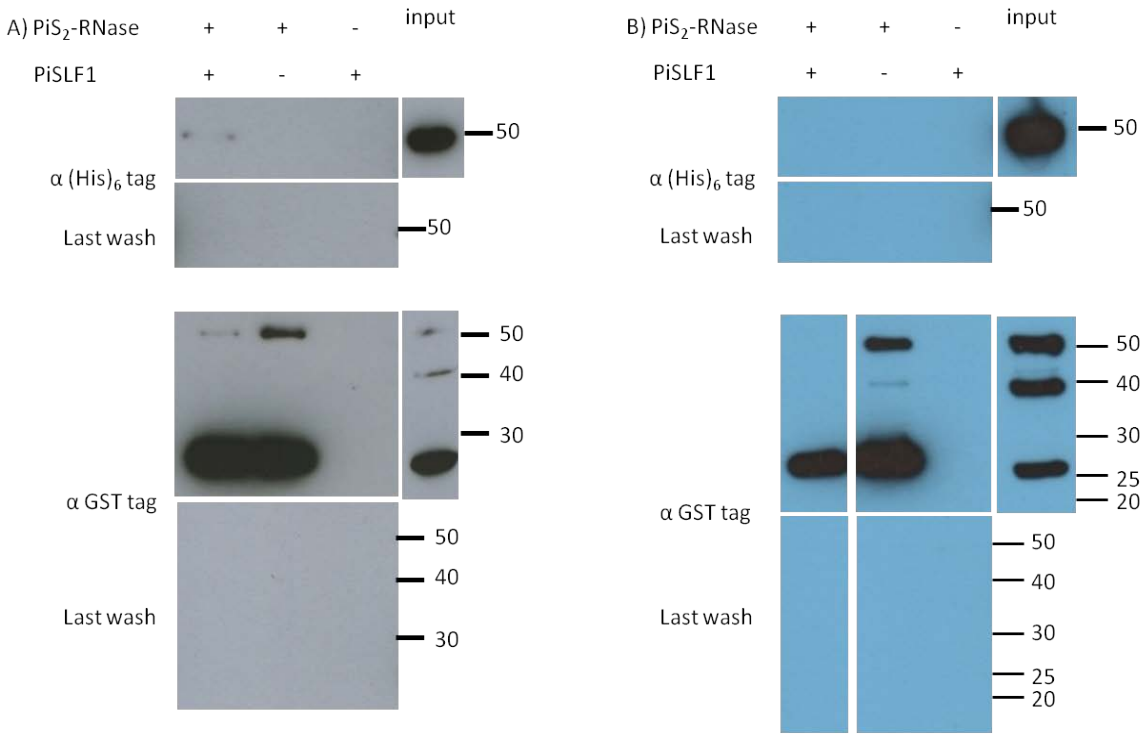


Figure 3.7: *In vitro* pull-down assays using $(His)_6$:PiSLF1 and GST:PiS_2-RNase.

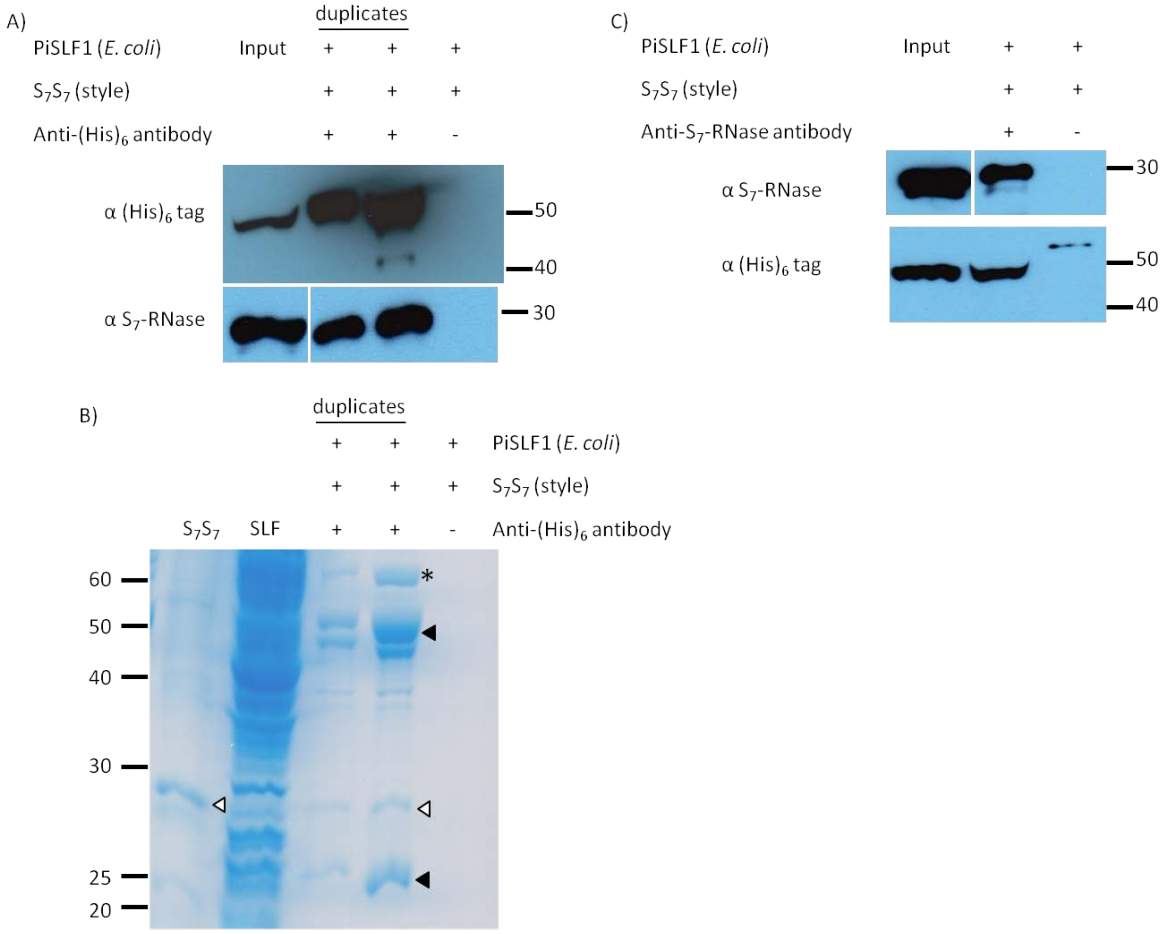
A) First example of a pull-down assay. Soluble lysates of *E. coli* cells expressing the indicated recombinant protein were incubated with glutathione Sepharose beads. Unbound proteins were removed by washing, the beads boiled in SDS loading buffer and the bound proteins separated by SDS-PAGE. The last wash was retained and also analysed. Immunoblots were probed with anti-GST antibody (to detect GST:PiS_2-RNase) or an anti-$(His)_6$ antibody (to detect $(His)_6$:PiSLF1). Input lanes show 1/10th the amount of cell lysate added to the pull-down assays.

B) Second example of a pull-down experiment done as described in A. Numbers to the right of the figure indicate the sizes (in kDa) of molecular weight standards.

Figure 3.8: Co-immunoprecipitation assays using $(His)_6$:PiSLF1 and native S_7-RNase from *N. alata* styles.

A) Co-IP performed with anti-$(His)_6$ tag antibody loaded onto the protein A beads analysed by immunoblotting; B) Duplicate gel of A stained with Coomassie; C) Co-IP performed with anti-S_7-RNase antibody loaded onto the protein A beads.

Extracts from the indicated source were incubated with protein A beads and the indicated antibody. Unbound proteins were removed by washing, the beads were boiled in SDS loading buffer and the bound proteins separated by SDS-PAGE. Immunoblots were probed with anti-S_7-RNase antibody (to detect S_7-RNase) or an anti-$(His)_6$ antibody (to detect $(His)_6$:PiSLF1). Selected proteins on the Coomassie-stained gel (B) are indicated: 60 kDa *E. coli* protein (*); antibody heavy and light chains (◀); and S_7-RNase (◁). Input lanes show $1/10^{th}$ the amount of extract added to the Co-IP assays. Numbers to the side of the figure indicate the sizes (in kDa) of molecular weight standards.


A)
duplicates
PiSLF1 (E. coli)
Input
+ + +
S7S7 (style)
+ + +
Anti-(His)6 antibody
+ + -
α (His)6 tag
α S7-RNase
50
40
30
B)
duplicates
+ + + PiSLF1 (E. coli)
+ + + S7S7 (style)
S7S7 SLF + + - Anti-(His)6 antibody
60
50
40
30
25
20
C)
PiSLF1 (E. coli)
Input
+ +
S7S7 (style)
+ +
Anti-S7-RNase antibody
+ -
α S7-RNase
α (His)6 tag
30
50
40

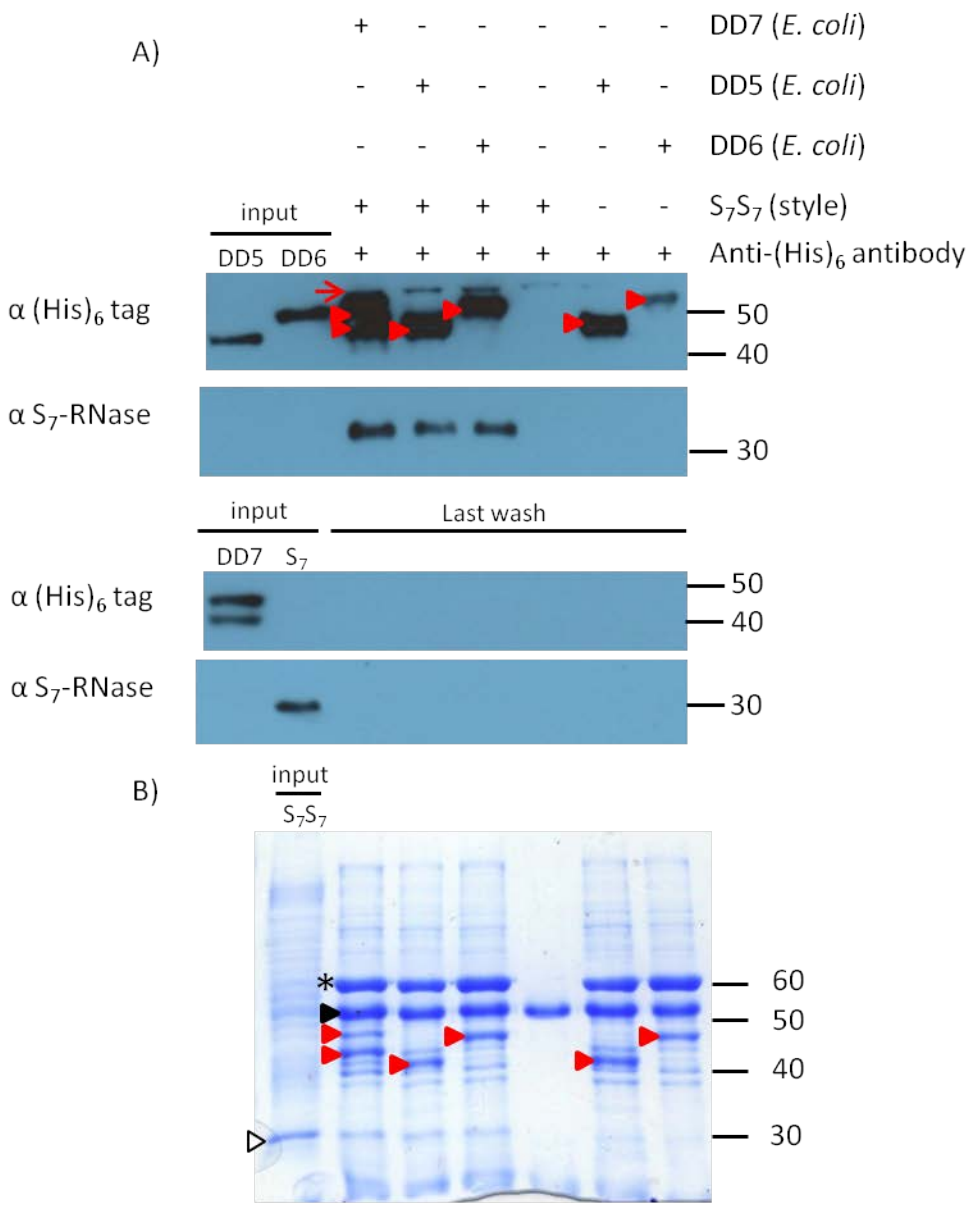


Figure 3.9: Co-immunoprecipitations using anti-$(His)_6$ tag antibody, various $(His)_6$-tagged DDs and native S_7-RNase.

Extracts from the indicated source were incubated with protein A beads loaded with anti-His tag antibody. Unbound proteins were removed by washing, the beads were boiled in SDS loading buffer and the bound proteins separated by SDS-PAGE. The last wash before elution was also retained for analysis. A) Immunoblots probed with anti-S_7-RNase antibody (to detect S_7-RNase) or anti-$(His)_6$ antibody (to detect $(His)_6$:DD proteins). B) Duplicated gel of A stained with Coomassie. Selected proteins in A and D are indicated: 60 kDa *E. coli* protein (*); antibody heavy chain (▶); $(His)_6$-tagged DDs (▶); and S_7-RNase (▷). Input lanes show 1/10th the amount of extract added to the Co-IP assays. Numbers to the side of the figure indicate the sizes (in kDa) of molecular weight standards.

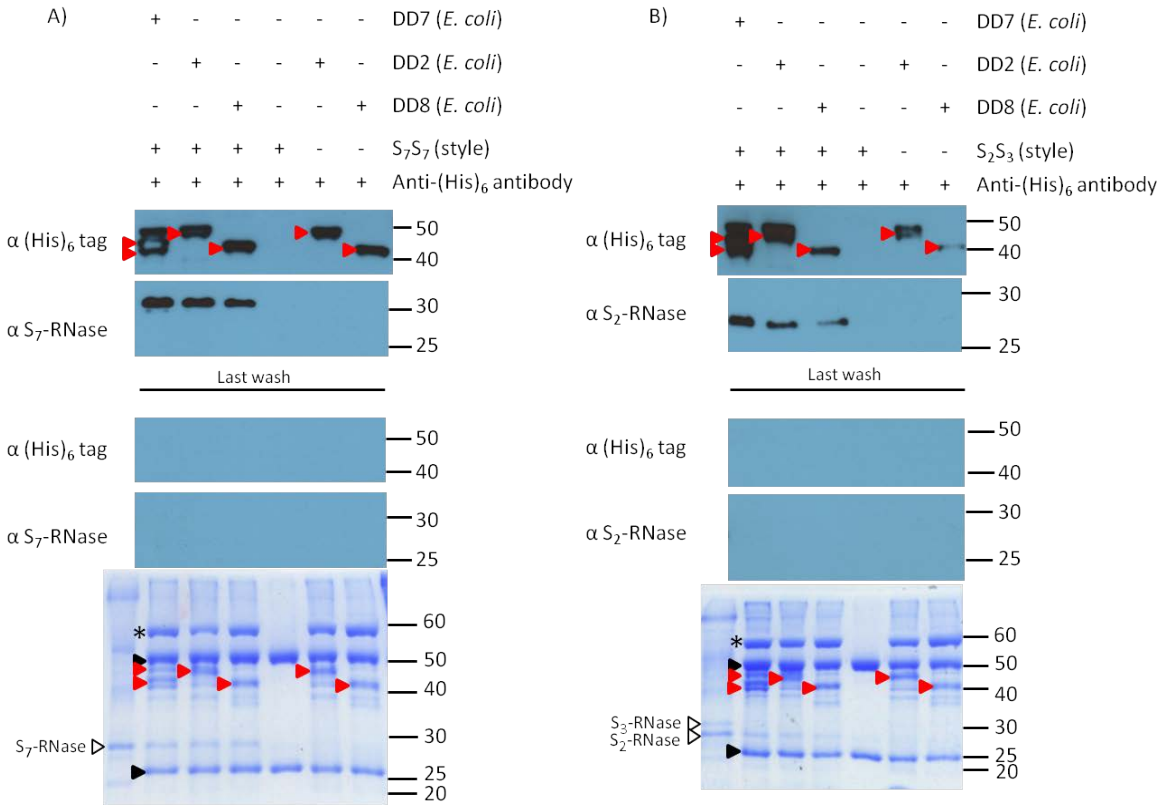


Figure 3.10: Co-immunoprecipitations using anti-$(His)_6$ tag antibody, various $(His)_6$-tagged DDs and native S_7-RNase (A) and S_2-RNase (B).

Extracts from the indicated source were incubated with protein A beads loaded with anti-His tag antibody. Unbound proteins were removed by washing (the last wash was retained for analysis), the beads were boiled in SDS loading buffer and the bound proteins separated by SDS-PAGE. Immunoblots were probed with anti-S_7-RNase antibody (to detect S_7-RNase; A), anti-S_2-RNase antibody (to detect S_2-RNase; B) or an anti-$(His)_6$ antibody (to detect $(His)_6$:DD proteins). Duplicate gels stained with Coomassie are shown below each figure. Selected proteins on the immunoblots and Coomassie-stained gel are indicated: 60 kDa *E. coli* protein (*); antibody heavy chain and light chain (▶); $(His)_6$-tagged DDs (▶); S_2-RNase and S_7-RNase (▷). Input lanes show 1/10th the amount of extract added to the Co-IP assays. Numbers to the side of the figure indicate the sizes (in kDa) of molecular weight standards.

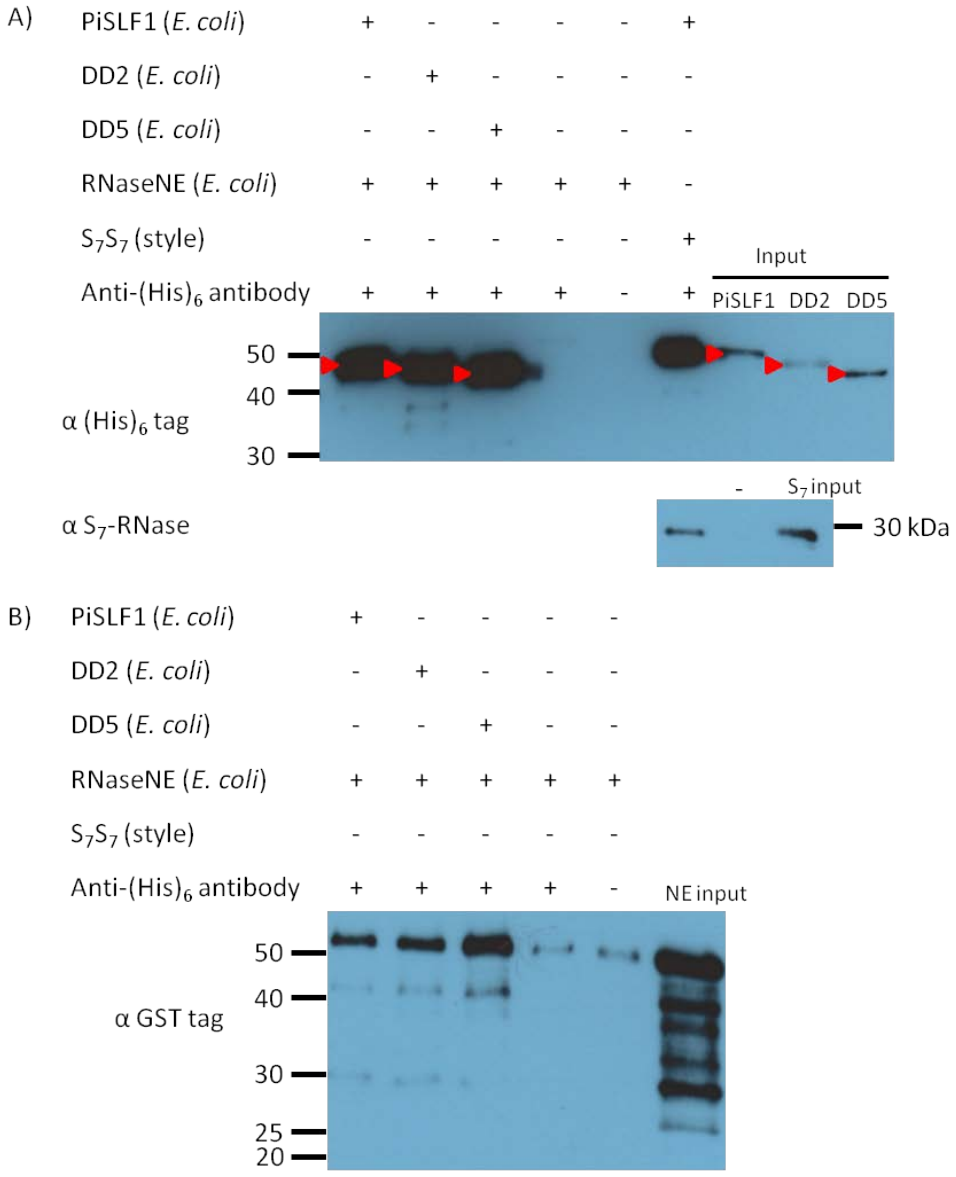


Figure 3.11: Co-immunoprecipitations using anti-$(His)_6$ tag antibody, various $(His)_6$-tagged proteins, S_7-RNase and recombinant GST:RNase NE.

Soluble cell lysates from the indicated source were incubated with protein A beads loaded with anti-$(His)_6$ tag antibody except as shown. Unbound proteins were removed by washing, the beads were boiled in SDS loading buffer and the bound proteins separated on an SDS gel. Immunoblots in A were probed with an anti-$(His)_6$ antibody (to detect $(His)_6$-tagged proteins) or anti-S_7-RNase antibody. Immunoblot in B was probed with an anti-GST antibody to detect GST:RNaseNE. Input lanes show $1/10^{th}$ the amount of lysate added to the Co-IP assays. Numbers to the side of the figure indicate the sizes (in kDa) of molecular weight standards. $(His)_6$-tagged DDs indicated by (▶). – indicates an empty lane.

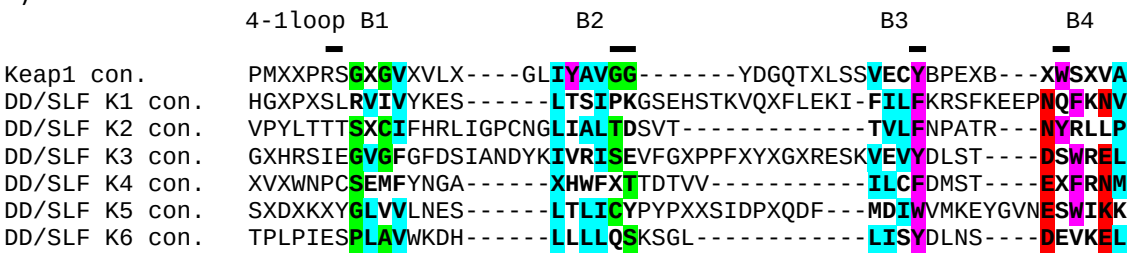
Figure 3.12: Predicted secondary structure of PiSLF1 and the DDs.

A) Alignment of the Kelch repeats in Keap1 and the SLF/DD proteins. First line of sequence shows a consensus of all six Kelch repeats in Keap1. The next six lines of sequence show the consensus of each of the six individual six Kelch-like repeats (K1 to K6) that make up PiSLF1 and DDs 1-10. B1, B2, B3 and B4 refer to the 4 β-strand 'blades' of each Kelch repeat and a loop separating the repeats (the 4-1 loop) is also shown. Black bars mark the conserved amino acids of the Kelch repeat (GG, R, Y and W) and the different coloured boxes indicate amino acids with similar physiochemical properties (cyan = hydrophilic residues; green = small amino acid residues; red = polar residues; purple = hydrophobic residues with an aromatic side chain). Colour coding is from Murphy et al., (2000).

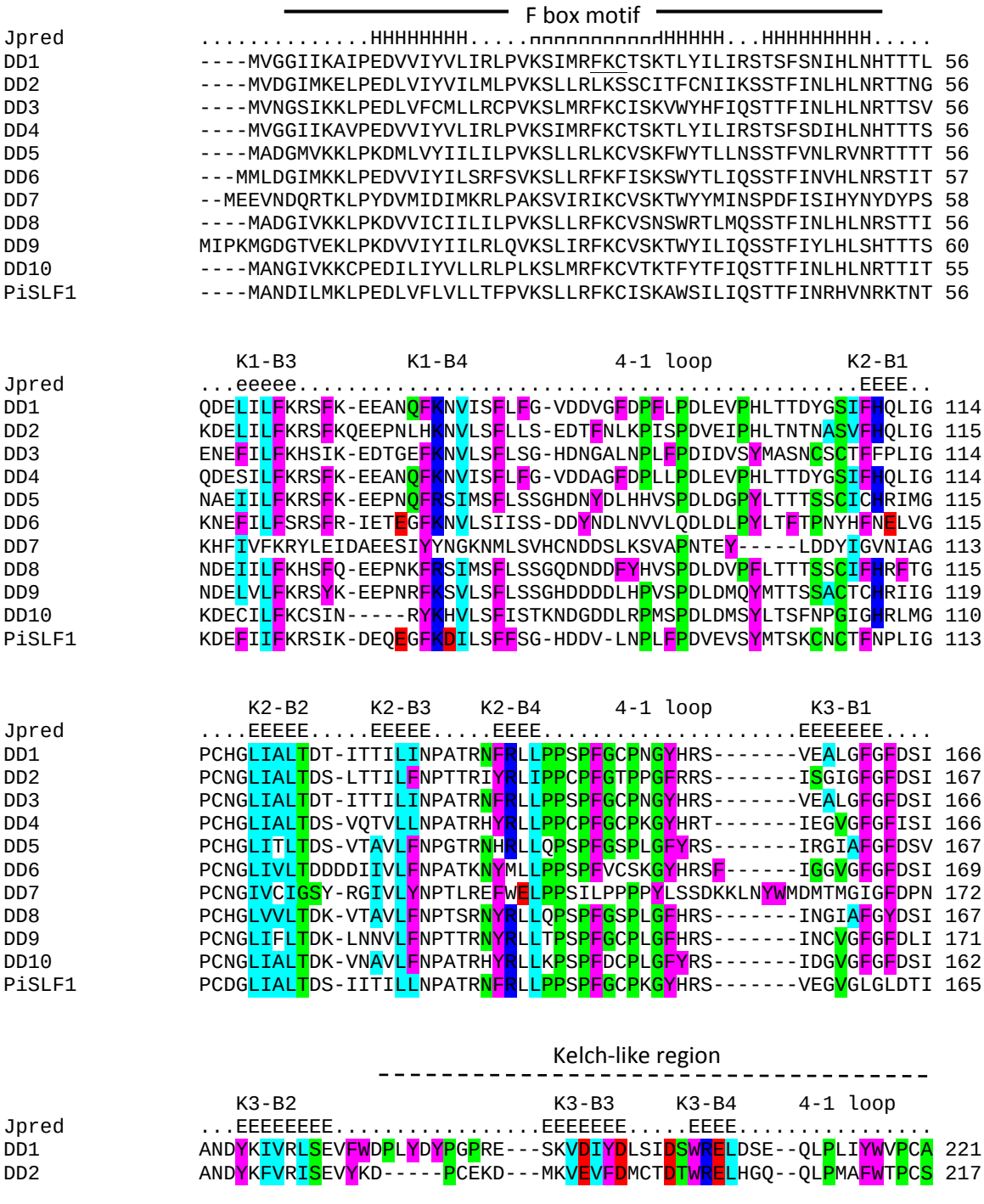
B) Alignment of PiSLF1 and the 10 DDs showing the Jpred prediction of the secondary structural. H and E indicate regions of α-helix and extended β-sheet, respectively, with lower case letters indicating region predicated with lower certainty. A solid line indicates the amino terminal F-box motif and a dotted line the partial Kelch repeat identified by a REP search (Andrade et al., 2000).

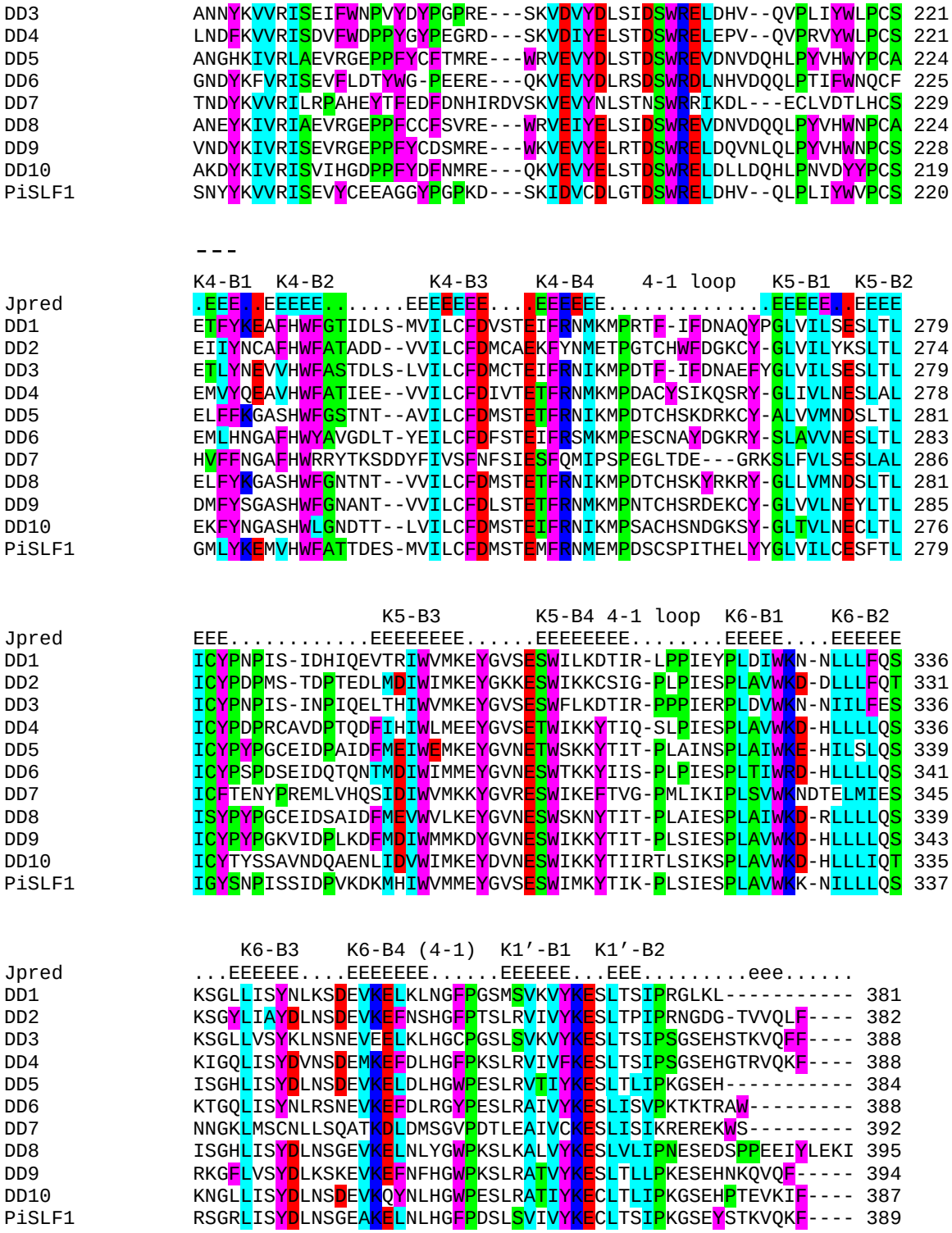
C) Structural models of PiSLF1 (top) and Keap1 (bottom). Modelling used the only β-propeller region of each protein. Each of the six repeats is colour-coded.

A)


4-1loop B1
B2
B3
B4
Keap1 con. PMXXPRSGXGVXVLX----GLIYAVGG-------YDGQTXLSSVECYBPEXB---XWSXVA
DD/SLF K1 con. HGXPXSLRVIVYKES------LTSIPKGSEHSTKVQXFLEKI-FILFKRSFKEEPNQFKNV
DD/SLF K2 con. VPYLTTTSXCIFHRLIGPCNGLIALTDSVT------------TVLFNPATR---NYRLLP
DD/SLF K3 con. GXHRSIEGVGFGFDSIANDYKIVRISEVFGXPPFXYXGXRESKVEVYDLST----DSWREL
DD/SLF K4 con. XVXWNPCSEMFYNGA------XHWFXTTDTVV----------ILCFDMST----EXFRNM
DD/SLF K5 con. SXDXKXYGLVVLNES------LTLICYPYPXXSIDPXQDF---MDIWVMKEYGVNESWIKK
DD/SLF K6 con. TPLPIESPLAVWKDH------LLLLQSKSGL----------LISYDLNS----DEVKEL

B)

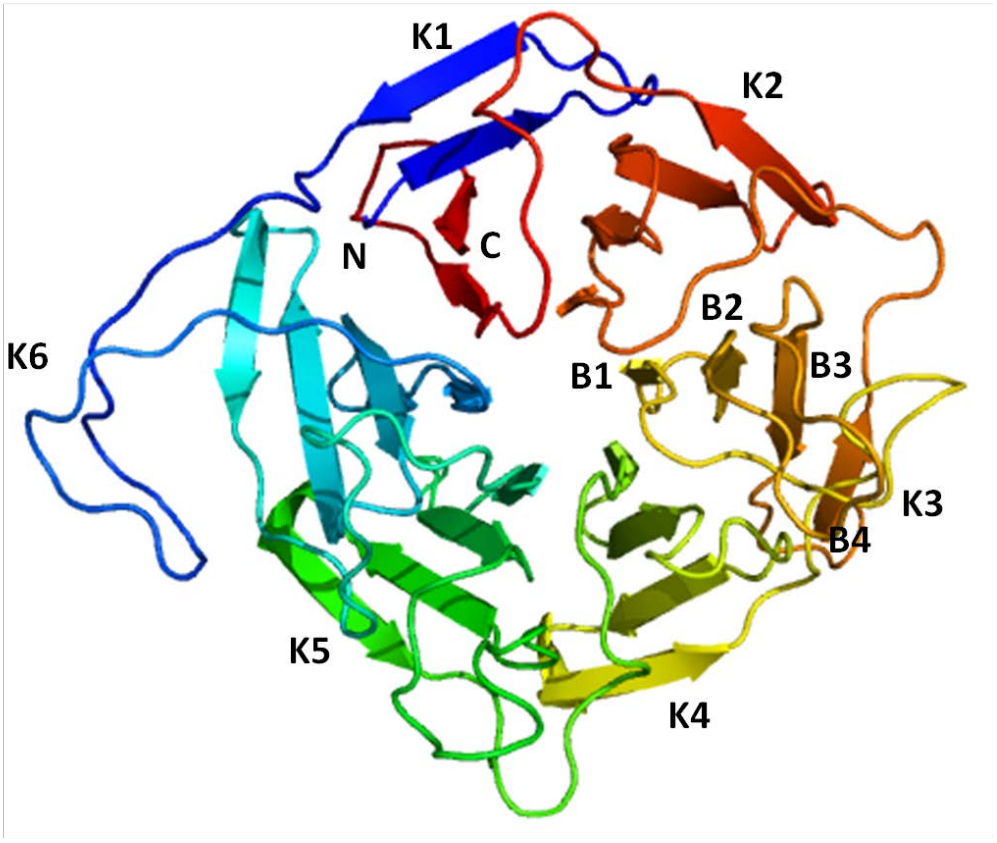

F box motif
Jpred .....HHHHHHHH.....hhhhhhhhhhHHHHH...HHHHHHHHH.....
DD1 ----MVGGIIKAIPEDVVIYVLIRLPVKSIMRFKCTSKTLYILIRSTSFSNIHLNHTTTL 56
DD2 ----MVDGIMKELPEDLVIYVILMLPVKSLLRLKSSCITFCNIIKSSTFINLHLNRTTNG 56
DD3 ----MVNGSIKKLPEDLVFCMLLRCPVKSLMRFKCISKVWYHFIQSTTFINLHLNRTTSV 56
DD4 ----MVGGIIKAVPEDVVIYVLIRLPVKSIMRFKCTSKTLYILIRSTSFSDIHLNHTTTS 56
DD5 ----MADGMVKKLPKDMLVYIILILPVKSLLRLKCVSKFWYTLLNSSTFVNLRVNRTTTT 56
DD6 ---MMLDGIMKKLPEDVVIYILSRFSVKSLLRFKFISKSWYTLIQSSTFINVHLNRSTIT 57
DD7 --MEEVNDQRTKLPYDVMIDIMKRLPAKSVIRIKCVSKTWYYMINSPDFISIHYNYDYPS 58
DD8 ----MADGIVKKLPKDVVICIILILPVKSLLRFKCVSNSWRTLMQSSTFINLHLNRSTTI 56
DD9 MIPKMGDGTVEKLPKDVVIYIILRLQVKSLIRFKCVSKTWYILIQSSTFIYLHLSHTTTS 60
DD10 ----MANGIVKKCPEDILIYVLLRLPLKSLMRFKCVTKTFYTFIQSTTFINLHLNRTTIT 55
PiSLF1 ----MANDILMKLPEDLVFLVLLTFPVKSLLRFKCISKAWSILIQSTTFINRHVNRKTNT 56
K1-B3 K1-B4 4-1 loop K2-B1
Jpred ...eeeee.....EEEE..
DD1 QDELILFKRSFK-EEANQFKNVISFLFG-VDDVGFDPFLPDLEVPHLTTDYGSIFHQLIG 114
DD2 KDELILFKRSFKQEEPNLHKNVLSFLLS-EDTFNLKPISPDVEIPHLTNTNASVFHQLIG 115
DD3 ENEFILFKHSIK-EDTGEFKNVLSFLSG-HDNGALNPLFPDIDVSYMASNCSCTFFPLIG 114
DD4 QDESILFKRSFK-EEANQFKNVISFLFG-VDDAGFDPLLPDLEVPHLTTDYGSIFHQLIG 114
DD5 NAEIILFKRSFK-EEPNQFRSIMSFLSSGHDNYDLHHVSPDLDGPYLTTTSSCICHRIMG 115
DD6 KNEFILFSRSFR-IETEGFKNVLSIISS-DDYNDLNVVLQDLDLPYLTFTPNYHFNELVG 115
DD7 KHFIVFKRYLEIDAEESIYYNGKNMLSVHCNDDSLKSVAPNTEY-----LDDYIGVNIAG 113
DD8 NDEIILFKHSFQ-EEPNKFRSIMSFLSSGQDNDDFYHVSPDLDVPFLTTTSSCIFHRFTG 115
DD9 NDELVLFKRSYK-EEPNRFKSVLSFLSSGHDDDDLHPVSPDLDMQYMTTSSACTCHRIIG 119
DD10 KDECILFKCSIN-----RYKHVLSFISTKNDGDDLRPMSPDLDMSYLTSFNPGIGHRLMG 110
PiSLF1 KDEFIIFKRSIK-DEQEGFKDILSFFSG-HDDV-LNPLFPDVEVSYMTSKCNCTFNPLIG 113
K2-B2 K2-B3 K2-B4 4-1 loop K3-B1
Jpred ....EEEEE.....EEEEE.....EEEE.....EEEEEEE....
DD1 PCHGLIALTDT-ITTILINPATRNFRLLPPSPFGCPNGYHRS-------VEALGFGFDSI 166
DD2 PCNGLIALTDS-LTTILFNPTTRIYRLIPPCPFGTPPGFRRS-------ISGIGFGFDSI 167
DD3 PCNGLIALTDT-ITTILINPATRNFRLLPPSPFGCPNGYHRS-------VEALGFGFDSI 166
DD4 PCHGLIALTDS-VQTVLLNPATRHYRLLPPCPFGCPKGYHRT-------IEGVGFGFISI 166
DD5 PCHGLITLTDS-VTAVLFNPGTRNHRLLQPSPFGSPLGFYRS-------IRGIAFGFDSV 167
DD6 PCNGLIVLTDDDDIIVLFNPATKNYMLLPPSPFVCSKGYHRSF------IGGVGFGFDSI 169
DD7 PCNGIVCIGSY-RGIVLYNPTLREFWELPPSILPPPPYLSSDKKLNYWMDMTMGIGFDPN 172
DD8 PCHGLVVLTDK-VTAVLFNPTSRNYRLLQPSPFGSPLGFHRS-------INGIAFGYDSI 167
DD9 PCNGLIFLTDK-LNNVLFNPTTRNYRLLTPSPFGCPLGFHRS-------INCVGFGFDLI 171
DD10 PCNGLIALTDK-VNAVLFNPATRHYRLLKPSPFDCPLGFYRS-------IDGVGFGFDSI 162
PiSLF1 PCDGLIALTDS-IITILLNPATRNFRLLPPSPFGCPKGYHRS-------VEGVGLGLDTI 165
Kelch-like region
K3-B2 K3-B3 K3-B4 4-1 loop
Jpred ...EEEEEEEE.....EEEEEEE.....EEEE.....
DD1 ANDYKIVRLSEVFWDPLYDYPGPRE---SKVDIYDLSIDSWRELDSE--QLPLIYWVPCA 221
DD2 ANDYKFVRISEVYKD-----PCEKD---MKVEVFDMCTDTWRELHGQ--QLPMAFWTPCS 217


DD3 ANNYKVVRISEIFWNPVYDYPGPRE---SKVDVYDLSIDSWRELDHV--QVPLIYWLPCS 221
DD4 LNDFKVVRISDVFWDPPYGYPEGRD---SKVDIYELSTDSWRELEPV--QVPRVYWLPCS 221
DD5 ANGHKIVRLAEVRGEPPFYCFTMRE---WRVEVYDLSTDSWREVDNVDQHLPYVHWYPCA 224
DD6 GNDYKFVRISEVFLDTYWG-PEERE---QKVEVYDLRSDSWRDLNHVDQQLPTIFWNQCF 225
DD7 TNDYKVVRILRPAHEYTFEDFDNHIRDVSKVEVYNLSTNSWRRIKDL---ECLVDTLHCS 229
DD8 ANEYKIVRIAEVRGEPPFCCFSVRE---WRVEIYELSIDSWREVDNVDQQLPYVHWNPCA 224
DD9 VNDYKIVRISEVRGEPPFYCDSMRE---WKVEVYELRTDSWRELDQVNLQLPYVHWNPCS 228
DD10 AKDYKIVRISVIHGDPPFYDFNMRE---QKVEVYELSTDSWRELDLLDQHLPNVDYYPCS 219
PiSLF1 SNYYKVVRISEVYCEEAGGYPGPKD---SKIDVCDLGTDSWRELDHV--QLPLIYWVPCS 220

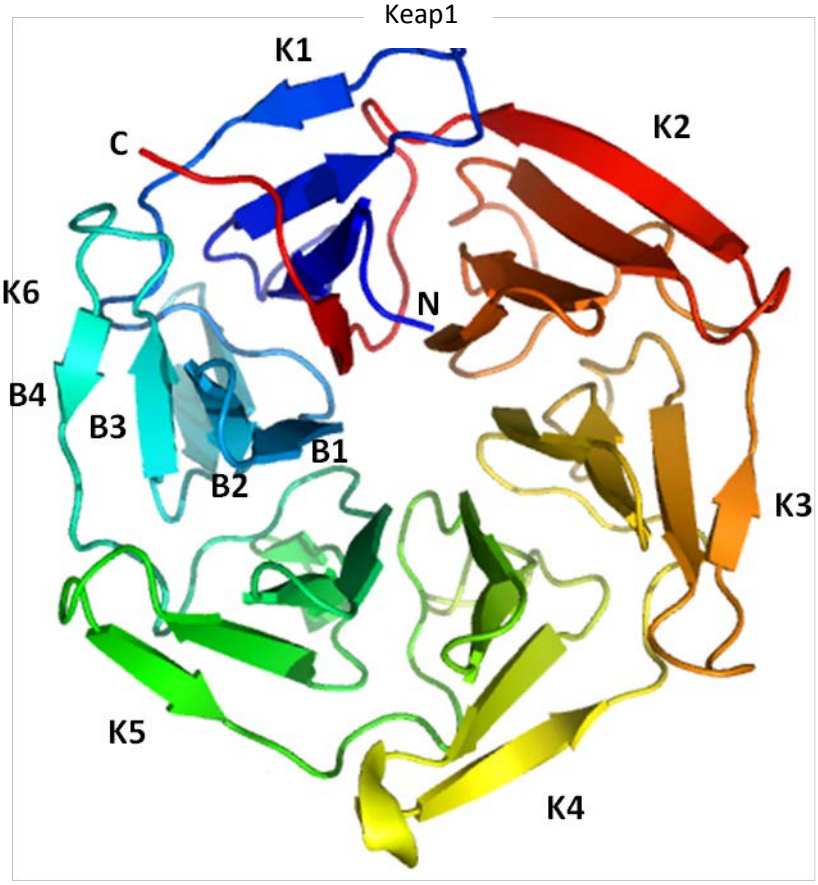
K4-B1 K4-B2 K4-B3 K4-B4 4-1 loop K5-B1 K5-B2
Jpred .EEE..EEEEE.....EEEEEEEE...EEEEEE.....EEEEE..EEEE
DD1 ETFYKEAFHWFGTIDLS-MVILCFDVSTEIFRNMKMPRTF-IFDNAQYPGLVILSESLTL 279
DD2 EIIYNCAFHWFATADD--VVILCFDMCAEKFYNMETPGTCHWFDGKCY-GLVILYKSLTL 274
DD3 ETLYNEVVHWFASTDLS-LVILCFDMCTEIFRNIKMPDTF-IFDNAEFYGLVILSESLTL 279
DD4 EMVYQEAVHWFATIEE--VVILCFDIVTETFRNMKMPDACYSIKQSRY-GLIVLNESLAL 278
DD5 ELFFKGASHWFGSTNT--AVILCFDMSTETFRNIKMPDTCHSKDRKCY-ALVVMNDSLTL 281
DD6 EMLHNGAFHWYAVGDLT-YEILCFDFSTEIFRSMKMPESCNAYDGKRY-SLAVVNESLTL 283
DD7 HVFFNGAFHWRRYTKSDDYFIVSFNFSIESFQMIPSPEGLTDE---GRKSLFVLSESLAL 286
DD8 ELFYKGASHWFGNTNT--VVILCFDMSTETFRNIKMPDTCHSKYRKRY-GLLVMNDSLTL 281
DD9 DMFYSGASHWFGNANT--VVILCFDLSTETFRNMKMPNTCHSRDEKCY-GLVVLNEYLTL 285
DD10 EKFYNGASHWLGNDTT--LVILCFDMSTEIFRNIKMPSACHSNDGKSY-GLTVLNECLTL 276
PiSLF1 GMLYKEMVHWFATTDES-MVILCFDMSTEMFRNMEMPDSCSPITHELYYGLVILCESFTL 279
K5-B3 K5-B4 4-1 loop K6-B1 K6-B2
Jpred EEE.....EEEEEEEE......EEEEEEEE........EEEEE....EEEEEE
DD1 ICYPNPIS-IDHIQEVTRIWVMKEYGVSESWILKDTIR-LPPIEYPLDIWKN-NLLLFQS 336
DD2 ICYPDPMS-TDPTEDLMDIWIMKEYGKKESWIKKCSIG-PLPIESPLAVWKD-DLLLFQT 331
DD3 ICYPNPIS-INPIQELTHIWVMKEYGVSESWFLKDTIR-PPPIERPLDVWKN-NIILFES 336
DD4 ICYPDPRCAVDPTQDFIHIWLMEEYGVSETWIKKYTIQ-SLPIESPLAVWKD-HLLLLQS 336
DD5 ICYPYPGCEIDPAIDFMEIWEMKEYGVNETWSKKYTIT-PLAINSPLAIWKE-HILSLQS 339
DD6 ICYPSPDSEIDQTQNTMDIWIMMEYGVNESWTKKYIIS-PLPIESPLTIWRD-HLLLLQS 341
DD7 ICFTENYPREMLVHQSIDIWVMKKYGVRESWIKEFTVG-PMLIKIPLSVWKNDTELMIES 345
DD8 ISYPYPGCEIDSAIDFMEVWVLKEYGVNESWSKNYTIT-PLAIESPLAIWKD-RLLLLQS 339
DD9 ICYPYPGKVIDPLKDFMDIWMMKDYGVNESWIKKYTIT-PLSIESPLAVWKD-HLLLLQS 343
DD10 ICYTYSSAVNDQAENLIDVWIMKEYDVNESWIKKYTIIRTLSIKSPLAVWKD-HLLLIQT 335
PiSLF1 IGYSNPISSIDPVKDKMHIWVMMEYGVSESWIMKYTIK-PLSIESPLAVWKK-NILLLQS 337
K6-B3 K6-B4 (4-1) K1′-B1 K1′-B2
Jpred ...EEEEEE....EEEEEEE......EEEEEE...EEE.........eee......
DD1 KSGLLISYNLKSDEVKELKLNGFPGSMSVKVYKESLTSIPRGLKL----------- 381
DD2 KSGYLIAYDLNSDEVKEFNSHGFPTSLRVIVYKESLTPIPRNGDG-TVVQLF---- 382
DD3 KSGLLVSYKLNSNEVEELKLHGCPGSLSVKVYKESLTSIPSGSEHSTKVQFF---- 388
DD4 KIGQLISYDVNSDEMKEFDLHGFPKSLRVIVFKESLTSIPSGSEHGTRVQKF---- 388
DD5 ISGHLISYDLNSDEVKELDLHGWPESLRVTIYKESLTLIPKGSEH----------- 384
DD6 KTGQLISYNLRSNEVKEFDLRGYPESLRAIVYKESLISVPKTKTRAW---------- 388
DD7 NNGKLMSCNLLSQATKDLDMSGVPDTLEAIVCKESLISIKREREKWS---------- 392
DD8 ISGHLISYDLNSGEVKELNLYGWPKSLKALVYKESLVLIPNESEDSPPEEIYLEKI 395
DD9 RKGFLVSYDLKSKEVKEFNFHGWPKSLRATVYKESLTLLPKESEHNKQVQF----- 394
DD10 KNGLLISYDLNSDEVKQYNLHGWPESLRATIYKECLTLIPKGSEHPTEVKIF---- 387
PiSLF1 RSGRLISYDLNSGEAKELNLHGFPDSLSVIVYKECLTSIPKGSEYSTKVQKF---- 389

C)

PiSLF1



Keap1



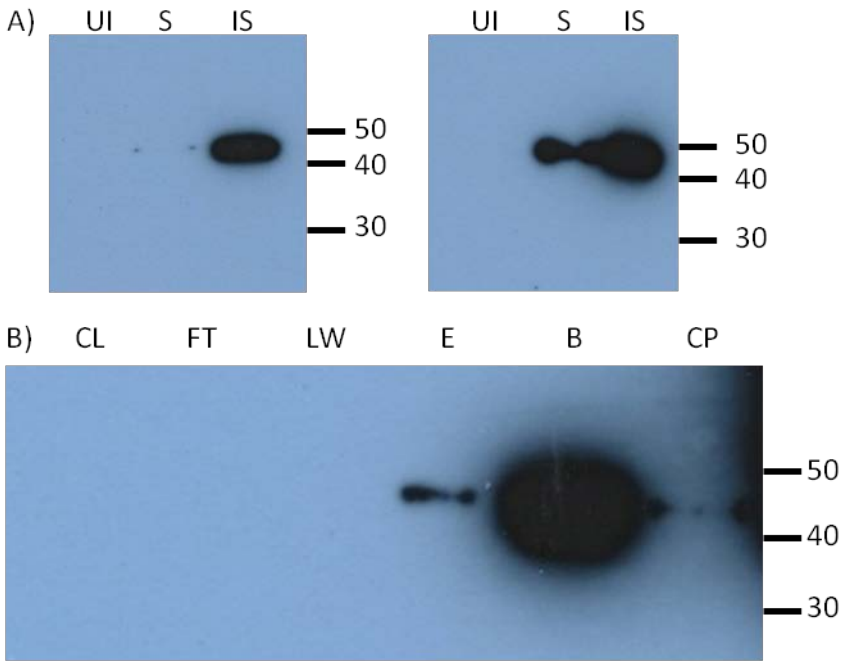


Figure 3.13: Expression of $(His)_6$:PiSLF1 FL.

A) Expression of $(His)_6$:PiSLF1 FL in *E. coli* BL21 codonplus RIL (left panel) or BL21 (right panel) cells was induced by IPTG addition and cells were subsequently grown for 16 hr at 16°C. UI indicates a total extract made from uninduced cells. S and IS indicate the soluble and insoluble fractions, respectively, prepared from cells 16 hr after induction. The same amount of protein was loaded in each lane. Fractions were separated on polyacrylamide gels and protein detection done by immunoblotting using an anti-$(His)_6$ tag antibody. Numbers to the right of the figures indicate sizes (in kDa) of molecular weight markers.

B) Soluble $(His)_6$:PiSLF1 FL was purified batch-wise using NTA affinity beads and the proteins present in the various fractions were detected by immunoblot analysis using an anti-$(His)_6$ tag antibody. CL: total soluble cell lysate; FT: flow-through; LW: last wash of beads before elution; E: elution fraction (250 mM imidazole); B: protein bound to the beads after elution; CP: Concentrated eluate fraction; - empty lane. Numbers to the right of the figures indicate sizes (in kDa) of molecular weight markers.

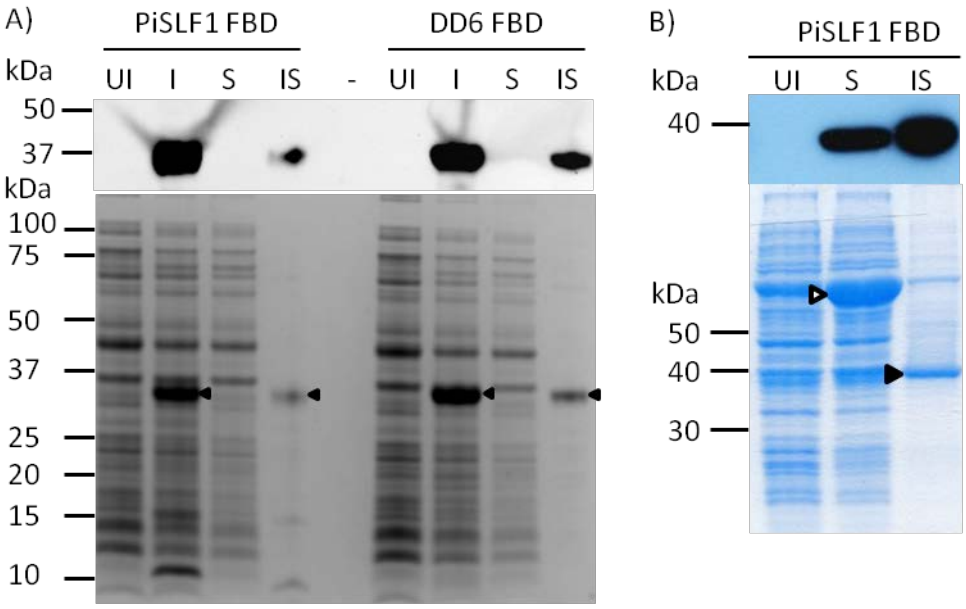


Figure 3.14: Expression of $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD in *E. coli*.

A) *E. coli* BL21 star cells expressing $(His)_6$:PiSLF1 FBD or $(His)_6$:DD6 FBD were grown at 37°C after induction and harvested 3 hr post-induction. UI indicates a total extract from uninduced cells and I indicates a total extract made from induced cells. S and IS indicate the soluble and insoluble fractions from induced cells, respectively. Fractions were prepared and separated by SDS-PAGE and immunoblot analysis (upper panel) was performed using an anti-$(His)_6$ antibody. Arrowheads indicate the position of $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD in a duplicate Coomassie stained gel (lower panel) and numbers to the left are the sizes (in kDa) of molecular weight markers. – indicates an empty lane. All lanes have equal amounts of protein except the IS lane, which has one-fifth the amount of protein present in other lanes.

B) ArticExpress cells expressing $(His)_6$:PiSLF1 FBD were grown at 13°C after induction and harvested 24 hr post-induction. Proteins were analysed as described in A. The open arrowhead indicates the cold-adapted chaperonin Cpn60 from *Oleispira antarctica* and the closed arrowhead indicates $(His)_6$:PiSLF1 FBD.

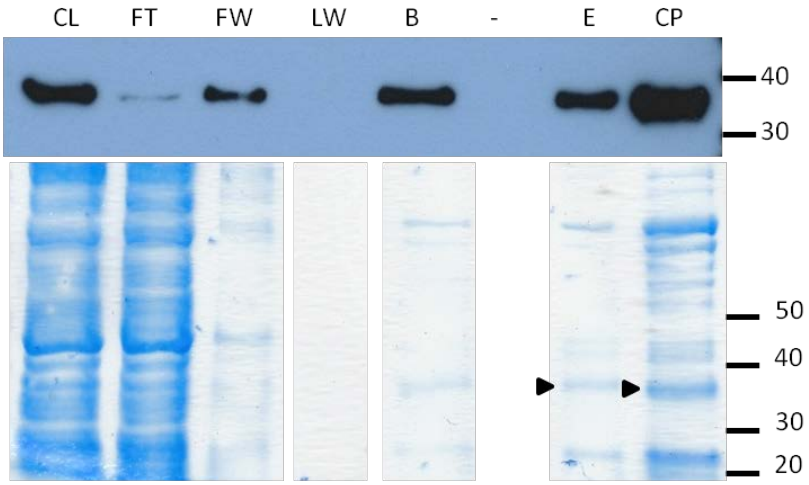


Figure 3.15: Expression and affinity purification of $(His)_6$:PiSLF1 FBD.

Soluble $(His)_6$:PiSLF1 FBD was batch-wise purified using Ni-NTA affinity beads as described in the legend to Figure 3.4B. Protein detection was by immunoblot analysis using an anti-$(His)_6$ tag antibody (upper panel) and Coomassie staining (lower panel). CL: total soluble cell lysate; FT: Flow-through fraction; FW: First wash; LW: last wash of beads before elution; E: elution fraction; B: protein bound to the beads after elution; CP: concentrated eluate fraction. Numbers to the right indicate sizes (in kDa) of molecular weight markers.

4.1: Introduction

Although recombinant protein expression in *E. coli* has many advantages over other expression systems, it often results in the production of inactive protein aggregates known as inclusion bodies or IBs (Baneyx and Mujacic, 2004). Two factors known to contribute to IB formation in *E. coli* are level of expression (exceeding the protein-folding capacity of the cell) and a requirement for post-translational modifications such as disulphide bonding or glycosylation (Lilie et al., 1998). While IB formation means the target protein is inactive and insoluble, it also means the protein exists in highly pure aggregates, as up to 90% of the content of an IB is the protein of interest. This can facilitate purification as IBs are readily isolated in a simple centrifugation step. More importantly, because the IB protein is protected from proteolysis, it is generally also intact (Ventura and Villaverde, 2006). Thus, if a simple and efficient renaturation procedure can be developed, deposition in IBs and subsequent isolation and renaturation are often the most straightforward way of producing large amounts of native protein in a mainly pure form (Lilie et al., 1998).

Generally, protein refolding is done in two main steps: In the first step, a strong denaturant such as urea or guanidine hydrochloride is used to solubilise the IB protein; and in the second step the denaturant is removed and conditions favourable to protein refolding are selected. A refolding protein may pass through one or more intermediate states before it is completely folded. At high protein concentrations these partially folded intermediates can interact with each other to form aggregates of denatured protein. To push the equilibrium towards production of soluble native protein rather than insoluble aggregates requires testing a range of additives such as metal ions, glycerol, detergents and amino acids, and refolding conditions such as temperature, buffer compatibility and pH (Burgess, 2009). Many methods have been developed for protein renaturation and these are documented in the REFOLD database (Chow et al., 2006a; Chow et al., 2006b).

As presented in Chapter 3, it was not possible to produce $(His)_6$:PiSLF1 FBD in a soluble and largely pure form, so work instead turned to refolding the insoluble form using a protocol developed based on general guidelines in the REFOLD database and other sources (Chow et al., 2006a; Burgess, 2009). This chapter reports *in vitro* refolding of $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD and preliminary biophysical characterization of the refolded proteins. Binding experiments using the refolded proteins, stylar extracts and purified S-RNases are also described.

4.2: Materials and Methods

4.2.1: Protein folding and purification

Recombinant $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD (clone from DD6 S_2-allele, accession no. EF420256.1) induction in *E. coli* BL21 star was done as described in section 3.2.1. Inclusion bodies were prepared and washed with Bugbuster master mix as described in the manufacturer's protocol for IB purification (Novagen). Detergent was removed from the IB pellet by washing three times in 25 mL of a buffer containing 50 mM bicine, 100 mM NaCl, 5% (v/v) glycerol and 1 mM TCEP ((2-carboxyethyl) phosphine, Sigma Aldrich) and the pellet was solubilised in 8 M guanidine hydrochloride (Sigma Aldrich) in 50 mM Bicine buffer (pH 6) with shaking at 37°C for between 2 hr and overnight. The solubilised protein was filtered through 0.22 µM filter (Millipore) and the protein concentration was determined as described in section 3.2.4 and adjusted to 1 mg/mL with 8 M guanidine hydrochloride in 50 mM Bicine buffer (pH 6).

An initial screen of refolding conditions suitable for $(His)_6$:PiSLF1 FBD was done in 24-well plates (Corning) as shown in Figure 4.1. Solubilised protein (10 µl) was added to 1 mL water, buffer or buffer containing an additive (a metal ion, detergent, glycerol or polyethylene glycol) as listed in Figure 4.1. The plate was gently shaken on an orbital mixer (Ratek) at 100 rpm at room temperature for 2-3 hr and each well was checked for precipitation by eye and photographed with a digital camera.

The final refolding protocol for $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD was as follows: solubilised $(His)_6$:PiSLF1 FBD or $(His)_6$:DD6 FBD in 8 M guanidine hydrochloride was dripped into folding solution (water with 50 µM ferric chloride ($FeCl_3$; Sigma Aldrich) and 0.5 mM TCEP). In a typical experiment, 20 mL of solubilised protein was dripped into 1 L of refolding solution (1:50 dilution) at 0.4 mL/min using peristaltic pump (Econo pump model EP1, Bio-Rad). After all the solubilized protein had been added, the diluted protein solution was allowed to stand at room temperature for 2 hr before being passed through 0.22 µM filter (Millipore) to remove any precipitate. At the same time, 5 mL (bed volume) of nitroloacetic acid (NTA) beads (Qiagen) was packed by gravity into a 2.5 × 10 cm disposable chromatography column (Bio-Rad). The column was washed with 50 mL water, loaded with the dilute protein solution and washed (2 column volumes per wash) with increasing amounts of imidazole (20, 50 and 100 mM) in 50 µM $FeCl_3$ and 500 µM TCEP. The column was washed five times (1 column volume per wash) with elution solution (200 mM imidazole, 50 µM $FeCl_3$ and 500 µM TCEP), the eluate fractions were combined and concentrated using an Amicon ultrafiltration unit with a 10,000 Dalton cut-off (Millipore**)** to a final volume of about 20 mL. The concentrated eluate was dialysed at room temperature overnight against 1 L of 50 mM Bicine (pH 7.6) containing 50 µM

$FeCl_3$ and 500 µM TCEP (dialysis buffer) using a dialysis cassette with a 10,000 Dalton cut-off (Thermo Scientific). A small volume of dialysis buffer was retained for use as a buffer blank in subsequent analyses. The protein solution was removed from the dialysis cassette, concentrated using an Amicon unit to about 1 mg/mL and stored at 4^{o}C for up to a week prior to use. Routinely, about 1 g of pelleted bacterial cells yielded about 20 mg of solubilized IB protein and about 1 mg of refolded and affinity purified protein (either $(His)_6$:PiSLF1 FBD or $(His)_6$:DD6 FBD).

4.2.2: Mass spectrometry analysis of $(His)_6$:PiSLF1 FBD

The molecular mass of intact and refolded $(His)_6$:PiSLF1 FBD was determined by electrospray ionization, time-of-flight mass spectrometry (ESI-TOF MS). The sample was injected onto a GE Healthcare µRPC C2/C18 ST 4.6/100 column (4.6 × 100 mm, 3 µm 120 Å) using an Agilent 1200 series HPLC and eluted directly into an Agilent 6220 ESI-TOF with a gradient of 95% solution A (0.1% formic acid) to 100% solution B (0.1% formic acid; 90% acetonitrile) at a flow rate of 400 µL/min. Mass spectrometry was performed with the assistance of Alexander Ray from the Bio21 Institute (University of Melbourne).

4.2.3: Circular dichroism spectroscopy

Circular dichroism (CD) spectra of refolded protein samples were recorded using an AVIV Model 410-SF CD spectrometer. Wavelength scans were performed between 190 and 260 nm with a sample concentration of 0.15 mg/mL in dialysis buffer in 10 mm x 0.1 mm quartz cuvettes at 25°C unless indicated otherwise. Three scans were performed for every sample and averaged readings were used for analysis. For thermal denaturation, protein samples were placed in a cuvette and scanned before and after heating to 90°C for 5 min. Samples were cooled to 25°C and rescanned. For chemical denaturation, protein samples were mixed with 8 M guanidine hydrochloride (dissolved in dialysis buffer) to a final protein concentration approximately 0.15 mg/mL. Data were analyzed using the CDPro software package (Sreerama et al., 2000).

4.2.4: Gel filtration chromatography

Gel filtration liquid chromatography was performed using S200 Sephacryl high resolution resin (GE Healthcare) in a 50 × 0.7 cm Bio-Rad glass column with a column volume of approximately 20 mL. Refolded protein (approximately 1 mg) was manually loaded onto the column and the column was developed using 2.5 column volumes of dialysis buffer plus 150 mM NaCI at a flow rate of 0.13 ml/min. Fractions of 1 mL were collected and the protein concentration of each fraction was determined using the Bradford assay (section 3.2.4). Protein standards used to calibrate the column

were human transferrin (76 kDa), chicken albumin (45 kDa) and horse myoglobin (17 kDa; all from Sigma). The column void volume was determined using dextran blue 2000 (Sigma).

4.2.5: Analytical ultracentrifugation

Sedimentation velocity experiments were conducted in a Beckman model XL-I analytical ultracentrifuge (Beckman Coulter, Fullerton, CA) at a temperature of 20°C. Samples were loaded into a conventional double sector quartz cell and mounted in a Beckman 4-hole An-60 Ti rotor at an initial concentration of 0.15 mg/mL in dialysis buffer. Radial absorbance data were acquired using a rotor speed of 15,000 rpm and a wavelength of 280 nm, with radial increments of 0.003 cm in continuous scanning mode. Sedimentation velocity data at multiple time points were fitted to a continuous sedimentation coefficient [c(s)] distribution and a continuous mass [c(M)] distribution model using the program SEDFIT (Schuck and Rossmanith, 2000), which is available at www.analyticalultracentrifugation.com. Data were fitted using a regularization parameter of p = 0.95, floating frictional ratios, and 150 sedimentation coefficient increments in the range of 0.1-300 S. Analytical ultracentrifugation was performed with the assistance of Dr Mok Yee Foong from the Bio21 Institute (University of Melbourne).

4.2.6: Co-immunoprecipitation

Co-immunoprecipitation (Co-IP) was performed as described in section 3.2.5 with the following changes; 1) Binding buffer used contain 50 mM Bicine, 100 mM NaCl, 5 mM $MgCl_2$, 50 µM ferric chloride, 0.5 mM TCEP and 0.01% nonidet P40; 2) 10 µg of enriched S-RNase(s) or total style extract and 10 µg of refolded $(His)_6$:PiSLF1 FBD/$(His)_6$:DD6 FBD was used unless otherwise stated. Purified NaPI protein used for Co-IP and anti-NaPI antibody used for immunoblotting was provided by Dr Simon Poon from Hexima Ltd (La Trobe University, Australia). Bound proteins were eluted by heating at 95°C for 5 min in 2× protein loading buffer, separated on 12% polyacrylamide gel and immunoblotted as described in section 3.2.2. Immunoblot images were obtained either by exposing to film or digitally scanned by ChemiDoc imager (Biorad).

4.2.7: S-RNase purification

S-RNases were purified from stylar extracts essentially as described in Jahnen et al., (1989) with some modifications. Styles frozen in liquid nitrogen were ground to a powder with a mortar and pestle and extracted with 10 ml per gram frozen tissue of 100 mM Tris (pH 8.0) and 14 mM β-mercaptoethanol and 0.1 gram insoluble Polyclar AT per gram of frozen tissue. The mixture was stirred for 10 min on ice, filtered through two layers of Miracloth (Calbiochem) and centrifuged (8,000g for 20 min, 4°C). The supernatant was adjusted to 50% fractional saturation by adding

saturated ammonium sulfate and stirred on ice for 30 min. The solution was centrifuged as above and the $(NH_4)_2SO_4$ concentration of the supernatant increased to 95% fractional saturation by addition of 0.316 g solid $(NH_4)_2SO_4$ per ml of supernatant. The solution was stirred slowly at 4°C for 20 min and the precipitate was collected by centrifugation (10,000g for 20 min). The pellet was dissolved in 50 mM sodium acetate ($NaCH_3CO_2$) buffer (pH 5.0) and desalted on a PD 10 column (GE healthcare) equilibrated with the same buffer. The protein solution was loaded onto a cation exchange column (Econo-Pac S cartridge, Bio-Rad) previously equilibrated in $NaCH_3CO_2$ buffer (pH 5.0). Bound protein was eluted with a salt gradient from 0 to 0.5 M NaCl in $NaCH_3CO_2$ buffer (pH 5.0) delivered by a Bio-Rad Econo liquid chromatography system. Fractions of 1 mL were collected and the protein concentration of each fraction was determined using the Bradford assay (section 3.2.4). Fractions were analysed by SDS-PAGE followed by coomassie staining and fractions containing S-RNase were kept in -80^{o}C until further use.

4.3: Results

4.3.1: Folding of insoluble $(His)_6$:PiSLF1 FBD

Chapter 3 illustrated the difficulties encountered in producing sufficient quantities of a soluble and enriched form of $(His)_6$:PiSLF1 suitable for biochemical and other studies. As the insoluble form of $(His)_6$:PiSLF1 was the most abundant protein upon expression in *E. coli*, an alternative was to isolate the insoluble form and then attempt to fold it into a soluble form. $(His)_6$:PiSLF1 FBD was chosen for this, as there had been some success at producing it in a soluble form (see section 3.3.4).

A screen was performed of small molecule additives and buffer conditions that could aid in folding of $(His)_6$:PiSLF1 FBD from the insoluble fraction. Included in the screen were various detergents and metal salts, and additives like glycerol and polyethylene glycol (PEG) that are suggested to stabilise folded proteins (Burgess, 2009). Briefly, the screen involved rapidly diluting 10 µL of a 1 mg/mL solution of resolubilised $(His)_6$:PiSLF1 FBD in 8 M guanidine HCl into 1 mL of a test solution (Figure 4.1). Folding was allowed to proceed for 2-3 h, at which point the presence of a visible precipitate was assessed, with the absence of a precipitate indicating that the protein was still in a soluble, and thus potentially folded, form. Figure 4.1A shows the protein precipitates that formed under unfavourable conditions (red arrows).

Figure 4.1B shows that $(His)_6$:PiSLF1 FBD remained soluble in water, either with or without added metal ions, but usually precipitated when a buffering agent was present with the only exception being 50 mM Bicine (pH 7.6) with 50 µM ferric chloride ($FeCl_3$). Similarly, protein aggregation was seen when $(His)_6$:PiSLF1 FBD was folded into buffer solutions that contained a detergent or an

additive like glycerol or PEG (Figure 4.1C). The final conditions chosen for folding $(His)_6$:PiSLF1 FBD were by drip-wise dilution of a 1 mg/ml solution into unbuffered $FeCl_3$. Insoluble $(His)_6$:DD6 FBD was also successfully refolded using this method.

Figure 4.2A shows the recovery of $(His)_6$:PiSLF1 FBD from folding solution by Ni-affinity chromatography. The anti-$(His)_6$ antibody and Coomassie staining detected a major protein of about 39 kDa, which is close to the expected size for $(His)_6$:PiSLF1 FBD. Most of the folded $(His)_6$:PiSLF1 FBD bound to the column with protein in the flow through (FT) fraction possibly because the column's capacity had been exceeded. The protein was eluted with 200 mM imidazole and the Coomassie-stained gel shows the 39 kDa $(His)_6$:PiSLF1 FBD protein was the most abundant protein in this fraction. ESI-TOF analysis indicated the major eluted protein had a molecular weight of 39,526 Da, which matches the predicted mass of 39,526 Da for $(His)_6$:PiSLF1 FBD. The pooled eluate fractions were dialysed against bicine buffer with $FeCl_3$ and concentrated to about 1 mg/mL. Under these conditions the protein was stable for about a week at 4°C. Routinely, 20 mg of solubilised protein (in 8 M guanidine hydrochloride) yielded about 1 mg of affinity purified $(His)_6$:PiSLF1 FBD.

Figure 4.2B shows purification of refolded $(His)_6$:DD6 FBD. Like $(His)_6$:PiSLF1 FBD, a single major protein was eluted with 200 mM imidazole and the size of this protein was consistent with the expected size of $(His)_6$:DD6 FBD (39.6 kDa). Similarly, about 1 mg of $(His)_6$:DD6 FBD was routinely obtained after affinity purification.

4.3.2: Biophysical characterisation of the refolded proteins

To estimate the molecular weight of the native refolded protein, size exclusion chromatography (SEC) was performed on $(His)_6$:DD6 FBD (Figure 4.3). Monomeric $(His)_6$:DD6 FBD (S_2 allele) should elute near the 42 kDa standard (chicken albumin). However, the apparent molecular weight of $(His)_6$:DD6 FBD was above 80 kDa, as it eluted earlier than the 76 kDa standard (transferrin), close to the void volume. Thus, it appeared that the refolded $(His)_6$:DD6 FBD did not exist as a monomer in its refolded state. SEC with $(His)_6$:PiSLF1 FBD produced the same result (data not shown).

Analytical ultracentrifugation (AUC) was performed to determine the size-distribution of refolded $(His)_6$:PiSLF1 FBD. Figure 4.4 shows the protein sedimented as a broad asymmetrical peak centered around 18S, suggesting that $(His)_6$:PiSLF1 FBD exists as a range of molecular sizes. Based on the molecular weight of $(His)_6$:PiSLF1 FBD and assuming a globular structure, monomers of this protein should have a sedimentation coefficient of between 3S and 4S (Schuck et al., 2002). Hence, this result indicated that refolded $(His)_6$:PiSLF1 FBD existed as oligomers composed of variable numbers of monomers.

Circular dichroism (CD) analysis in the far ultraviolet region (190-260 nm) was used to investigate whether folded $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD were mostly composed of β-sheet, as predicted if kelch repeats are the major structural element. Figure 4.5A shows that the CD spectrum for folded $(His)_6$:PiSLF1 FBD at room temperature (25°C) had positive ellipticity at 196 nm, a crossover of the baseline at about 202 nm and a minimum ellipticity value at around 218 nm, all features that are consistent with a protein composed mostly of β-sheet (Kelly et al., 2005). Although absorbance below 200 nm could be due to the presence of chloride ions in the buffer, their concentration was too low to interfere with the CD spectrum. Ellipticity at wavelengths >245 nm was also close to zero, indicating interpretation was not affected by light scattering from protein precipitates. The percentage of β-sheet in the protein was not determined as the presence of $FeCl_3$ in the buffer interfered with accurate determination of the protein concentration.

Heating $(His)_6$:PiSLF1 FBD to 90°C for 5 min did not markedly alter the CD spectrum, indicating the protein was resistant to heat denaturation (Figure 4.5A). However, $(His)_6$:PiSLF1 FBD was disordered when incubated in 8 M guanidine hydrochloride with little or no spectral reading from 210 to 260 nm (Figure 4.5B). $(His)_6$:DD6 FBD had a similar CD spectrum to $(His)_6$:PiSLF1 FBD and could be chemically denatured but was thermally stable (Figures 4.6A and 4.6B).

4.3.3: Binding of folded $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD to native *N. alata* S-RNases

The Co-IP assay described in Chapter 3 was used to determine whether the folded proteins interacted with *N. alata* S-RNases. Another *N. alata* style-expressed protein, the 8 kDa NaPI proteinase inhibitor (Atkinson et al., 1993), was chosen as a negative control for these experiments, as NaPI is not involved in self-incompatibility and an antibody for this protein was available.

The ability of rabbit polyclonal antibodies raised to either deglycosylated S_2-RNase or S_7-RNase from *N. alata* to recognize other *N. alata* S-RNases in the Melbourne collection was tested using total style extracts of each *S* genotype. Figure 4.7 shows that anti-S_7-RNase antibody detected the S_3-, S_6- and S_7-RNases, weakly detected S_1-RNase and only detected S_2-RNase when long exposure times were used. The anti-S_2-RNase antibody detected the S_2-, S_6- and S_7-RNases but only weakly detected the S_1- and S_3-RNases. A Coomassie-stained gel loaded with equal amounts of stylar protein shows that the S-RNases were not equally abundant in the various extracts (Figure 4.7). However, differences in S-RNase abundance do not explain the differences in antibody binding. For example, despite being an abundant protein, S_2-RNase was barely detectable with the anti-S_7-RNase antibody (Figure 4.7).

Figure 4.8 shows a series of Co-IP assays with $(His)_6$:PiSLF1 FBD and total stylar extracts from S_1S_1, S_2S_2, S_3S_3, S_6S_6 and S_7S_7 *N. alata* plants. An immunoblot with anti-S_2-RNase antibody showed that S_1-RNase was retained on the beads when $(His)_6$:PiSLF1 FBD was present but not when it was omitted from the Co-IP, suggesting an interaction between S_1-RNase and $(His)_6$:PiSLF1 FBD. Additional bands of about 50 kDa and 25 kDa seen in the immunoblots developed with the anti-S_2- and anti-S_7-RNase antibodies presumably represent the IgG heavy and light chains of anti-$(His)_6$ antibody, respectively. Binding to IgG light chain in the anti-S_7-RNase antibody immunoblot obscured the binding to S_1-RNase. Similar interactions were seen with the S_2-, S_3-, S_6- and S_7-RNases, which were all retained on the beads when $(His)_6$:PiSLF1 FBD was present but not when it was omitted. By contrast, binding of the stylar protein NaPI was independent of $(His)_6$:PiSLF1 FBD, although more NaPI appeared to bind when this protein was present than when it was omitted.

Further investigations were done using S-RNases fractions enriched from style extracts by ammonium sulphate precipitation followed by cation exchange chromatography. This was done for the S_2-, S_3- and S_7-RNases but not for the S_1- and S_6-RNases, as insufficient plant material was available.

Figure 4.9 shows a series of Co-IP assays with $(His)_6$:PiSLF1 FBD, NaPI and enriched S_2-, S_3- and S_7-RNase fractions. The S_2- and S_3-RNases did not interact non-specifically with the protein A beads when $(His)_6$:PiSLF1 FBD was omitted from the Co-IP but weak binding of S_7-RNase to the beads was observed in the anti-S_7-RNase immunoblot. Similarly, weak binding of NaPI was detected in the presence of $(His)_6$:PiSLF1 FBD and anti-$(His)_6$ antibody but not binding of $(His)_6$:PiSLF1 FBD or S_2-RNase was detected when the anti-$(His)_6$ antibody was omitted.

Of the S-RNases tested, both S_2-RNase and S_7-RNase interacted with $(His)_6$:PiSLF1 FBD. Binding of these S-RNases to $(His)_6$:PiSLF1 FBD was detected in the anti-S_2-RNase immunoblot; as expected only S_7-RNase binding was detected in the anti-S_7-RNase immunoblot. Neither immunoblot detected an interaction between $(His)_6$:PiSLF1 FBD and S_3-RNase despite both antibodies recognizing this protein. This indicated that $(His)_6$:PiSLF1 FBD and S_3-RNase do not interact.

Figure 4.10 shows a series of Co-IP assays with $(His)_6$:DD6 FBD (S_2 allele), purified NaPI and fractions enriched in the S_2-, S_3- and S_7-RNases. None of the S-RNases detectably bound to the beads in the absence of $(His)_6$:DD6 FBD and neither S_2-RNase nor $(His)_6$:DD6 FBD remained on the beads when the anti-$(His)_6$ tag antibody was omitted. $(His)_6$:DD6 FBD did not detectably interact with NaPI.

Co-IP found interactions between $(His)_6$:DD6 FBD and S_2- and S_7-RNase but not S_3-RNase. Binding of S_2- and S_7-RNase to $(His)_6$:DD6 FBD was observed in the anti-S_2-RNase antibody immunoblot and S_7-

RNase was observed in the anti-S_7-RNase antibody immunoblot (Figure 4.10, upper panel). These interactions were also observed when the Co-IPs were examined on a Coomassie-stained gel (Figure 4.10, lower panel). As with $(His)_6$:PiSLF1 FBD, no interaction was detected between $(His)_6$:DD6 FBD and S_3-RNase, either on immunoblots or a Coomassie-stained gel.

Co-IP assays were also done with pairs of S-RNases that were present in similar amounts (Figure 4.10). Differences in protein size and in binding to the anti-S-RNase antibodies made it possible to determine which S-RNase had been retained in the assay.

When S_2- and S_3-RNase were both present, only S_2-RNase bound to $(His)_6$:DD6 FBD. S_3-RNase migrates slower than S_2-RNase on protein gels and was detected by both anti-S-RNase antibodies. The S-RNase that bound to $(His)_6$:DD6 FBD co-migrated with S_2-RNase and was detected with anti-S_2-RNase antibody but not anti-S_7-RNase antibody, consistent with the expected behaviour of S_2-RNase. S_3-RNase was not detected on the anti-S_7-RNase antibody immunoblot. Similarly, when the S_3- and S_7-RNases were present together in the Co-IP, $(His)_6$:DD6 FBD only bound to the faster migrating S_7-RNase and did not detectably bind the S_3-RNase. Finally, when S_2- and S_7-RNase were present, both likely bound to $(His)_6$:DD6 FBD. S_7-RNase was certainly bound, as a band of the expected size was detected by the anti-S_7-RNase antibody (which detects S_7-RNase but not S_2-RNase). Although S_2- and S_7-RNases are close to each other in size, the band in the Coomassie-stained gel was broader than the corresponding band in lanes where only S_7-RNase was bound, suggesting the slightly faster migrating S_2-RNase had also bound.

Collectively, both $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD interacted with S_2- and S_7-RNase but not with S_3-RNase or NaPI. These data suggest that the *in vitro* folded $(His)_6$:DD6 FBD and $(His)_6$:PiSLF1 FBD were functional and interacted specifically with some S-RNases.

The binding of S-RNases and $(His)_6$:PiSLF1 FBD was further characterized in a series of Co-IP assays that were loaded with increasing amounts of these two proteins in a constant ratio and a fixed amount of anti-$(His)_6$ antibody. Figure 4.11 (upper panel) shows that the binding of S_2-RNase and $(His)_6$:PiSLF1 FBD to the protein A beads increased in parallel as the amount of input proteins added to the Co-IP assay increased. This was seen in immunoblots probed with anti-$(His)_6$ and anti-S_2-RNase antibody and in Coomassie-stained gels and indicated that retention of $(His)_6$:PiSLF1 FBD and S_2-RNase on the beads depended on the amount of these proteins present in the Co-IP, as expected for a specific interaction. A similar linear response to increasing protein was seen between $(His)_6$:PiSLF1 FBD and S_7-RNase (Figure 4.11, lower panel).

Figure 4.12 shows a repeat experiment of the one shown in Figure 4.11, except that the amount of $(His)_6$:PiSLF1 FBD and anti-$(His)_6$ antibody was held constant and only the amount of S-RNase added to the Co-IP varied (i.e., the ratio of S-RNase to $(His)_6$:PiSLF1 FBD changed). Figure 4.12 (upper panel) shows a series of Co-IP assays done using from 2 µg to 10 µg of S_2-RNase and 10 µg $(His)_6$:PiSLF1 FBD. In immunoblots probed with anti-S_2-RNase antibody and in Coomassie-stained gels, equal amounts of $(His)_6$:PiSLF1 FBD were seen in each Co-IP but the level of S_2-RNase appeared not to increase when more than 2 µg of S_2-RNase was added. A similar result was seen with S_7-RNase, with no increase in band intensity detected when more than 4 µg of S-RNase was added (Figure 4.12, lower panel). This indicate that the binding capacity of $(His)_6$:PiSLF1 FBD for each S-RNase tested was saturable.

4.4: Discussion

The experiments in this chapter show that insoluble SLF proteins produced in *E. coli* can be refolded into functional SLF proteins. Functional in this context means that the refolded SLFs were capable of selectively binding S-RNases. Binding selectivity was shown through the ability of $(His)_6$:PiSLF1 FBD to interact with the *N. alata* S_1-, S_2-, S_6- and S_7-RNases and $(His)_6$:DD6 FBD (S_2 allele) to interact with the S_2- and S_7-RNases. Neither protein could interact with S_3-RNase or with other stylar proteins such as NaPI. Selectivity occurred whether the S-RNases were added to the Co-IP in a stylar extract (Figure 4.8) or in a largely pure form (Figures 4.9-4.12). According to the collaborative non-self recognition model each SLF gene encodes a unique protein capable of recognizing some of the numerous different S-RNase forms present in a species (Kubo et al., 2010). The ability of each type of SLF to interact with a subset of non-self S-RNases (i.e., S-RNases from other *S* haplotypes), but never with self S-RNases (i.e., the S-RNase encoded by the same *S* haplotype), is an essential feature of this model.

Very few conditions were found where $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD could be solubilised and refolded from IBs, as both proteins usually precipitated when placed in neutral pH solutions (Figure 4.1). Unfortunately, there is no "universal" refolding buffer and selection and optimization of buffer conditions must be empirically determined for each protein, with the general advice being to use a buffer pH that is at least one pH unit away from the protein's isoelectric point (pI), as at this pH the protein will have zero net charge and be most prone to precipitation (Middelberg 2002). Thus PiSLF1 and DD6, which have theoretical pIs of 5.6 and 5.1 respectively, should have been soluble in neutral buffers, which are often recommended as a 'good starting point' for resolubilising and refolding IB proteins. However a recent study found that most IB proteins precipitated when placed at pH 7.4 (Coutard et al., 2012): these workers found that proteins with acidic pIs refolded best when placed in alkaline buffers and proteins with alkaline pIs were more stable in acidic buffers. Thus testing a

broader range of pH values, up to four units away from the pIs of these recombinant proteins, may lead to the identification of a bigger range of conditions suitable for refolding.

Similarly the presence of metal ions such as Fe^{3+} also appeared beneficial to solubility. More than 30% of all proteins coordinate a metal, although most often this is to fulfil some structural or enzymatic role (Gray, 2003). While it is unknown if PiSLF1 or DD6 do this, F-box proteins are known to be modified, including by iron coordination (such as with the F-Box and Leucine-Rich Repeat Protein FBXL5), in response to stimuli that increase F-box protein stability and substrate degradation (Skaar et al., 2013). In addition, metals can also interact with proteins during polypeptide folding and under these circumstances help guide and sometimes even be essential for the folding process (Sedlak et al., 2008). A role for metals such as Fe^{3+} in SLF function is a hypothesis that requires further investigation.

Although soluble, functional proteins were produced, it was evident that refolded $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD existed in solution as higher order structures, not as monomers. Over time (weeks) the recombinant proteins would slowly precipitate from the solution. Most likely the oligomers formed through a process called 'domain swapping', in which two identical proteins exchange a part of their structure to form an intertwined dimer or higher-order structure (Bennett et al., 1995). The overall organization of a domain-swapped oligomer is identical to that of the monomer except it is formed through inter-molecular rather than intra-molecular folding. Additionally, there may also be changes to the structure of the region that connects the exchanged domain to the rest of the protein, the so-called "hinge loop" region (Bennett and Eisenberg, 2004). Domain swapping proteins commonly exchange only a single secondary structure element such as a β-strand or α-helix at one end of the protein. If this exchange occurs in a reciprocal manner between two monomers, a dimer is formed. Trimers, tetramers and so on are formed through cyclical exchanges of the structural element between three or more monomers. As the number of monomers increases, the oligomer essentially takes on an open-ended structure, where one end is left free to interact with additional monomers and further extension occurs autocatalytically (Liu and Eisenberg, 2002). Plausibly the $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD oligomers seen by SEC and AUC analysis (Figures 4.3 and 4.4) were formed through such a domain swapping mechanism. Another example where this phenomenon has been seen is with a thermostable α-amylase from the extremophile bacteria, *Halothermothrix orenii* (Sivakumar et al., 2006). When no salt is present, the protein has a strong tendency to form very large poly-dispersed aggregates of around 5,000 kDa that are still enzymatically active. The addition of low concentrations of NaCl reverses aggregation, leading to the formation of α-amylase monomers. Both the poly-dispersed aggregates and

monomers have identical secondary structures, as measured by CD spectroscopy.Some of the physical properties of the refolded $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD, such as their oligomeric nature, high content of β-sheet and resistance to heating, were similar to those of amyloids, misfolded globular proteins that are also thermally stable and rich in β-content due to the presence of intermolecular arrays of parallel β-sheets, the so-called cross-β structure (Nelson et al., 2005). In one sense the refolded proteins can be considered amyloids, although not in the sense of being 'misfolded' toxic protein structures; rather, the proteins have formed low-energy quaternary structures that are still functional. Indeed, despite being associated with neurodegenerative diseases such as Alzheimer, Parkinson and Huntington disease, functional amyloids, biologically active proteins that use the amyloids' unique mechanical properties, have been reported in a wide range of organisms, from bacteria to mammals (Bennett et al., 2006; Fowler et al., 2007; Pham et al., 2014). Examples of plant-encoded amyloids are much less uncommon although some have been reported (Villar-Piqué et al., 2010). Further investigation of these SLF amyloids, for example determining whether they stain with the amyloidophilic fluorophores thioflavin and Congo red, was not done; nor was it investigated whether other refolding conditions would help circumvent the formation of these amyloids.

The *Petunia* protein PiSLF1 (formerly called $PiSLF_1$, the SLF from the *P. inflata* S_1 allele) was used as a positive control in these experiments, as earlier *in vitro* assays showed it to be an S-RNase interactor that preferentially binds non-self S-RNases more than self S-RNases (Hua et al., 2007). This ability of recombinant PiSLF1 to bind S-RNases was confirmed, but whether it bound its cognate *Petunia* S_1-RNase with lower avidity than non-self S-RNases was not tested. DD6 was also shown to bind S-RNases in a manner similar to PiSLF1. Thus the project's initial aim, which had been to use the *in vitro* binding assay to identify which *DD* gene encoded the *N. alata* homolog of PiSLF1, was achieved. However, little difference was noted in interactions between DD6 and self (S_2) and non-self (S_7) RNases. Indeed, experiments with crude *E. coli* extracts reported in Chapter 3 indicated that a range of $(His)_6$-tagged DDs could bind native *N. alata* S_2- and S_7-RNases with similar affinity (Figures 3.9 and 3.10). Earlier work had used an *in vitro* assay to distinguished SLFs from SLF-Likes (SLFLs), with SLFLs either failing to interact with or unable to compete with SLFs for binding to an S-RNase (Hua et al., 2007). A major point of departure, therefore, between the results of earlier papers and those reported here, was this ability of all tested F-box proteins to bind to all S-RNases other than the *N. alata* S_3-RNase.

According to the collaborative non-self recognition model, each functional SLF protein interacts with one or two different S-RNase alleles but never interacts with its cognate S-RNase (Kubo et al., 2010).

This model is based on two types of evidence, the first arising from genetic interactions in transgenic plants (called *in vivo* experiments) and the second from immunoprecipitation experiments similar to those done here (*in vitro* experiments). The *in vivo* experiment is based on a phenomenon called 'competitive interaction', where the SI response of pollen in a self-incompatible plant breaks down when all or part of an additional S allele is present in the plant's genome (Golz et al., 2000, 2001; see Chapter 1). In the *in vivo* assay, a plant is transformed with an *SLF* transgene of a particular class and the presence of a competitive interaction monitored by placing the plant's pollen on an otherwise incompatible stigma (such as a self-pollination). When the pollination is compatible, this is taken to mean that that the particular SLF used can detoxify one of the plant's stylar S-RNases (see Kubo et al., 2010; Williams et al., 2014b for examples). For instance, competitive interaction was noted when a transgene containing the *Petunia* S_7 allele of a type-2 SLF (*S_7-SLF2*) was expressed in a *Petunia* plant with the S_9 allele, as a result the S_7-SLF2 was said to mediate detoxification of the S_9-RNase. Using this *in vivo* approach S_7-SLF2 was shown to be able to detoxify the *Petunia* S_9- and S_{11}-RNases but not the S_5- and S_7-RNases. So far S_2-SLF1 (the S_2 allele of the PiSLF1 protein used in this thesis) competitively interacts with the largest number of tested *Petunia* S alleles (four out of seven; Williams et al., 2014b).

Obviously the *in vitro* assay explores detoxification further by demonstrating a direct interaction between an SLF and its target S-RNase. Kubo et al. (2010) for instance used an extract of transgenic pollen expressing FLAG-tagged S_7-SLF2 to show binding with the *Petunia* S_9- and S_{11}-RNases but not the S_5- and S_7-RNases (all S-RNases were native proteins present in a stylar extract). In a follow-up study, an *in vitro* ubiquitination assay using extracts of FLAG:S_7-SLF2-expressing pollen showed mono- and polyubiquitination of the S_9- and S_{11}-RNases but not the S_5- and S_7-RNases (Entani et al., 2014). In the case of *Petunia* S_7-SLF2, therefore, the *in vivo* and *in vitro* evidence both support an interaction with some S-RNases but not others that is consistent with the collaborative non-self recognition model. But *in vitro* assays have not been done for most SLFs and where they have, are not consistent with the current model. For example *Petunia* S_2-SLF1 (formerly known as Pi SLF_2) *in vitro* interacted with the S_1- and S_2-RNases (Hua and Kao 2006) but *in vivo* showed a competitive interaction with the S_1 allele but not the S_2 allele (Sijacic et al., 2004).

Here $(His)_6$:DD6 FBD interacted with the *N. alata* S_2- and S_7-RNases but not the S_3-RNase, a result that is potentially consistent with collaborative non-self recognition model and could be further explored in an *in vivo* experiment. However the DD6 protein used was encoded by the S_2 allele (Wheeler and Newbigin 2007), and the collaborative non-self recognition model predicts it should not interact with its cognate S_2-RNase. Indeed, the collaborative non-self recognition model requires

that the expression of any *SLF* types that recognise cognate S-RNases is suppressed or that the S allele has either a divergent form of the *SLF* type or carries a deletion in this gene. *DD6* was expressed in pollen from all *N. alata S* allele backgrounds tested and sequence differences between alleles were ≤5% (Wheeler and Newbigin, 2007).

There are a further three reasons to question the conclusion that the DD6/S-RNase interaction satisfies the expectations of the collaborative non-self recognition model. First, $(His)_6$:DD6 FBD interacted with the same S-RNases as $(His)_6$:PiSLF1 FBD. That is, even though PiSLF1 and DD6 are from different groups of SLFs (DD6 sits with the *Petunia* SLF8 clade; see Figure 2.10 and Williams et al., 2014a), there was no evidence for each type of SLF interacting with a different subset of S-RNases.

The second reason for questioning this conclusion is that the failure of $(His)_6$:DD6 FBD and $(His)_6$:PiSLF1 FBD to interact with S_3-RNase could plausibly be due to glycosylation. All solanaceous S-RNases so far described are glycosylated and carbohydrate chains are likely to be a major feature on their surface (Oxley and Bacic 1995). But there is considerable variation in the number and structure of attached glycan sidechains. For example the *N. alata* S_1-RNase has a single potential *N*-glycosylation site (Asn-X-Ser/Thr) whereas four sites exist on the *N. alata* S_2-, S_6- and S_7-RNases (Woodward et al., 1992; Vissers et al., 1995). The *N. alata* S_3-RNase is the most heavily glycosylated of these, having the four potential sites present in the S_2-, S_6- and S_7-RNases plus an additional site located centrally in the protein between conserved domain 3, which is part of the active site, and conserved domain 4 (Oxley et al., 1996). If this region is important for interactions with SLFs, then the presence of an *N*-linked glycan could prevent binding and explain the observed selectivity. Deletion of the F-box domain should not affect its ability to interact with S_3-RNase as this domain interacts with Skp1 in known SCF^{E3} ligases (Viestra, 2003). Moreover, as interaction assays were performed with mixtures of S-RNases, it is unlikely the oligomerisation of $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD affected the ability to interact with S_3-RNase. Although plausibly the failure of the recombinant proteins to interact with S_3-RNase could be an artefact of their non-native and aggregate state, their ability to interact specifically with other S-RNases tends to suggest that this is not the case.

The final reason for raising doubts about the conclusion is that *DD6* is one of a cluster of three *DD* genes that are at least 0.9 cM from the S locus (Wheeler and Newbigin 2007). Historically, SI researchers have understood the need for recombination within the S locus to be suppressed in order to maintain functional associations between the pollen and stylar components of an S allele, with any change in stylar specificity not being selectively advantageous unless it occurs

simultaneously with changes affecting the pollen specificity (and vice versa, e.g., see Lewis 1960). For this reason Wheeler and Newbigin (2007) excluded *DD6* from consideration as the *N. alata* ortholog of *pollen S*, instead focussing on *DD2*, *DD7* and *DD10*, which map to the same region of the chromosome as *pollen S*, and *DD4*, *DD5* and *DD8*, which could not be mapped because of a lack of suitable polymorphisms. As rejecting this historic view about recombination at this *S* locus would be challenging, excluding DD6 as a pollen S candidate suggests that the ability to bind S-RNases can no longer be considered a defining property of proteins involved in RNase-based SI systems.

Interestingly, a similar issue is raised in a recent paper on a reproductive barrier related to SI called unilateral incompatibility (UI; Li and Chetelat 2015). Under UI pollen from one species is rejected by the styles of a related species, whereas in the reciprocal cross, no pollen rejection occurs. Typically the pollen from a self-compatible species is rejected by the styles of a related self-incompatible species whereas pollen rejection rarely occurs when the reciprocal cross is performed (self-compatible species pollinated by the self-incompatible one). This unidirectional pattern of pollen rejection is referred to as the "SI × SC rule' and it is known that several SI-related factors, including S-RNase and CUL1, are involved (Murfett et al., 1996; Li and Chetelat 2014).

In the solanaceous species *Solanum pennellii* the *ui1.1* locus encodes one of two pollen factors that are required for UI (the other locus *ui6.1* encodes CULLIN1). Pollen lacking *ui1.1* are incompatible on styles that express S-RNases, suggesting that *ui1.1* encodes a factor that is required for resistance to S-RNase–based rejection. The *ui1.1* locus maps to a 43.2-Mbp interval at the *S. pennellii* S locus, an interval that includes 23 genes encoding pollen-expressed SLFs. Transformations into transgenic plants were used to test for *ui1.1* function and of the *S. pennellii SLF* genes tested only one, *SpSLF-23*, showed the compatible pollen phenotype consistent with *ui1.1* function. Moreover, the pollen compatibility phenotype was seen when transgenic pollen was placed on styles expressing various S-RNases, suggesting that the SpSLF-23 protein is capable of recognising many different S-RNases (Li and Chetelat 2015).

As shown in chapter 3, soluble recombinant PiSLF produced in *E. coli* is likely in a non-native form and therefore the pull down assay results cannot be interpreted without ambiguity. Hence, further work focus on exploring an alternative to obtain functional protein. A protein refolding protocol was developed for PiSLF1 and used on DD6, both were shown to be functional despite not in their monomer form. Both interacts with a specific subset of S-RNases but not control protein, NaPI. It is likely the same protocol can be used to obtain functional DD2, 5, 7 and 8. The interaction results do not satisfy the collaborative non-self recognition theory as explained earlier. The genetics of SLF and S-RNase on the *S* locus is an essential criteria used to determine the inclusion or exclusion of a

candidate gene. DD6 is considered as a non-SLF due to its position at the *S* locus is found to interact with S-RNase meant biochemical interaction test may no longer be a conclusive test for SLF. Hence, it is important for *in vivo* experiment to be performed using PiSLF and DD6 in *N. alata* various *S* background plants to understand DD6 role in SI.

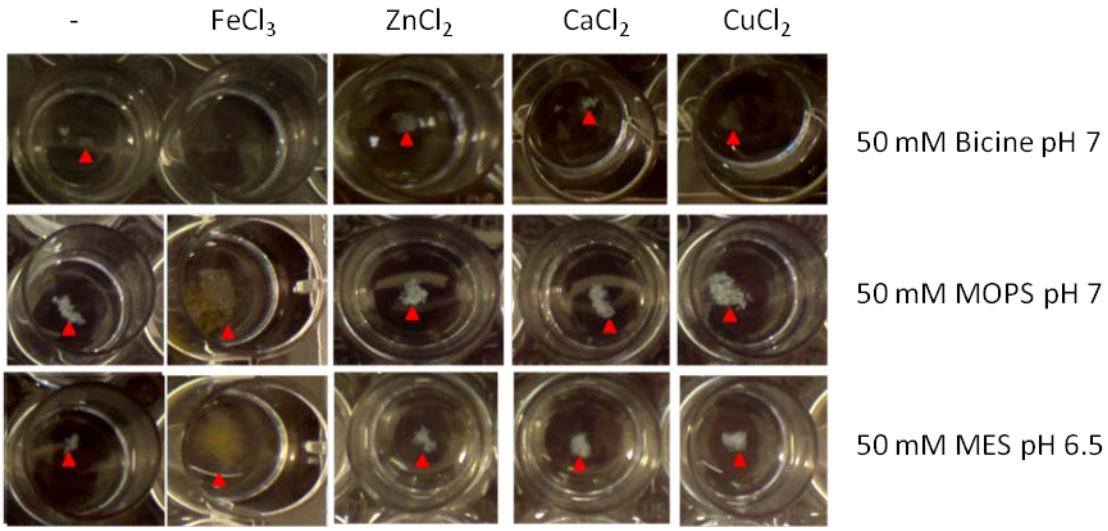
Figure 4.1: Screen of conditions for folding $(His)_6$:PiSLF1 FBD.

A) Assay used to detect folding of $(His)_6$:PiSLF1 FBD. If the protein was compatible with the buffer and additive combination being tested, it remained soluble and no precipitate was visible after 3 hr. Red arrows indicate wells in which a visible precipitate has formed.

B) List of buffers and metal salts tested for compatibility with folding of $(His)_6$:PiSLF1 FBD. + indicates a precipitate was observed and a – indicates no precipitate was observed. NT: conditions not tested.

C) List of buffers and non-metal additives tested for compatibility with folding of $(His)_6$:PiSLF1 FBD. All metals were at 50 μM. + indicates a precipitate was observed.

A)



B	-	$CaCl_2$	$ZnCl_2$	$CuCl_2$	$FeCl_3$	KCl	$MgCl_2$
Water	-	-	-	-	-	NT	NT
50 mM Bicine, pH 7.6	+	+	+	+	-	+	+
50 mM MOPS, pH 7	+	+	+	+	+	+	+
50mM MES, pH 6.5	+	+	+	+	+	+	+
50 mM Tricine, pH 7.6	+	+	+	+	+	+	+
50 mM HEPES, pH 7	+	+	+	+	+	+	+

C	10% glycerol	20% glycerol	PEG	Tween 20	Tween 80	Nonidet P40	Triton X100
50 mM Bicine, pH 7.6	+	+	+	+	+	+	+
50 mM MOPS, pH 7	+	+	+	+	+	+	+
50mM MES, pH 6.5	+	+	+	+	+	+	+
50 mM Tricine, pH 7.6	+	+	+	+	+	+	+
50 mM HEPES, pH 7	+	+	+	+	+	+	+

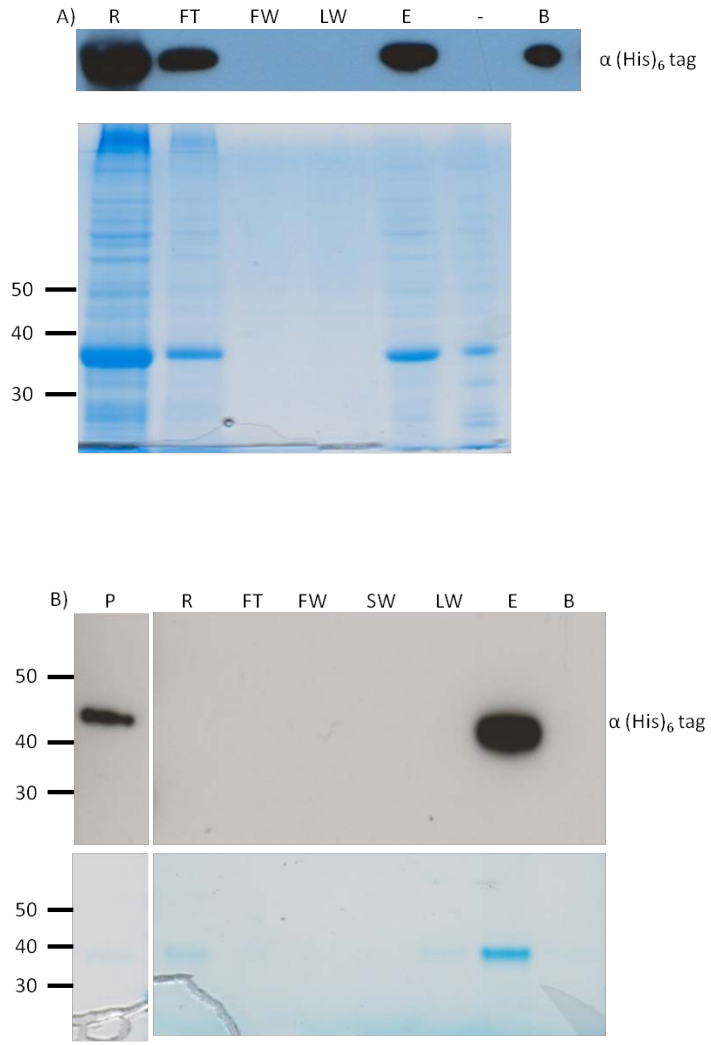


Figure 4.2: Recovery of folded $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD protein using Ni-affinity chromatography.

A) Ni-affinity chromatography of folded $(His)_6$:PiSLF1 FBD. The proteins present in the various fractions were detected by immunoblot analysis using an anti-$(His)_6$ tag antibody and Coomassie staining. R: folded protein before purification; FT: column flow through; FW: first column wash with 20 mM imidazole; LW: last column wash with 100 mM imidazole; E: protein eluate with 200 mM imidazole; B: protein remaining on the resin after elution. Numbers to the left of the figure are sizes in kDa.

B) Ni-affinity chromatography of folded $(His)_6$:DD6 FBD. The lanes are the same as in A except for P: insoluble pellet; and SW: Second column wash with 50 mM imidazole.

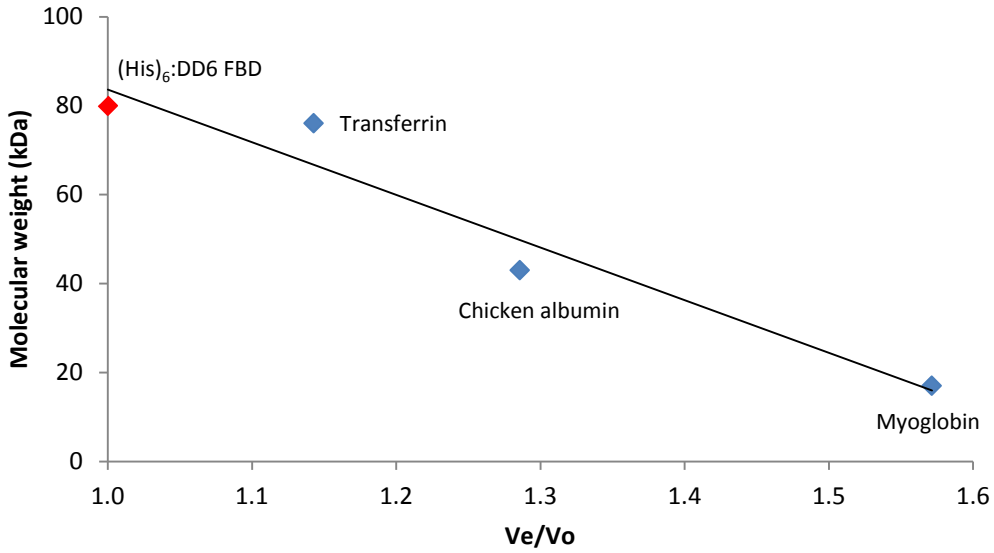


Figure 4.3: Elution profile of protein standards and $(His)_6$:DD6 FBD on the sephacryl S200 column. Protein standards used were transferrin (76 kDa), chicken albumin (45 kDa) and myoglobin (17 kDa). Ve/Vo is elution volume divided by void volume. Void volume was determined using dextran blue 2000.

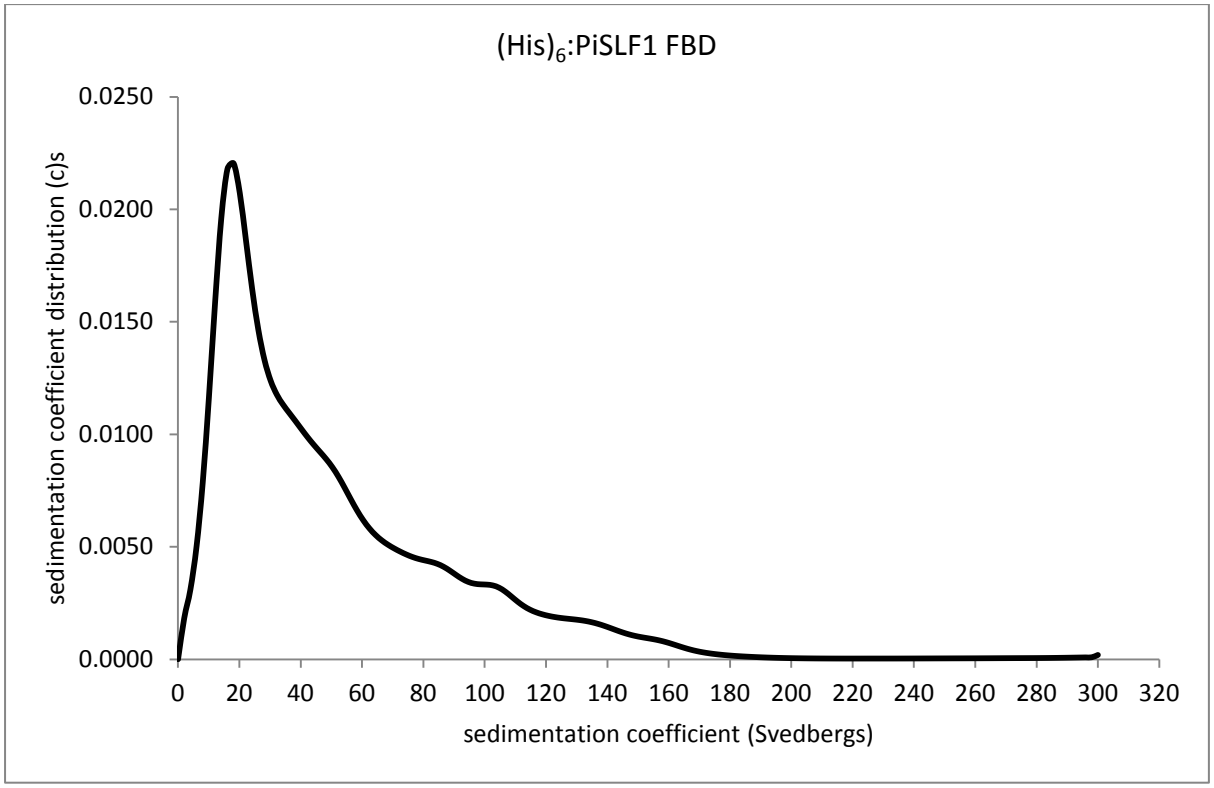
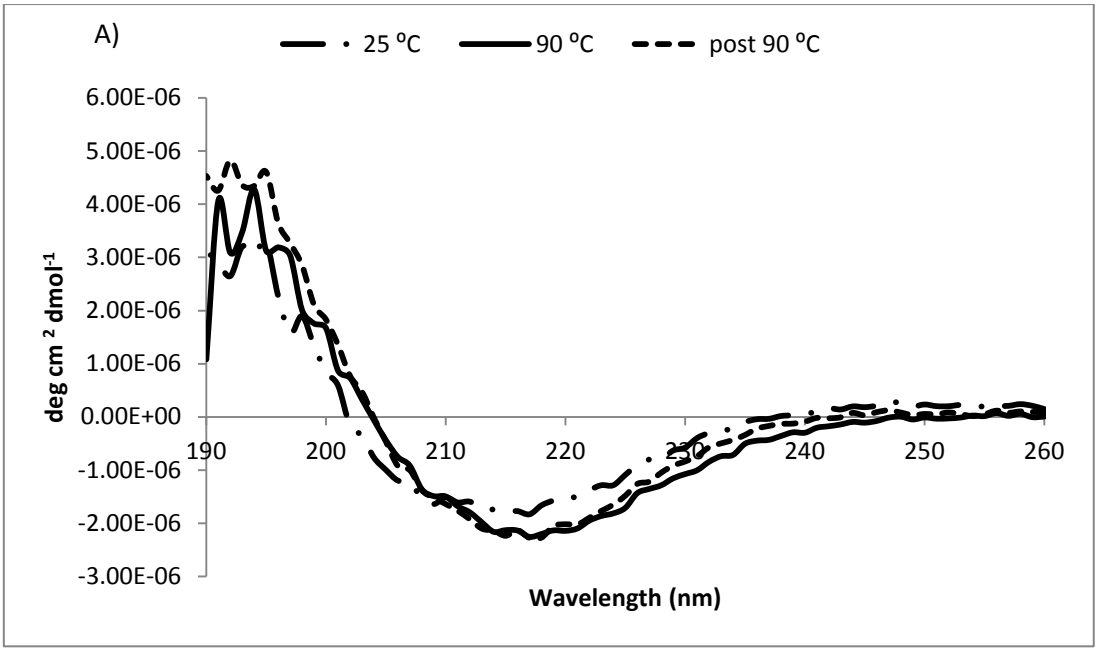


Figure 4.4: Analytical ultracentrifuge analysis of $(His)_6$:PiSLF1 FBD.

Migration of protein in the ultracentrifuge cell was monitored by UV absorbance at 280 nm. X axis shows the sedimentation coefficient based on the migration of protein species at 15,000 rpm and the Y axis shows the sedimentation coefficient distribution of protein.



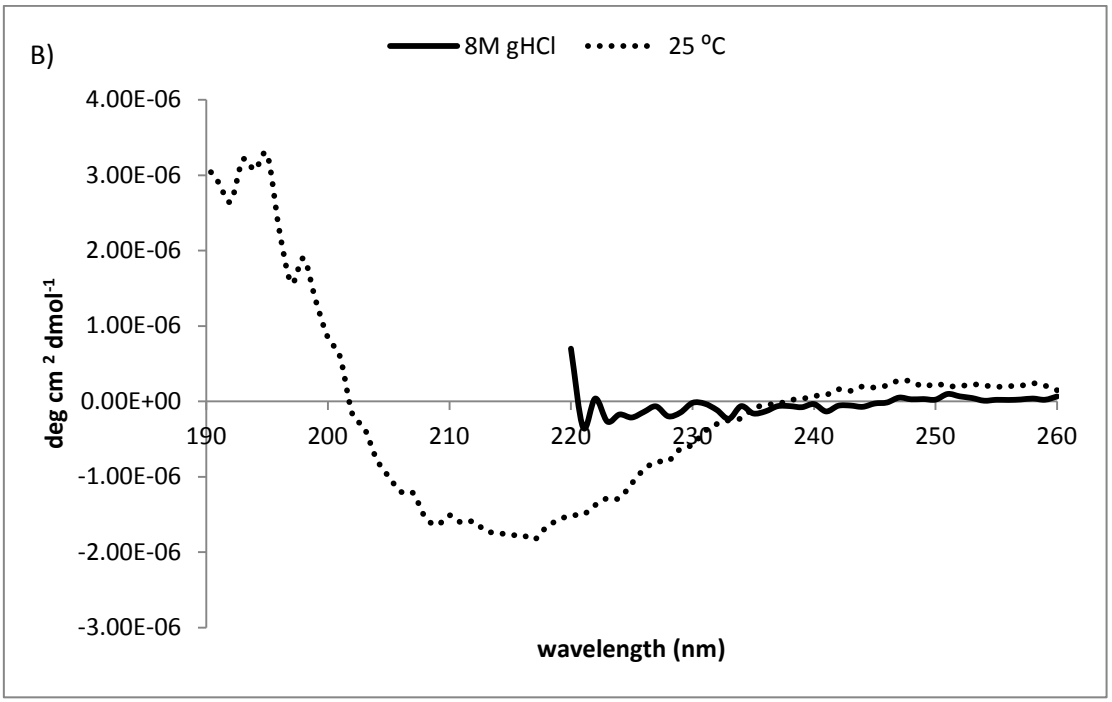


Figure 4.5: Far UV CD spectrum of folded $(His)_6$:PiSLF1 FBD.

A) CD spectrum at 25°C, 90°C and 5 min after incubation at 90°C (post 90°C). The buffer used was 50 mM Bicine (pH 7.6), 0.5 mM TCEP and 50 µM $FeCl_3$.

B) CD spectrum at 25°C and after treatment with 8 M guanidine hydrochloride (gHCl). Wavelengths <220 nm were not used because guanidine hydrochloride absorbs strongly in this region of the spectrum.

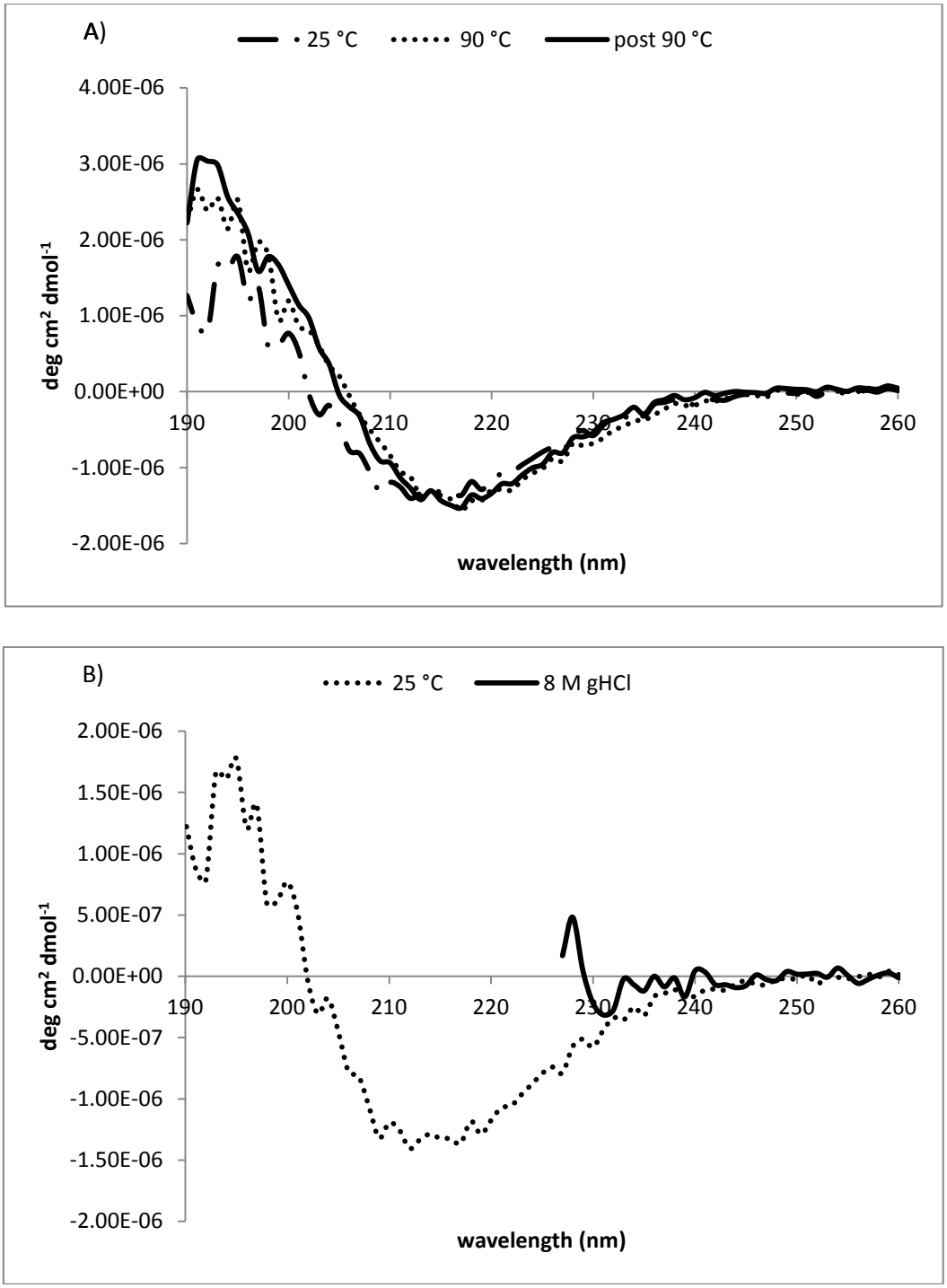


Figure 4.6: Far UV CD spectrum of folded $(His)_6$:DD6 FBD

A) CD spectrum at 25°C, 90°C and 5 min after incubation at 90°C (post 90°C). The buffer used was 50 mM Bicine (pH 7.6), 0.5 mM TCEP and 50 µM $FeCl_3$.

B) CD spectrum at 25°C or after treatment with 8 M guanidine hydrochloride. Wavelengths <220 nm were not used because guanidine hydrochloride absorbs strongly in this region of the spectrum.

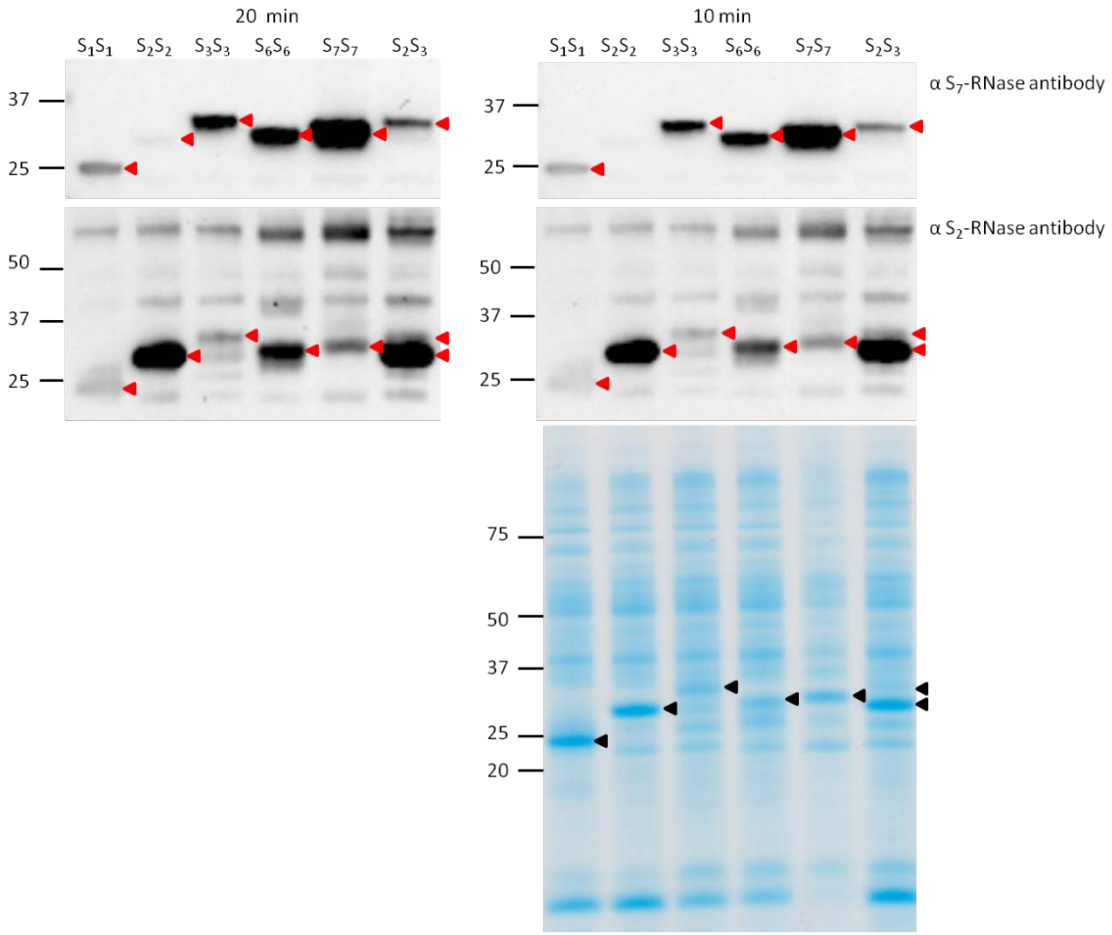


Figure 4.7: Specificity of two anti-S-RNase antibodies for various *N. alata* S-RNases. Equal amounts of total style extract (3 µg protein) from *N. alata* plants of the indicated *S* genotype were separated by SDS-PAGE and either stained with Coomassie blue or immunoblotted and probed with the indicated antibody. Red and black arrowheads indicate the position of each S-RNase on the immunoblot and Coomassie-stained gel, respectively. The digital imaging exposure time (10 or 20 min) is indicated above the immunoblots. Numbers to the left are sizes of molecular weight markers in kDa.

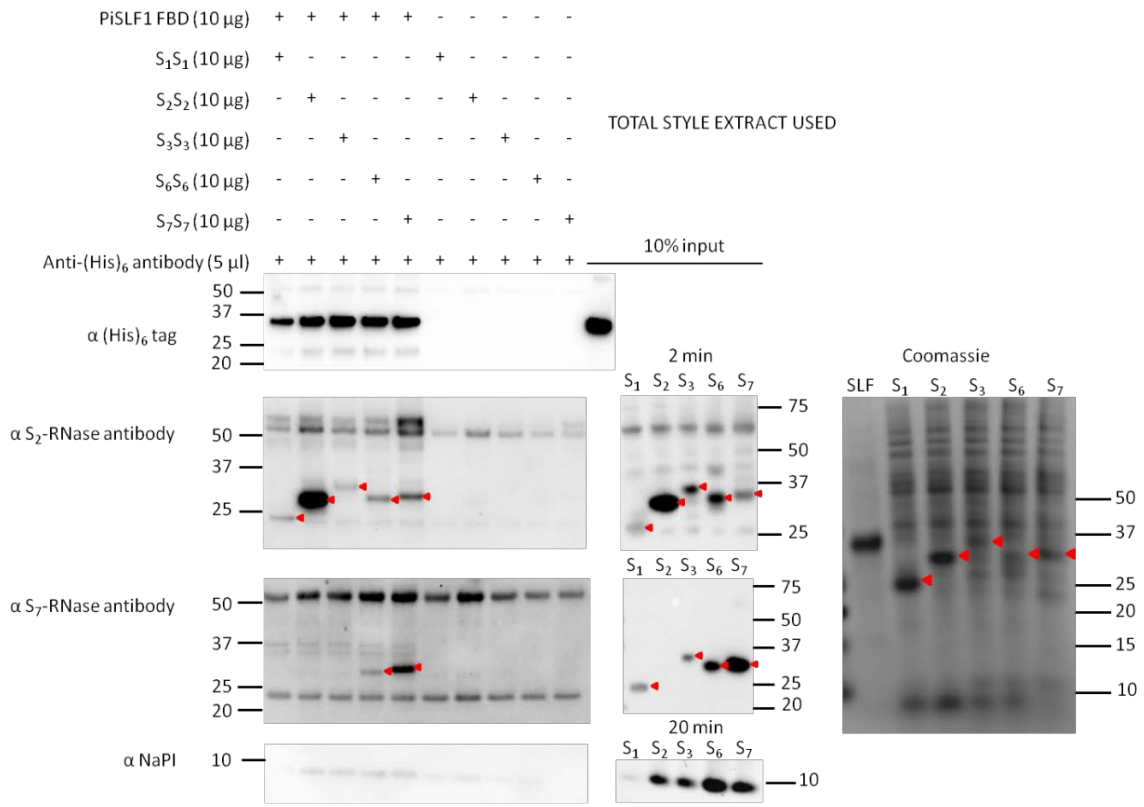


Figure 4.8: Co-IP assays using total *N. alata* style extracts and refolded, purified $(His)_6$:PiSLF1 FBD.

The components present in a Co-IP assay are indicated above the relevant lane: +, component added; -, component not added. All Co-IP assays contain equal amounts of style extract (10 µg) and anti-$(His)_6$ tag antibody. When added the same amount of refolded $(His)_6$:PiSLF1 FBD (10 µg) was used. Immunoblots and a Coomassie-stained gel of the input proteins are also shown. Position of each S-RNase is indicated by a red arrowhead. Numbers to the left or right are sizes of molecular weight markers in kDa.

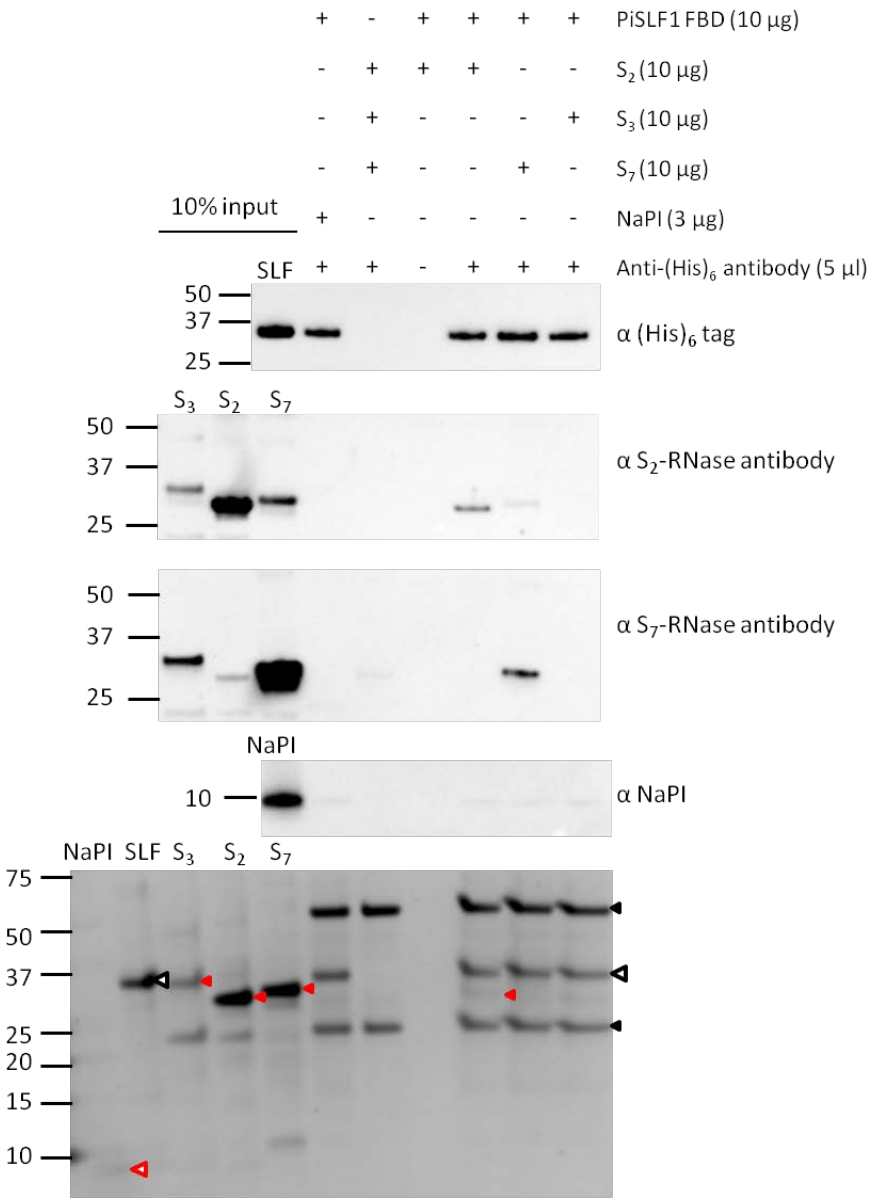


Figure 4.9: Co-IP assay using enriched *N. alata* S-RNases and $(His)_6$:PiSLF1 FBD.

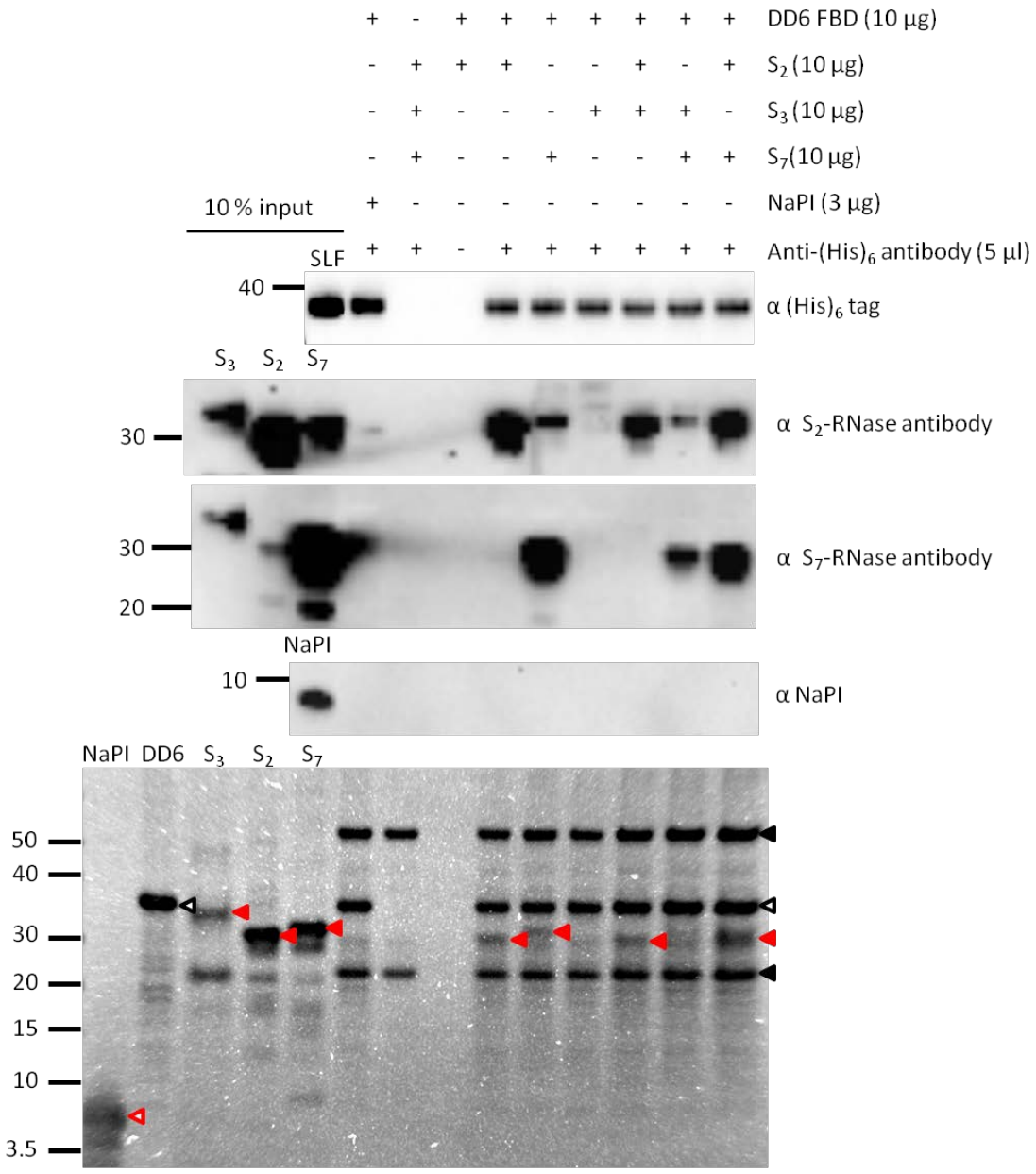
Co-IP assays were performed with the indicated S-RNase, NaPI (*N. alata* proteinase inhibitor) and $(His)_6$:PiSLF1 FBD. The components present in a Co-IP assay are indicated above the relevant lane: +, component added; -, component not added.

Replicate immunoblots were probed with the anti-$(His)_6$ tag antibody, the anti-S_7-RNase antibody, the anti-S_2-RNase antibody and the anti-NaPI antibody. Numbers to the left are sizes of molecular weight standards (in kDa). The black open arrowhead indicates $(His)_6$:PiSLF1 FBD in the lower panel and the red closed arrowheads indicate the S_2-, S_3- and S_7-RNases; black closed arrowheads indicate the anti-$(His)_6$ antibody heavy and light chains. An open red arrowhead indicates NaPI.

Figure 4.10: Co-IP assay using enriched *N. alata* S-RNases and $(His)_6$:DD6 FBD

Co-IP assays were performed with the indicated S-RNases, NaPI and $(His)_6$:DD6 FBD. The components present in each Co-IP are indicated above each lane: +, component added; -, component not added.

Replicate immunoblots (upper panels) were probed with the anti-$(His)_6$ tag antibody, the anti-S_7-RNase antibody, the anti-S_2-RNase antibody and the anti-NaPI antibody. Numbers to the left are sizes of molecular weight standards (in kDa). The lower panel is a replicate Coomassie-stained gel. The black open arrowhead indicates $(His)_6$:DD6 FBD in the lower panel and the red closed arrowheads indicate the S_2-, S_3- and S_7-RNases; black closed arrowheads indicate the anti-$(His)_6$ antibody heavy and light chains. An open red arrowhead indicates NaPI.


+ - + + + + + + + DD6 FBD (10 µg)
- + + + - - + - + S2 (10 µg)
- + - - - + + + - S3 (10 µg)
- + - - + - - + + S7(10 µg)
10 % input + - - - - - - - - NaPI (3 µg)
+ + - + + + + + + Anti-(His)6 antibody (5 µl)
SLF
40
α (His)6 tag
S3 S2 S7
30
α S2-RNase antibody
30
20
α S7-RNase antibody
NaPI
10
α NaPI
NaPI DD6 S3 S2 S7
50
40
30
20
15
10
3.5

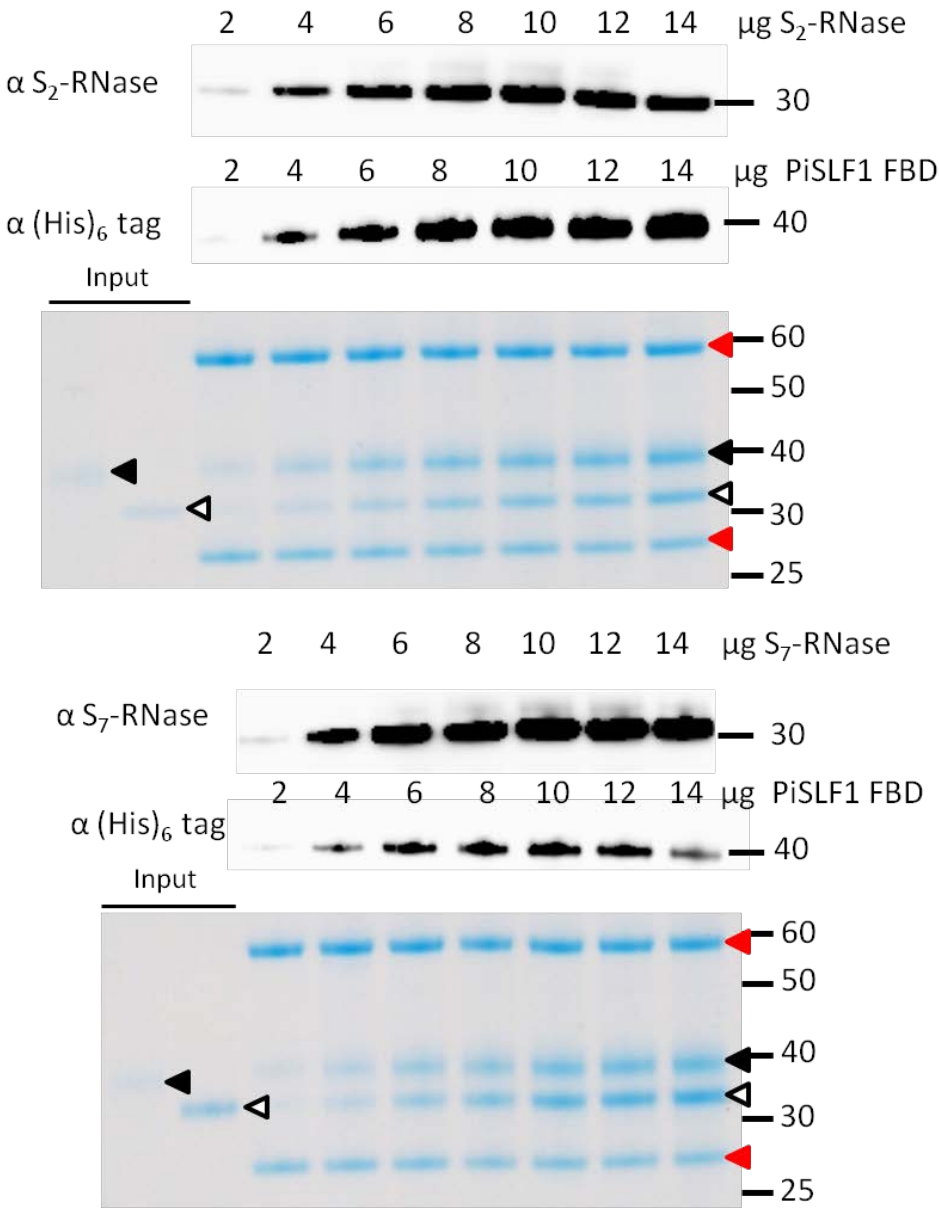


Figure 4.11: A series of Co-IP was performed with increasing amounts of $(His)_6$:PiSLF1 FBD and either S_2-RNase (upper panel) or S_7-RNase (lower panel). The amount of anti-$(His)_6$ antibody (3 µL) used in each Co-IP was held constant. Closed black arrowhead: $(His)_6$:PiSLF1 FBD; Open black arrowhead: S_2- or S_7-RNase; Closed red arrowhead: antibody heavy and light chains.

Replicate immunoblots were probed with the anti-$(His)_6$ tag antibody and either anti-S_2-RNase antibody (upper panel) or anti-S_7-RNase antibody (lower panel). Numbers to the right are sizes molecular weight standards (in kDa). Replicate Coomassie-stained gels are also shown.

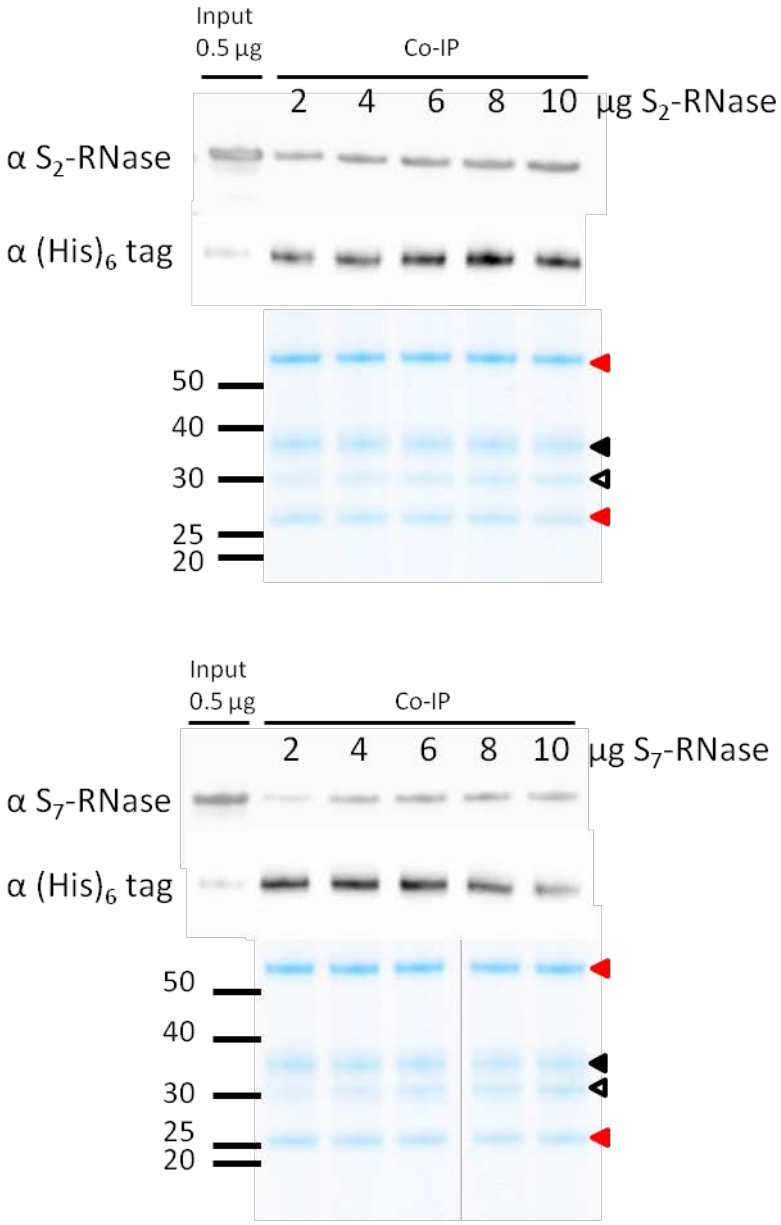


Figure 4.12: S-RNase titration assay.

Co-IP was performed using 10 µg of refolded $(His)_6$:PiSLF1 FBD with varying amounts of S_2- and S_7-RNase. Closed black arrowhead: $(His)_6$:PiSLF1 FBD; Opened black arrowhead: S_2- or S_7-RNase. Red arrowhead: antibody heavy (~55 kDa) and light chain (~26 kDa). Replicate Coomassie-stained gels are also shown. Numbers to the right are sizes molecular weight standards (in kDa).

5: Conclusions and future work

One aim of this thesis was to use next generation sequencing and de novo transcript assembly to perform a transcriptomic analysis on RNA extracted from *N. alata* pollen grains as a means of isolating previously unidentified *SLFs* and other RNase-based SI related transcripts reported by other studies. The results of this study were presented and discussed in chapter 2.

The second aim was to study interactions between the DDs and *Petunia* SLFs and the S-RNases using either pull down or co-immunoprecipitation assays. The approach taken was to express DDs/SLF and S-RNases in *E. coli* as recombinant proteins with tags to facilitate later purification steps. Interactions between SLFs and S-RNases were studied using purified recombinant proteins and the results are presented and discussed in chapters 3 and 4.

5.1: *Nicotiana alata* pollen transcriptome

Next generation sequencing successfully isolated new *SLF* candidates and other RNase-based SI-related genes including *Skp1*, *Cullin1*, *RBX1*, *SBP1* and these genes were shown to be expressed in pollen grains. Bioinformatics and molecular validation results showed that ~80% of *Nicotiana* pollen expressed transcripts were present in the reassembled *N. alata* transcriptome. The number of contigs present in pollen transciptomes of *A. thaliana* and *N. alata* are about the same and indicate that the *N. alata* assembly is likely not over or under represented, a conclusion supported by GO annotations which showed that contigs represent transcripts from various functional categories. Therefore, although partial the transcriptome of 6,800 contigs is considered to be representative of the transcripts present in *N. alata* pollen grains.

The discovery of chimeric contigs and detection of differential splicing were not unexpected. Chimera contigs were expected as a consequence of the approach used by the assembler. Chimeric contigs were, however, easily detected by searching against the GenBank database with matches to two or more unrelated proteins indicating a chimeric sequence. The presence of chimeric contigs did not affect the overall utility of transcriptome. Differential splicing of mRNA has also been detected in other plant transcriptome and is contributing to further understanding of mRNA regulation. Differential splicing is a form of mRNA regulation that is more prevalent than initially thought (Staiger 2015).

Surveying the *N. alata* pollen transcriptome revealed transcripts that encode components of a SCF^{SLF} E3 ligase complex are present and may have an important role in RNase-based SI, similar to the SCF^{SLF} E3 ligase complex reported in *Petunia* which also belongs to the Solanaceae family (Zhao et al., 2010). *In vivo* transgenic plant work has shown that tomato Cullin1 is required for pollen resistance

against non-self S-RNases and it is likely *N. alata* Cullin1 would have similar function, since the two species both possess RNase-based SI system and both are Solanaceous plants. With the identification of a possible SCF^{SLF} E3 ligase complex present in *N. alata*, an important next step is to perform *in vitro* and *in vivo* study to show that this complex is formed in pollen grain and is required for RNase-based SI in *N. alata*.

The first aim of this thesis was achieved as new DDs and SI-related transcripts which encode for products that are known to form a complex that is required for RNased-based SI were isolated from *N. alata* pollen transcriptome. Based on the results obtained, further work can be performed *in vitro* and *in vivo* which will contribute to the better understanding of RNase-based SI.

5.2: Protein-protein interaction study

The approach used to investigate which of the DDs interact with S-RNase was to express DD6, S_6-RNase, RNaseNE, *Petunia* SLF1 and S_2-RNase, as recombinant tagged proteins in *E. coli* as it is one of the most frequently used expression system for the production of recombinant protein. The other important reason for choosing this expression system was because similar expression work performed for PiSLF1 and *Petunia* S_2-RNase had successfully produced purified proteins. The production of a very small amount of soluble "sticky" $(His)_6$:PiSLF1 suggested the recombinant protein was likely not in its native form and hence any interactions obtained using "sticky" PiSLF1 and DD6 cannot be considered conclusive. This demonstrated that the full length F-box proteins are difficult to express in a prokaryote like *E. coli* and possibly a different expression system is needed to obtain functional recombinant protein. Alternatively removing domains that are not essential for interaction can aid in the production of soluble protein in *E. coli*. Indeed, this was shown to be possible as removing the F-box domain in $(His)_6$:PiSLF1 FBD resulted in a protein with improved solubility when expressed in BL21 star, an *E. coli* strain design for protein expression with enhanced mRNA stability. However, due to the low amount of soluble $(His)_6$:PiSLF1 FBD, further purification steps did not recover sufficient pure protein for an interaction assay. Although $(His)_6$:PiSLF1 FBD is still largely insoluble, a relatively pure protein can be obtained in sufficient amounts using, a protein refolding protocol. As shown by circular dichroism, refolded $(His)_6$:PiSLF1 FBD and $(His)_6$:DD6 FBD were β-structured proteins in agreement with theoretical structural predictions. Most importantly, the refolded proteins were functional as determined by their ability to specifically interact with only some of the S-RNases tested and not interact with a range of control proteins such as NaPI. This suggests that refolded SLFs proteins possess native functional characteristics and hence can interact with a specific subset of S-RNases, in agreement with the collaborative non-self recognition model.

$(His)_6$:DD6 FBD (derived from the *N. alata* S_2 allele) interacts with *N. alata* S_2-RNase, which is not expected as the collaborative non-self recognition model states that an SLF will not interact with its cognate S-RNase and must not be present in the same *S* haloptype. The polydispered nature of $(His)_6$:DD6 FBD presumably does not interfere with the specificity of its interaction with S-RNases. Hence, it is important to confirm whether DD6-S_2 is located near the S_2 allele as it was one of the putative SLFs amplified by PCR using degenerate primers (Wheeler and Newbigin, 2007).

S-RNases were degraded when expressed in *E. coli* making it impossible to obtained full-length protein in quantities suitable for further work. As S-RNases require disulphide bonding for protein activity, it is unlikely the recombinant protein produced in *E. coli* would be active since this expression system does not support disulphide bond formation. Since S-RNase is an abundant protein in the style, it is easier to obtain enriched native S-RNase from style tissue than to produce it in *E. coli*, hence further work used native S-RNases from total stylar extract or enriched S-RNases obtained from *N. alata* styles from different *S* background plants.

5.3: Future work

Overall, this study isolated transcripts which encode for individual components of a SCF^{SLF} E3 ligase complex in *N. alata* pollen grain. However, the complete suite of *DDs* remains to be isolated. Based on the interaction study refolded PiSLF1 and DD6 interact with S-RNases with specificity and importantly did not interact non-specifically with control proteins. Although the data suggest that PiSLF and DD6 interact with a specific subset of S-RNases, in agreement to the collaborative non-self recognition model, there are reasons to doubt this conclusion as discussed in chapter 4. Further *in vivo* examination of DD6 and a few other DDs is essential to conclude if this finding agrees with other published findings. Using the transcriptome analysis and biochemical study reported here, it is possible to speculate that in *N. alata*, a SCF^{SLF} E3 ligase targets S-RNase for degradation as predicted by the inhibitor model. Based on these results, some future experiment are possible.

Multiple *DDs* are present on the *S* locus but the exact number present in *N. alata* remains unknown. Next generation sequencing technology makes it possible to identify the complete suite of *DDs* present on the *S* locus. However, due to the low coverage of *N. alata* pollen transcriptome, the new *DDs* are identified are not full length and the full suite of *DDs* was not obtained. 5’ and 3’ rapid amplification of cDNA ends (5’ and 3’ RACE) can be performed to obtain full-length DD sequences. Alternatively deeper coverage of the transcriptome and repeat RNA sequencing can be used to identify further *DDs*. In addition, sequencing the genomic region of various *S* background plants would also provide the genomic position of each *DD* on each *S* haplotype. The expression and

genomic location of each *DD* will reveal if the distribution and expression of *DDs* is in agreement with collaborative non-self recognition model. With the identification of most if not all DDs present in each *S* background, phylogenetic analysis of new and already identified DDs would reveal how many types of *DDs* are present and allow predictions to be done for each DDs in which *S* background will each caused the alteration of SI to SC phenotype.

Linkage analysis is necessary to determine which *DDs* are linked to the *S* locus. Other studies have reported that some rosaceous *SFBBs* recombine with the *S-RNase* gene and hence should not be considered as *pollen S* although it is still not known if SFBB interacts with S-RNase (De Franceschi et al., 2011b; Kakui et al., 2011). It would be necessary to find out if an F-box that is not tightly linked to *S* locus also interacts with S-RNase, and if it can be used as a definitive test to be *pollen S*. *DD6* would be a good candidate to do so.

As part of the results obtained from *N. alata* pollen transcriptome analysis, *in vitro* and *in vivo* functional study of SI-related transcripts should be performed to determine if Cullin and SSK are functionally important for RNase-based SI. Transgenic analysis of transmission of T-DNA insertion that knocks down the expression of either Cullin or SSK has been performed by Zhao et al., (2010) and Li and Chetelat et al., (2014) are functional evidence to show that these proteins are essential for RNase-based SI, smilar work could be performed in *N .alata*.

As reported in chapter 4, $(His)_6$:DD6 FBD and $(His)_6$:PiSLF1 FBD interact with the same subset of S-RNases (interact with S_2- and S_7-RNase but not S_3-RNase). It is important to perform *in vivo* experiment for DD6 and PiSLF1 in *N. alata* and investigate which S-RNase each can neutralize *in vivo* based on altered SI to SC behavior in transgenic *N. alata* plants. This experiment will reveal if PiSLF and DD6 belongs to the similar type of SLF and if it supports the collaborative non-self recognition model. The current expectation is that *in vitro* interaction result will agree with the *in vivo* altered SI response in transgenic plant expressing an additional *SLF*.

If *in vitro* interaction assay is to be used in future work as control for other DDs that are not tesed in this study, an RNase not involved in RNase-based SI would be the ideal candidate. NaPI was chosen as negative control as pure recombinant full length RNase NE could not be obtained. Further work should focus on obtaining pure functional SI non-related RNase to ensure folded $(His)_6$:DD6 FBD and $(His)_6$:PiSLF1 FBD do not interact with SI non-related RNase. This may be possible by exploring other eukaryote protein expression system such as yeast (*Pichia pastoris*), tobacco or transiently in *N. benthamiana* leaves. Future interaction assay may also include a reciprocal interaction experiment to make sure any interaction occurs both ways. A SI non-related F-box/kelch protein can be used as

negative control. Another possible control is a construct which consists of only the F-box motif, to show that the COOH terminus is not required for interaction with S-RNase.

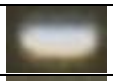
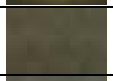
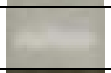
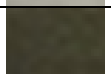
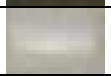
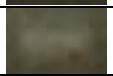
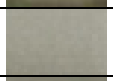
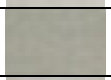
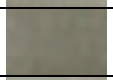
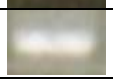
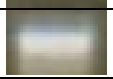
The three dimension protein structure of S-RNase showed that S-RNase has separate catalytic domain (for hydrolysis of RNA) and interacting domain (hypervariable region) which are found in close proximity. The hypervariable region's location on protein surface makes it the region most likely responsible for interacting with specific SLF (Ida et al., 2001; Matsuura et al., 2001). Although SLFs are highly conserved within a class, sequence divergence is detected among the different types of SLF (Kubo et al., 2010, 2015) and it is speculated that recognition specificity may be controlled by the regions that are diverged. The three dimension protein structure of a few types of SLF should reveal if the diverged region is responsible for interaction with S-RNase. It would also be interestingly to study the co-crystallised protein structure between different combinations of S-RNase and SLF. This would show where interaction occurs and reveal how *S* specificity is achieved at protein level. Site-directed mutagenesis targeted at specific amino acids required for protein-protein interaction can be performed to determine which are the critical amino acids required for binding and if swapping amino acids can change a protein's interaction specificity. Combining the molecular, biochemical and protein structural analysis data would provide an important step forward towards better understanding how RNase-based SI systems function.

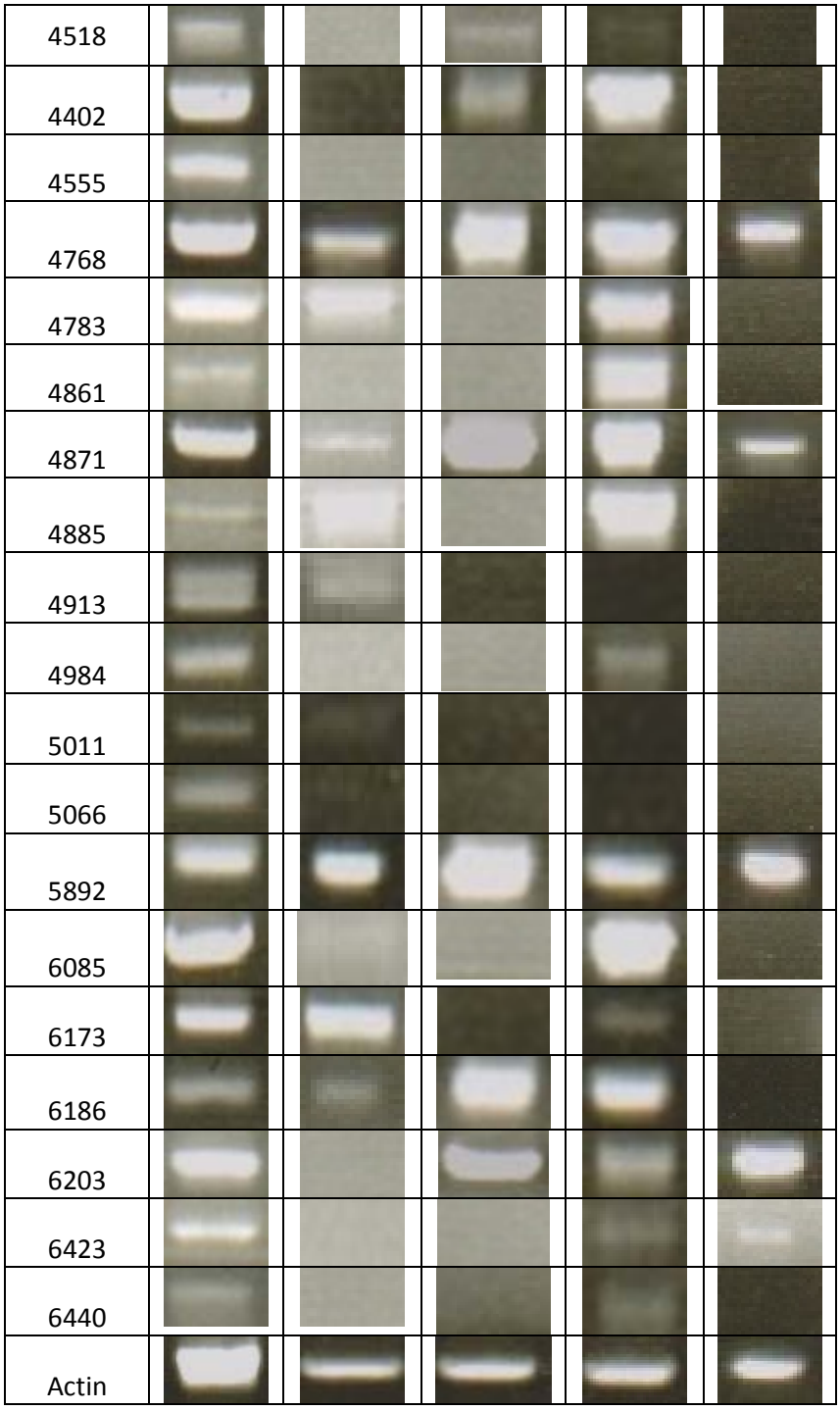
6. Appendices

Appendix I: Transcriptome validation PCR primers used for RT-PCR

Contig	Primer sequence 5′ to 3′		
12	F: TCAGGGACTGCTCGGAGATGGTTT R: TGGTGTTGGCAGTGTTGCTGCT	4885	F: AATGGAGGCTGCACGAAACCCT R: AACTGGGCTAACTGAGTGGCGT
36	F: TGGCTTCGTAGTTGCTGCGCTT R: TCCATGGTCGTCACACGTCGAA	4913	F: AAGAGCGGTTCAGCAGGAATGC R: TGCGCTGCGTTTCTTCCACCAT
593	F: TTGCGCAACTTTTGTGGCACCC R: ACCGGTTTTTGCCGCCCGATTA	4984	F: TTCCCACTAATTAGGCATCACAATGAC R: AAATCATCTTCATCTTCGTCGTCGC
599	F: ATGCTTGCCTGGCTCACCTT R: TGTGGCACGGGTAACCAAACGA	5011	F: GCAAACTCTTCGGTCTGCGGTT R: TCGGCACTGTGTGTGTGTGTGT
615	F: TGAGGACTCCACCATGCGACAA R: TCACGACGACGCGGATAACA	5066	F: TGGTTCAGCAGCATCTTTGGCCT R: ACCTTGTGCTGTTACCATGCAACC
637	F: ATCTCACCGGCGAAAACCGT R: TGCAGCAAATGCAGCAACGCT	5892	F: TCAGCAGGAATGCCGAGCAGAA R: TGCCGAGTGCCACTTTGAACACA
700	F: TTCCCACTAATTAGGCATCACAATGAC R: AAATCATCTTCATCTTCGTCGTCGC	6085	F: TCCAGCAAAATGCCTGGTGCCAA R: TGCACATAACAACCCCGCCA
887	F: TAAGCGAATCGCCAAGGCCGTA R: TGGTGTCTCGGCTTCCAACTCA	6173	F: TGGCATTTCCCCCATTTCACCCT R: TGCAAGCGCCATGTTGATGCT
1026	F: AGCAAGACCCGACTCAAGCTGT R: ACGCGAATCACAGCAGTCACCT	6186	F: AAGAGCGGTTCAGCAGGAATGC R: ATGCCCAAGCATTCTCCCTCCT
1354	F: TGTCACTGCCATGGTTGGTGCT R: TGGGCATTGCGAAAACCTGG	6203	F: TCTTCAGCTGGGGTGATTGCGA R: TTCCCACCATTGCTCGCCGTAA
2401	F: TTGTGGTGTTCTGACGGGCA R: GCTGTCAATTGGCCTGTGGAAGCA	6423	F: TGCCGCCAGCATCAATTCCT R: TTGAACGAGCCAGGGCAGAAGA
2403	F: AACTTCCCGTCTTGAGGCCGTT R: AGGTTTGGTTCTGTCGCCCGTT	4402	F: AGCAGCCTCGCGCAAATTCT R: TCTGCAGTGAGTTTCCTCGGCT
2417	F: AAGGTGGCTGTGTTGCACCTGT R: TCAGACTCGTACTGCACGCACA	4871	F: ACCGGCATGAGCATACGCTATTGTG R: ACCCAAGCAGTGGCTTCTGGAAT
2423	F: TTTTCCAGGTTTCCACGCGCCA R: TCGTGCTTGTAATGCCTCCTGC	6440	F: TTGGCCCATGTTGGTCTTGGCT R: TTTGCTGCGTCTGGCTATCTGCAC
2550	F: AGCAGAGTCGCTTGAACTGCCT R: TGCACACCCCACCATCTGGATT		
2637	F: TGTGGAAGAGCCATTGCTGAGGA R: GCATGGGCTTCTTTCAGTGGCA	F-box sequences	Primer sequence 5′ to 3′
2845	F: AGCATGGGCTTTGCCCACATCA R: AAAACGGCAGTTTCTGGCGG	452	F: CACAACGAAACACCATTTTCCC R: oligo dT $_{(15)}$
2848	F: TGCAACCCTTGCAGCCATTGT R: AGTGTCTGCGACTTGCTGCT	607	F: AGGGGATGAATTGGCCGAACCT R: GCGCTCTGTACCAAAGTCAGCGT R2: CTTTGTCAAATTTCTCTATTCTGTTCC
2904	F: TGCAGATGCCGAGTTTTGCAGC R: TGGCGATCGTCAAACCAGAGA	1067	F: TGGGATGTCTAGCCCCACCGA R: CCGCCTCTTCCGGTATTGATGGA R2: ACAACCCATCACAGCACAATTGTATAC
3116	F: AGGGAAGCTGCCAAGCTGCAAT R: AGTGCTGCCGTGTAAGGAAGGT	3796	F: TTAGATTGCGCGAAACCCAATCGT R: GAGTTGTTCTTTGTGTTGGTGAATCCG
4365	F: TTTGCGCTGGCTGTCAACCT R: TGTGCTTCCACCAATGCAGCCA	4426	F: GCCTGCAAGCGGTGGAGGTC R: AGGCACGCACCAGATGGCTT
4392	F: GGCTGTGGTCATGCCGGGTTTTAT R: ACGCCCATATGTCCATGTCGCA	5325	F: CGGGGAGAGTGCATGTTTGGCA R: TGGGAAGGAAGATGAAGAGGAAAGGAG
4394	F: TTGCTGGGAAGCAGCTTGAGGA R: TCAAAAGCCACCACGGAGACGA	6546	F: TCCCATCTTTCACTGAGCGGAGATCA R: CTTCCTTCCTTCTTTCTTTCTCCGTCC
4398	F: AACTGCGCACGCTGTGTCAA R: TGTGTTCCCTCCTGCCAACCAA	6623	F: CCCGCACCCCGAATCATCACC R: GCAAAGGCAGACTGCTGAGGGT
4422	F: AAATGCAACGGCATCGGGCA R: TGCTTCTTCAGTAACCAGCCATCT	3684 (DD4)	F: ACCATCGTACCATTGAAGGTGTTGGG R: ACTCTCAAACTTTTGGGAAAGCCATGT
4451	F: AACCAAACACACCGCTGCCCAA R: TGCCAATGTACCACCATGCGTCA		
4518	F: TACACACGCCAGACACTCGTCA R: AGCCTGAGAGGGAACATGCACT	E3 ligase associated	Primer sequence 5′ to 3′
4555	F: AGCTCTGTCGGAAACTCACCACCA R: ACTTCCGTGCCCATCACATCCA	3497 (Cullin G)	F: GGTTATTGAGCTGCATGACAAG R: TGGGGAAGCACCTTCCGGCT
4783	F: ACACCAAGTGGCTCCAAGCTGA R: TGGGGACAGGTTCTTTGGTTTCGT	4884 (Cullin C)	F: TGCTGGCTGTTCTAGTGCAGAGC R: TGGGGAAGCACCTTCCGGCT
4768	F: ACATGCATTCGGGGTGCACTT R: TGTGGCACGGGTAACCAAACGA	6029 (RBX1)	F: ACGCTGTCGCTCTTTGGGCT R: TGGGCACACTTGACGGGTTTTG
4861	F: TCTGGGGCTGTACCGTGCAAAT R: TGGGCCCATTGAAGCTCACGTT	SBP1	F: GCGTGAGTTGCAGAGACAAG R: TGGACTGACACAAGGGACAA

Appendix 2: Validation of contigs expression in different plant tissues.

Contig No	Pollen	Leaf	Petal	Style	Seedling
12					
36					
593					
599					
615					
637					
700					
887					
1026					
1354					
2401					
2403					
2417					
2423					
2550					
2637					
2845					
2848					
2904					
3116					
4365					
4392					
4394					
4398					
4422					
4451					


4518
4402
4555
4768
4783
4861
4871
4885
4913
4984
5011
5066
5892
6085
6173
6186
6203
6423
6440
Actin

7. References

Allen, A.M., and Hiscock, S.J. (2008). Self-Incompatibility in Flowering Plants - Evolution, Diversity, and Mechanisms. In Evolution and Phylogeny of Self-Incompatibility Systems in Angiosperms (Springer Berlin Heidelberg).

Allen, A.M., Thorogood, C.J., Hegarty, M.J., Lexer, C., and Hiscock, S.J. (2011). Pollen-pistil interactions and self-incompatibility in the Asteraceae: new insights from studies of *Senecio squalidus* (Oxford ragwort). Ann Bot **108,** 687-698.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25,** 3389-3402.

Anderson, M.A., Cornish, E.C., Mau, S.-L., Williams, E.G., Hoggart, R., Atkinson, A., Bonig, I., Grego, B., Simpson, R., Roche, P.J., Haley, J.D., Penschow, J.D., Niall, H.D., Tregear, G.W., Coghlan, J.P., Crawford, R.J., and Clarke, A.E. (1986). Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata* **321,** 38-44.

Anderson, M.A., McFadden, G.I., Bernatzky, R., Atkinson, A., Orpin, T., Dedman, H., Tregear, G., Fernley, R., and Clarke, A.E. (1989). Sequence variability of three alleles of the self-incompatibility gene of *Nicotiana alata*. Plant Cell **1,** 483-491.

Andrade, M.A., Gonzalez-Guzman, M., Serrano, R., and Rodriguez, P.L. (2001). A combination of the F-box motif and kelch repeats defines a large *Arabidopsis* family of F-box proteins. Plant Mol Biol **46,** 603-614.

Aoki, K., Yano, K., Suzuki, A., Kawamura, S., Sakurai, N., Suda, K., Kurabayashi, A., Suzuki, T., Tsugane, T., Watanabe, M., Ooga, K., Torii, M., Narita, T., Shin, I.T., Kohara, Y., Yamamoto, N., Takahashi, H., Watanabe, Y., Egusa, M., Kodama, M., Ichinose, Y., Kikuchi, M., Fukushima, S., Okabe, A., Arie, T., Sato, Y., Yazawa, K., Satoh, S., Omura, T., Ezura, H., and Shibata, D. (2010). Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. BMC Genomics **11,** 210.

Arango, M., Gevaudant, F., Oufattole, M., and Boutry, M. (2003). The plasma membrane proton pump ATPase: the significance of gene subfamilies. Planta **216,** 355-365.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics **22,** 195-201.

Atkinson, A.H., Heath, R.L., Simpson, R.J., Clarke, A.E., and Anderson, M.A. (1993). Proteinase inhibitors in *Nicotiana alata* stigmas are derived from a precursor protein which is processed into five homologous inhibitors. Plant Cell **5,** 203-213.

Baker, H.G. (1955). Self-Compatibility and Establishment After 'Long-Distance' Dispersal. Evolution **9,** 347-349.

Baneyx, F., and Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. Nat Biotech **22,** 1399-1408.

Barrett, S.C.H. (2002). The evolution of plant sexual diversity. Nat Rev Genet **3,** 274-284.

Beardsell, D., Knox, R., and Williams, E. (1993). Breeding System and Reproductive Success of *Thryptomene calycina* (Myrtaceae). Australian Journal of Botany **41,** 333-353.

Bell, P.R. (1995). Incompatibility in Flowering Plants: Adaptation of an Ancient Response. Plant Cell **7,** 5-16.

Bennett, M.J., and Eisenberg, D. (2004). The Evolving Role of 3D Domain Swapping in Proteins. Structure **12,** 1339-1341.

Bennett, M.J., Sawaya, M.R., and Eisenberg, D. (2006). Deposition Diseases and 3D Domain Swapping. Structure **14,** 811-824.

Bennett, M.J., Schlunegger, M.P., and Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. Protein Sci **4,** 2455-2468.

Boothby, T.C., Zipper, R.S., van der Weele, C.M., and Wolniak, S.M. (2013). Removal of retained introns regulates translation in the rapidly developing gametophyte of *Marsilea vestita*. Dev Cell **24,** 517-529.

Boutry, M., Michelet, B., and Goffeau, A. (1989). Molecular cloning of a family of plant genes encoding a protein homologous to plasma membrane H+-translocating ATPases. Biochem Biophys Res Commun **162,** 567-574.

Bredemeijer, G.M.M., and Blaas, J. (1981). S-specific proteins in styles of self-incompatible *Nicotiana alata*. TAG Theoretical and Applied Genetics **59,** 185-190.

Brohawn, S.G., Leksa, N.C., Spear, E.D., Rajashankar, K.R., and Schwartz, T.U. (2008). Structural evidence for common ancestry of the nuclear pore complex and vesicle coats. Science **322,** 1369-1373.

Burgess, R.R. (2009). Chapter 17 Refolding Solubilized Inclusion Body Proteins. In Methods in Enzymology, R.B. Richard and P.D. Murray, eds (Academic Press), Volume **463**, pp. 259-282

Cahais, V., Gayral, P., Tsagkogeorga, G., Melo-Ferreira, J., Ballenghien, M., Weinert, L., Chiari, Y., Belkhir, K., Ranwez, V., and Galtier, N. (2012). Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. Mol Ecol Resour **12,** 834-845.

Chen, C.K., Chan, N.L., and Wang, A.H. (2011). The many blades of the beta-propeller proteins: conserved but versatile. Trends Biochem Sci **36,** 553-561.

Chen, G., Zhang, B., Liu, L., Li, Q., Zhang, Y.e., Xie, Q., and Xue, Y. (2012). Identification of a Ubiquitin-Binding Structure in the S Locus F-Box Protein Controlling S-RNase-Based Self-Incompatibility. Journal of Genetics and Genomics **39,** 93-102.

Cheng, J., Han, Z., Xu, X., and Li, T. (2006). Isolation and identification of the pollen-expressed polymorphic F-box genes linked to the S locus in apple (*Malus* × *domestica*). Sexual Plant Reproduction **19,** 175-183.

Chow, M.K., Amin, A.A., Fulton, K.F., Whisstock, J.C., Buckle, A.M., and Bottomley, S.P. (2006a). REFOLD: an analytical database of protein refolding methods. Protein Expr Purif **46,** 166-171.

Chow, M.K., Amin, A.A., Fulton, K.F., Fernando, T., Kamau, L., Batty, C., Louca, M., Ho, S., Whisstock, J.C., Bottomley, S.P., and Buckle, A.M. (2006b). The REFOLD database: a tool for the optimization of protein expression and refolding. Nucleic Acids Res **34,** D207-212.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res **36,** W197-201.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21,** 3674-3676.

Coutard, B., Danchin, E.G., Oubelaid, R., Canard, B., and Bignon, C. (2012). Single pH buffer refolding screen for protein from inclusion bodies. Protein Expr Purif **82,** 352-359.

De Franceschi, P., Dondini, L., and Sanzol, J. (2012). Molecular bases and evolutionary dynamics of self-incompatibility in the *Pyrinae* (Rosaceae). J Exp Bot **63,** 4015-4032.

De Franceschi, P., Pierantoni, L., Dondini, L., Grandi, M., Sanzol, J., and Sansavini, S. (2011a). Cloning and mapping multiple S locus F-box genes in European pear (*Pyrus communis L.*). Tree Genetics & Genomes **7,** 231-240.

De Franceschi, P., Pierantoni, L., Dondini, L., Grandi, M., Sansavini, S., and Sanzol, J. (2011b). Evaluation of candidate F-box genes for the pollen S of gametophytic self-incompatibility in the *Pyrinae* (Rosaceae) on the basis of their phylogenomic context. Tree Genetics & Genomes *7*, 663-683.

Deshpande, R.A., and Shankar, V. (2002). Ribonucleases from T2 family. Crit Rev Microbiol **28,** 79-122.

De Smet, I., Voss, U., Jurgens, G., and Beeckman, T. (2009). Receptor-like kinases shape the plant. Nat Cell Biol **11,** 1166-1173.

Dodds, P., Clarke, A., and Newbigin, E. (1996). Molecular characterisation of an S-like RNase of *Nicotiana alata* that is induced by phosphate starvation. Plant Molecular Biology **31,** 227-238.

East, E.M., and Mangelsdorf, A.J. (1925). A New Interpretation of the Hereditary Behavior of Self-Sterile Plants. PNAS **11,** 166-171.

Edwards, K.D., Bombarely, A., Story, G.W., Allen, F., Mueller, L.A., Coates, S.A., and Jones, L. (2010). TobEA: an atlas of tobacco gene expression from seed to senescence. BMC Genomics **11,** 142.

Entani, T., Kubo, K., Isogai, S., Fukao, Y., Shirakawa, M., Isogai, A., and Takayama, S. (2014). Ubiquitin-proteasome-mediated degradation of S-RNase in a solanaceous cross-compatibility reaction. Plant J **78,** 1014-1021.

Ernst, A. (1955). Self-fertility in monomorphic *Primulas*. Genetica **27,** 391-448.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.K., and Mockler, T.C. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome Res **20,** 45-58.

Fletcher, S., Bowden, S.E.H., and Marrion, N.V. (2003). False interaction of syntaxin 1A with a Ca2+-activated K+ channel revealed by co-immunoprecipitation and pull-down assays: implications for identification of protein–protein interactions. Neuropharmacology **44,** 817-827.

Ford, C.S., and Wilkinson, M.J. (2012). Confocal observations of late-acting self-incompatibility in *Theobroma cacao L*. Sex Plant Reprod *25,* 169-183.

Foote, H.C., Ride, J.P., Franklin-Tong, V.E., Walker, E.A., Lawrence, M.J., and Franklin, F.C. (1994). Cloning and expression of a distinctive class of self-incompatibility (S) gene from *Papaver rhoeas L*. Proc Natl Acad Sci U S A **91,** 2265-2269.

Fowler, D.M., Koulov, A.V., Balch, W.E., and Kelly, J.W. (2007). Functional amyloid--from bacteria to humans. Trends Biochem Sci **32,** 217-224.

Franklin-Tong, V.E., and Gourlay, C.W. (2008). A role for actin in regulating apoptosis/programmed cell death: evidence spanning yeast, plants and animals. Biochem J **413,** 389-404.

Gagne, J.M., Downes, B.P., Shiu, S.H., Durski, A.M., and Vierstra, R.D. (2002). The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. Proc Natl Acad Sci USA **99,** 11519-11524. Epub 12002 Aug 11518.

Georgiou, G., and Valax, P. (1996). Expression of correctly folded proteins in *Escherichia coli*. Current Opinion in Biotechnology **7,** 190-197.

Goldraij, A., Kondo, K., Lee, C.B., Hancock, C.N., Sivaguru, M., Vazquez-Santana, S., Kim, S., Phillips, T.E., Cruz-Garcia, F., and McClure, B. (2006). Compartmentalization of S-RNase and HT-B degradation in self-incompatible *Nicotiana*. Nature **439,** 805-810.

Golz, J.F., Clarke, A.E., and Newbigin, E. (2000). Mutational Approaches to the Study of Self-incompatibility: Revisiting the Pollen-part Mutants. Ann Bot **85,** 95-103.

Golz, J.F., Clarke, A.E., Newbigin, E., and Anderson, M. (1998). A relic S-RNase is expressed in the styles of self-compatible *Nicotiana sylvestris*. Plant Journal **16**.

Golz, J.F., Oh, H.Y., Su, V., Kusaba, M., and Newbigin, E. (2001). Genetic analysis of *Nicotiana* pollen-part mutants is consistent with the presence of an S-ribonuclease inhibitor at the S locus. Proc Natl Acad Sci U S A **98,** 15372-15376.

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

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol **29,** 644-652.

Gray, H.B. (2003). Biological inorganic chemistry at the beginning of the 21st century. Proc Natl Acad Sci U S A **100,** 3563-3568.

Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis **18,** 2714-2723.

Haiser, H.J., Karginov, F.V., Hannon, G.J., and Elliot, M.A. (2008). Developmentally regulated cleavage of tRNAs in the bacterium *Streptomyces coelicolor*. Nucleic Acids Research **36,** 732-741.

Haud, N., Kara, F., Diekmann, S., Henneke, M., Willer, J.R., Hillwig, M.S., Gregg, R.G., MacIntosh, G.C., Gärtner, J., Alia, A., and Hurlstone, A.F.L. (2011). rnaset2 mutant zebrafish model familial cystic leukoencephalopathy and reveal a role for RNase T2 in degrading ribosomal RNA. Proceedings of the National Academy of Sciences **108,** 1099-1103.

Hebsgaard, S.M. Korning, P.G. Tolstrup, Engelbrecht, N. J. Rouze, Brunak P. S. (1996) Splice site prediction in *Arabidopsis thaliana* DNA by combining local and global sequence information. Nucleic Acids Research 24, 3439-3452.

Heslop-Harrison, J. (1975). Incompatibility and the Pollen-Stigma Interaction. Annual Review of Plant Physiology **26,** 403-425.

Hildebrand, P.W., Goede, A., Bauer, R.A., Gruening, B., Ismer, J., Michalsky, E., and Preissner, R. (2009). SuperLooper--a prediction server for the modeling of loops in globular and membrane proteins. Nucleic Acids Res **37,** W571-574.

Hillwig, M.S., Contento, A.L., Meyer, A., Ebany, D., Bassham, D.C., and Macintosh, G.C. (2011). RNS2, a conserved member of the RNase T2 family, is necessary for ribosomal RNA decay in plants. Proc Natl Acad Sci U S A **108,** 1093-1098.

Hiscock, S.J., and Tabah, D.A. (2003). The different mechanisms of sporophytic self-incompatibility. Philos Trans R Soc Lond B Biol Sci **358,** 1037-1045.

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. Plant Cell **18,** 2531-2553.

Hua, Z., and Kao, T.-H. (2008). Identification of major lysine residues of S_3-RNase of *Petunia inflata* involved in ubiquitin-26S proteasome-mediated degradation in vitro. The Plant Journal: For Cell And Molecular Biology **54,** 1094-1104.

Hua, Z., Meng, X., and Kao, T.-h. (2007). Comparison of *Petunia inflata* S Locus F-Box Protein (Pi SLF) with PiSLF Like Proteins Reveals Its Unique Function in S-RNase Based Self-Incompatibility. Plant Cell **19,** 3593-3609.

Huang, J., Zhao, L., Yang, Q., and Xue, Y. (2006). AhSSK1, a novel SKP1-like protein that interacts with the S locus F-box protein SLF. The Plant Journal: For Cell And Molecular Biology **46,** 780-793.

Huang, S., Lee, H.S., Karunanandaa, B., and Kao, T.H. (1994). Ribonuclease activity of *Petunia inflata* S proteins is essential for rejection of self-pollen. Plant Cell **6,** 1021-1028.

Huang, X., and Madan, A. (1999). CAP3: A DNA Sequence Assembly Program. Genome Research **9,** 868-877.

Hudson, M., and Cooley, L. (2008). "The Coronin Family of Proteins". (Landes Bioscience and Springer Science).

Ida, K., Norioka, S., Yamamoto, M., Kumasaka, T., Yamashita, E., Newbigin, E., Clarke, A.E., Sakiyama, F., and Sato, M. (2001). The 1.55 Å resolution structure of *Nicotiana alata* SF11-RNase associated with gametophytic self-incompatibility. Journal of Molecular Biology **314,** 103-112.

Igic, B., and Kohn, J.R. (2006). The distribution of plant mating systems: study bias against obligately outcrossing species. Evolution **60**, 1098-1103.

Ikeda, K., Igic, B., Ushijima, K., Yamane, H., Hauck, N.R., Nakano, R., Sassa, H., Iezzoni, A.F., Kohn, J.R., and Tao, R. (2004). Primary structural features of the S haplotype-specific F-box protein, SFB, in *Prunus*. Sexual Plant Reproduction **16,** 235-243.

Ioerger, T.R., Clark, A.G., and Kao, T.H. (1990). Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc Natl Acad Sci U S A **87,** 9732-9735.

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

Ito, N., Phillips, S.E., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D., and Knowles, P.F. (1991). Novel thioether bond revealed by a 1.7 A crystal structure of galactose oxidase. Nature **350,** 87-90.

Iwano, M., and Takayama, S. (2012). Self/non-self discrimination in angiosperm self-incompatibility. Current Opinion in Plant Biology **15,** 78-83.

Jahnen, W., Batterham, M.P., Clarke, A.E., Moritz, R.L., and Simpson, R.J. (1989). Identification, isolation, and N-terminal sequencing of style glycoproteins associated with self-incompatibility in *Nicotiana alata*. The Plant Cell Online **1,** 493-499.

Kane, J.F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Current Opinion in Biotechnology **6,** 494-500.

Kakui, H., Tsuzuki, T., Koba, T., and Sassa, H. (2007). Polymorphism of SFBB-gamma and its use for S genotyping in Japanese pear (*Pyrus pyrifolia*). Plant Cell Rep **26,** 1619-1625.

Kakui, H., Kato, M., Ushijima, K., Kitaguchi, M., Kato, S., and Sassa, H. (2011). Sequence divergence and loss-of-function phenotypes of S locus F-box brothers genes are consistent with non-self recognition by multiple pollen determinants in self-incompatibility of Japanese pear (*Pyrus pyrifolia*). The Plant Journal **68,** 1028-1038.

Kelly, S.M., Jess, T.J., and Price, N.C. (2005). How to study proteins by circular dichroism. Biochim Biophys Acta **1751,** 119-139.

Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. Nat Rev Mol Cell Biol **13,** 700-712.

Kessler, S.A., and Grossniklaus, U. (2011). She's the boss: signaling in pollen tube reception. Curr Opin Plant Biol **14,** 622-627.

Klaas, M., Yang, B., Bosch, M., Thorogood, D., Manzanares, C., Armstead, I.P., Franklin, F.C., and Barth, S. (2011). Progress towards elucidating the mechanisms of self-incompatibility in the grasses: further insights from studies in *Lolium*. Ann Bot **108,** 677-685.

Kowyama, Y., Kunz, C., Lewis, I., Newbigin, E., Clarke, A.E., and Anderson, M.A. (1994). Self-compatibility in a *Lycopersicon peruvianum* variant (LA2157) is associated with a lack of style S-RNase activity. TAG Theoretical and Applied Genetics **88,** 859-864.

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.

Lai, Z., Ma, W., Han, B., Liang, L., Zhang, Y., Hong, G., and Xue, Y. (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.

Lampugnani, E.R., Moller, I.E., Cassin, A., Jones, D.F., Koh, P.L., Ratnayake, S., Beahan, C.T., Wilson, S.M., Bacic, A., and Newbigin, E. (2013). In Vitro Grown Pollen Tubes of *Nicotiana alata* Actively Synthesise a Fucosylated Xyloglucan. PLoS One **8,** e77140.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol **10,** R25.

Lawrence, M.J. (1975). The Genetics of Self-Incompatibility in *Papaver rhoeas*. Proceedings of the Royal Society of London. Series B. Biological Sciences **188,** 275-285.

Lee, C.B., Swatek, K.N., and McClure, B. (2008). Pollen Proteins Bind to the C-terminal Domain of *Nicotiana alata* Pistil Arabinogalactan Proteins. Journal of Biological Chemistry **283,** 26965-26973.

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

Lee, S.R., and Collins, K. (2005). Starvation-induced Cleavage of the tRNA Anticodon Loop in *Tetrahymena thermophila*. Journal of Biological Chemistry **280,** 42744-42749.

Lewis, D. (1960). Genetic Control of Specificity and Activity of the S Antigen in Plants. Proceedings of the Royal Society of London. Series B. Biological Sciences **151,** 468-477.

Lewis, D., and Crowe, L.K. (1958). Unilateral interspecific incompatibility in flowering plants. Heredity **12,** 233-256.

Lewis, D. (1949). Structure of the incompatibility gene; induced mutation rate. Heredity (Edinb) **3,** 339-355.

Lewis, D., and Crowe, L.K. (1954). Structure of the incompatibility gene. Heredity **8,** 357-363.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12,** 323-323.

Li, J., Webster, M.A., Smith, M.C., and Gilmartin, P.M. (2011). Floral heteromorphy in *Primula vulgaris*: progress towards isolation and characterization of the S locus. Ann Bot **108,** 715-726.

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

Li, W., and Chetelat, R.T. (2010). A Pollen Factor Linking Inter- and Intraspecific Pollen Rejection in Tomato. Science **330,** 1827-1830.

Li, W., and Chetelat, R.T. (2014). The role of a pollen-expressed Cullin1 protein in gametophytic self-incompatibility in *Solanum*. Genetics **196,** 439-442.

Li, W., and Chetelat, R.T. (2015). Unilateral incompatibility gene ui1.1 encodes an S locus F-box protein expressed in pollen of *Solanum* species. Proc Natl Acad Sci U S A **112,** 4417-4422.

Li, X., Zhang, D., Hannink, M., and Beamer, L.J. (2004). Crystal Structure of the Kelch Domain of Human Keap1. Journal of Biological Chemistry **279,** 54750-54758.

Lilie, H., Schwarz, E., and Rudolph, R. (1998). Advances in refolding of proteins produced in *E. coli*. Curr Opin Biotechnol **9,** 497-501.

Liu, Y.H. (1993) Structural and Functional relationship of S-RNases. PhD Thesis School of Botany, University of Melbourne.

Liu, Y., and Eisenberg, D. (2002). 3D domain swapping: As domains continue to swap. Protein Science **11,** 1285-1299.

Liu, Z.-q., Xu, G.-h., and Zhang, S.-l. (2007). *Pyrus pyrifolia* stylar S-RNase induces alterations in the actin cytoskeleton in self-pollen and tubes in vitro. Protoplasma **232,** 61-67.

Lo, S.C., Li, X., Henzl, M.T., Beamer, L.J., and Hannink, M. (2006). Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling. EMBO J **25,** 3605-3617.

Luhtala, N., and Parker, R. (2010). T2 Family ribonucleases: ancient enzymes with diverse roles. Trends Biochem Sci **35,** 253-259.

Lundquist, A. (1965). The genetics of incompatibility. In:Proc. llth Int. Congr. Genet. **3**, 637-647.

Lush, W.M., and Clarke, A.E. (1997). Observations of pollen tube growth in *Nicotiana alata* and their implications for the mechanism of self-incompatibility. Sexual Plant Reproduction **10,** 27-35.

Luu, D.T., Qin, X., Morse, D., and Cappadocia, M. (2000). S-RNase uptake by compatible pollen tubes in gametophytic self-incompatibility. Nature **407,** 649-651.

Mable, B.K., Schierup, M.H., and Charlesworth, D. (2003). Estimating the number, frequency, and dominance of S alleles in a natural population of *Arabidopsis lyrata* (Brassicaceae) with sporophytic control of self-incompatibility. Heredity (Edinb) **90,** 422-431.

MacIntosh, G.C., Bariola, P.A., Newbigin, E., and Green, P.J. (2001). Characterization of Rny1, the *Saccharomyces cerevisiae* member of the T2 RNase family of RNases: unexpected functions for ancient enzymes? Proc Natl Acad Sci U S A **98,** 1018-1023.

Mackay, J.P., Sunde, M., Lowry, J.A., Crossley, M., and Matthews, J.M. (2007). Protein interactions: is seeing believing? Trends in Biochemical Sciences **32,** 530-531.

Marquez, Y., Brown, J.W., Simpson, C., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. Genome Res **22,** 1184-1195.

Mascarenhas, N.T., Bashe, D., Eisenberg, A., Willing, R.P., Xiao, C.M., and Mascarenhas, J.P. (1984). Messenger RNAs in corn pollen and protein synthesis during germination and pollen tube growth. Theor Appl Genet **68,** 323-326.

Matsubayashi, Y., and Sakagami, Y. (2006). Peptide hormones in plants. Annu Rev Plant Biol **57,** 649-674.

Matsuura, T., Sakai, H., Unno, M., Ida, K., Sato, M., Sakiyama, F., and Norioka, S. (2001). Crystal Structure at 1.5-Å Resolution of *Pyrus pyrifolia* Pistil Ribonuclease Responsible for Gametophytic Self-incompatibility. Journal of Biological Chemistry **276,** 45261-45269.

Matsumoto, D., Yamane, H., Abe, K., and Tao, R. (2012). Identification of a Skp1-like protein interacting with SFB, the pollen S determinant of the gametophytic self-incompatibility in *Prunus*. Plant Physiol **159,** 1252-1262.

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. Plant Cell **9,** 1757-1766.

McClure, B. (2004). S-RNase and SLF determine S haplotype-specific pollen recognition and rejection. Plant Cell **16,** 2840-2847.

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

McClure, B.A., Gray, J.E., Anderson, M.A., and Clarke, A.E. (1990). Self-incompatibility in *Nicotiana alata* involves degradation of pollen rRNA. Nature **347,** 757-760.

McClure, B.A., Haring, V., Ebert, P.R., Anderson, M.A., Simpson, R.J., Sakiyama, F., and Clarke, A.E. (1989). Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. Nature **342,** 955-957.

McCubbin, A. (2008). Self-Incompatibility in Flowering Plants. Heteromorphic Self-Incompatibility in *Primula*: Twenty-First Century Tools Promise to Unrevel a Classic Nineteeth Century Model System, 289-308.

McCubbin, A.G., Chung, Y.Y., and Kao, T. (1997). A Mutant S_3-RNase of *Petunia inflata* Lacking RNase Activity Has an Allele-Specific Dominant Negative Effect on Self-Incompatibility Interactions. Plant Cell **9,** 85-95.

Middelberg, A.P.J. (2002). Preparative protein refolding. Trends in Biotechnology **20,** 437-443.

Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., and Marshall, D. (2010). Tablet—next generation sequence assembly visualization. Bioinformatics **26,** 401-402.

Minamikawa, M., Kakui, H., Wang, S., Kotoda, N., Kikuchi, S., Koba, T., and Sassa, H. (2010). Apple S locus region represents a large cluster of related, polymorphic and pollen-specific F-box genes. Plant Mol Biol **74,** 143-154.

Mu, J. and Kao, T.-h. (1992). Expression of two S-ribonucleases of *Petunia inflata* using baculovirus expression system. Biochemical and Biophysical Research Communications **187,** 299-304.

Murfett, J., Strabala, T.J., Zurek, D.M., Mou, B., Beecher, B., and McClure, B.A. (1996). S-RNase and Interspecific Pollen Rejection in the Genus *Nicotiana*: Multiple Pollen-Rejection Pathways Contribute to Unilateral Incompatibility between Self-Incompatible and Self-Compatible Species. Plant Cell **8,** 943-958.

Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riekel, C., Grothe, R., and Eisenberg, D. (2005). Structure of the cross-beta spine of amyloid-like fibrils. Nature **435,** 773-778.

Newbigin, E., Anderson, M.A., and Clarke, A.E. (1993). Gametophytic Self-Incompatibility Systems. Plant Cell **5,** 1315-1324.

Newbigin, E., Paape, T., and Kohn, J.R. (2008). RNase-Based Self-Incompatibility: Puzzled by Pollen S. Plant Cell **20,** 2286-2292.

O'Brien, M., Major, G., Chantha, S.C., and Matton, D.P. (2004). Isolation of S-RNase binding proteins from *Solanum chacoense*: identification of an SBP1 (RING finger protein) orthologue. Sexual Plant Reproduction *17*, 81-87.

Okada, K., Tonaka, N., Taguchi, T., Ichikawa, T., Sawamura, Y., Nakanishi, T., and Takasaki-Yasuda, T. (2011). Related polymorphic F-box protein genes between haplotypes clustering in the BAC contig sequences around the S-RNase of Japanese pear. J Exp Bot **62,** 1887-1902.

Okada, K., Tonaka, N., Moriya, Y., Norioka, N., Sawamura, Y., Matsumoto, T., Nakanishi, T., and Takasaki-Yasuda, T. (2008). Deletion of a 236 kb region around S_4-RNase in a stylar-part mutant S_4^{sm} haplotype of Japanese pear. Plant Mol Biol **66,** 389-400.

Olmstead, R.G., Bohs, L., Migid, H.A., Santiago-Valentin, E., Garcia, V.F., and Collier, S.M. (2008). A molecular phylogeny of the Solanaceae. Taxon **57,** 1159-1181.

Oxley, D., and Bacic, A. (1995). Microheterogeneity of N-glycosylation on a stylar self-incompatibility glycoprotein of *Nicotiana alata*. Glycobiology **5,** 517-523.

Oxley, D., Munro, S.L., Craik, D.J., and Bacic, A. (1996). Structure of N-glycans on the S_3- and S_6-allele stylar self-incompatibility ribonucleases of *Nicotiana alata*. Glycobiology **6,** 611-618.

Paape, T., Igic, B., Smith, S.D., Olmstead, R., Bohs, L., and Kohn, J.R. (2008). A 15-Myr-old genetic bottleneck. Mol Biol Evol **25,** 655-663.

Padmanabhan, B., Nakamura, Y., and Yokoyama, S. (2008). Structural analysis of the complex of Keap1 with a prothymosin alpha peptide. Acta Crystallogr Sect F Struct Biol Cryst Commun **64,** 233-238.

Padmanabhan, B., Tong, K.I., Ohta, T., Nakamura, Y., Scharlock, M., Ohtsuji, M., Kang, M.I., Kobayashi, A., Yokoyama, S., and Yamamoto, M. (2006). Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. Mol Cell **21,** 689-700.

Perez, C., Michelet, B., Ferrant, V., Bogaerts, P., and Boutry, M. (1992). Differential expression within a three-gene subfamily encoding a plasma membrane H(+)-ATPase in *Nicotiana plumbaginifolia*. J Biol Chem **267,** 1204-1211.

Petroski, M.D., and Deshaies, R.J. (2005). Function and regulation of cullin-RING ubiquitin ligases. Nat Rev Mol Cell Biol **6,** 9-20.

Pham, C.L., Kwan, A.H., and Sunde, M. (2014). Functional amyloid: widespread in Nature, diverse in purpose. Essays Biochem **56,** 207-219.

Qiao, H., Wang, H., Zhao, L., Zhou, J., Huang, J., Zhang, Y., and Xue, Y. (2004). 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.

Qiao, H., Wang, F., Zhao, L., Zhou, J., Lai, Z., Zhang, Y., 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. Epub 2004 Aug 2312.

Qin, Y., Leydon, A.R., Manziello, A., Pandey, R., Mount, D., Denic, S., Vasic, B., Johnson, M.A., and Palanivelu, R. (2009). Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. PLoS Genet **5,** e1000621.

Rahman, M.H., Uchiyama, M., Kuno, M., Hirashima, N., Suwabe, K., Tsuchiya, T., Kagaya, Y., Kobayashi, I., Kakeda, K., and Kowyama, Y. (2007). Expression of stigma- and anther-specific genes located in the S locus region of *Ipomoea trifida*. Sexual Plant Reproduction **20,** 73-85.

Read, S.M., Newbigin, E., Clarke, A.E., McClure, B.A., and Kao, T. (1995). Disputed Ancestry: Comments on a Model for the Origin of Incompatibility in Flowering Plants. Plant Cell **7,** 661-664.

Reddy, A.S., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. Plant Cell **25,** 3657-3683.

Ride, J.P., Davies, E.M., Franklin, F.C., and Marshall, D.F. (1999). Analysis of *Arabidopsis* genome sequence reveals a large new gene family in plants. Plant Mol Biol **39,** 927-932.

Risk, J.M., Day, C.L., and Macknight, R.C. (2009). Reevaluation of Abscisic Acid-Binding Assays Shows That G-Protein-Coupled Receptor2 Does Not Bind Abscisic Acid. Plant Physiology **150,** 6-11.

Roldan, J.A., Rojas, H.J., and Goldraij, A. (2012). Disorganization of F-actin cytoskeleton precedes vacuolar disruption in pollen tubes during the in vivo self-incompatibility response in *Nicotiana alata*. Ann Bot **110,** 787-795.

Sage, T.L., Bertin, R.I., and William, E.G. (1994). Ovarian and other late-acting self-incompatibility systems. In Genetic control of self-incompatibility and reproductive development in flowering plants, G.W. Elizabeth, E.C. Adrienne, and K. Bruce, eds (Dordrecht: Kluwer Academic), pp. 116-140.

Saito, T., Sato, Y., Sawamura, Y., Shoda, M., Takasaki-Yasuda, T., and Kotobuki, K. (2012). Dual recognition of S_1 and S_4 pistils by S_4^{sm} pollen in self-incompatibility of Japanese pear (*Pyrus pyrifolia Nakai*). Tree Genetics & Genomes *8*, 689-694.

Sambrook, J. and Russell D.W. (2001). Molecular cloning. A laboratary manual. Third edition, Cold Spring Habour Laboratory Press. USA.

Samuel, M.A., Chong, Y.T., Haasen, K.E., Aldea-Brydges, M.G., Stone, S.L., and Goring, D.R. (2009). Cellular pathways regulating responses to compatible and self-incompatible pollen in *Brassica* and *Arabidopsis* stigmas intersect at Exo70A1, a putative component of the exocyst complex. Plant Cell **21,** 2655-2671.

Sassa, H., Hirano, H., and Ikehashi, H. (1992). Self-Incompatibility-Related RNases in Styles of Japanese Pear (*Pyrus serotina Rehd*.). Plant Cell Physiol. **33,** 811-814.

Sassa, H., Hirano, H., and Ikehashi, H. (1993). Identification and characterization of stylar glycoproteins associated with self-incompatibility genes of Japanese pear, *Pyrus serotina Rehd*. Mol Gen Genet **241,** 17-25.

Sassa, H., Hirano, H., Nishio, T., and Koba, T. (1997). Style-specific self-compatible mutation caused by deletion of the S-RNase gene in Japanese pear (*Pyrus serotina*). The Plant Journal **12,** 223-227.

Sassa, H., Kakui, H., Miyamoto, M., Suzuki, Y., Hanada, T., Ushijima, K., Kusaba, M., Hirano, H., and Koba, T. (2007). S locus F-box brothers: multiple and pollen-specific F-box genes with S haplotype-specific polymorphisms in apple and Japanese pear. Genetics **175,** 1869-1881.

Sassa, H., Nishio, T., Kowyama, Y., Hirano, H., Koba, T., and Ikehashi, H. (1996). Self-incompatibility S alleles of the Rosaceae encode members of a distinct class of the T2/S ribonuclease superfamily. Mol Gen Genet **250,** 547-557.

Schuck, P., Perugini, M.A., Gonzales, N.R., Howlett, G.J., and Schubert, D. (2002). Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. Biophys J **82,** 1096-1111.

Schuck, P., and Rossmanith, P. (2000). Determination of the sedimentation coefficient distribution by least-squares boundary modeling. Biopolymers **54,** 328-341.

Schopfer, C.R., Nasrallah, M.E., and Nasrallah, J.B. (1999). The male determinant of self-incompatibility in *Brassica*. Science **286,** 1697-1700.

Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-Seq assembly across the dynamic range of expression levels. Bioinformatics **28,** 1086-1092.

Sedlak, E., Ziegler, L., Kosman, D.J., and Wittung-Stafshede, P. (2008). In vitro unfolding of yeast multicopper oxidase Fet3p variants reveals unique role of each metal site. Proc Natl Acad Sci U S A **105,** 19258-19263.

Seavey SR, Bawa KS. 1986. Late-acting self-incompatibility in Angiosperms. Botanical Review **52**, 195-219.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7,** 539.

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

Sivakumar, N., Li, N., Tang, J.W., Patel, B.K., and Swaminathan, K. (2006). Crystal structure of AmyA lacks acidic surface and provide insights into protein stability at poly-extreme condition. FEBS Lett **580,** 2646-2652.

Skaar, J.R., Pagan, J.K., and Pagano, M. (2013). Mechanisms and function of substrate recruitment by F-box proteins. Nat Rev Mol Cell Biol **14,** 369-381.

Sonneveld, T., Tobutt, K.R., Vaughan, S.P., and Robbins, T.P. (2005). Loss of pollen S function in two self-compatible selections of *Prunus avium* is associated with deletion/mutation of an S-haplotype-specific F-box gene. Plant Cell **17,** 37-51.

Sorensen, H.P., and Mortensen, K.K. (2005). Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J Biotechnol **115,** 113-128.

Sreerama, N., Venyaminov, S.Y., and Woody, R.W. (2000). Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. Anal Biochem **287,** 243-251.

Staiger, D. (2015). Shaping the *Arabidopsis* Transcriptome through Alternative Splicing. Advances in Botany **2015,** 13.

Stein, J.C., Howlett, B., Boyes, D.C., Nasrallah, M.E., and Nasrallah, J.B. (1991). Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. Proc Natl Acad Sci U S A **88,** 8816-8820.

Stone, S.L., Anderson, E.M., Mullen, R.T., and Goring, D.R. (2003). ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible *Brassica* pollen. Plant Cell **15,** 885-898.

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

Takayama, S., Shimosato, H., Shiba, H., Funato, M., Che, F.S., Watanabe, M., Iwano, M., and Isogai, A. (2001). Direct ligand-receptor complex interaction controls *Brassica* self-incompatibility. Nature **413,** 534-538.

Takeuchi, H., and Higashiyama, T. (2011). Attraction of tip-growing pollen tubes by the female gametophyte. Curr Opin Plant Biol **14,** 614-621.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol **28,** 2731-2739.

Tantikanjana, T., Nasrallah, M.E., and Nasrallah, J.B. (2010). Complex networks of self-incompatibility signaling in the Brassicaceae. Curr Opin Plant Biol **13,** 520-526.

Tao, R., Yamane, H., Sassa, H., Mori, H., Gradziel, T.M., Dandekar, A.M., and Sugiura, A. (1997). Identification of stylar RNases associated with gametophytic self-incompatibility in almond (*Prunus dulcis*). Plant Cell Physiol **38,** 304-311.

Thompson, D.M., and Parker, R. (2009). The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in *Saccharomyces cerevisiae*. The Journal of Cell Biology **185,** 43-50.

Thompson, R.D., and Kirch, H.H. (1992). The S locus of flowering plants: when self-rejection is self-interest. Trends Genet **8,** 381-387.

Twell, D., Wing, R., Yamaguchi, J., and McCormick, S. (1989). Isolation and expression of an anther-specific gene from tomato. Molecular and General Genetics MGG **217,** 240-245.

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 Mol Biol **57,** 141-153.

Ushijima, K., Sassa, H., Dandekar, A.M., Gradziel, T.M., Tao, R., and Hirano, H. (2003). Structural and transcriptional analysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. Plant Cell **15,** 771-781.

Ushijima, K., Sassa, H., Tao, R., Yamane, H., Dandekar, A.M., Gradziel, T.M., and Hirano, H. (1998). Cloning and characterization of cDNAs encoding S-RNases from almond (*Prunus dulcis*): primary structural features and sequence diversity of the S-RNases in Rosaceae. Mol Gen Genet **260,** 261-268.

Ushijima, K., Yamane, H., Watari, A., Kakehi, E., Ikeda, K., Hauck, N.R., Iezzoni, A.F., and Tao, R. (2004). The S haplotype-specific F-box protein gene, SFB, is defective in self-compatible haplotypes of *Prunus avium* and *P. mume.* Plant Journal **39,** 573-586.

Varkey, J.P., Muhlrad, P.J., Minniti, A.N., Do, B., and Ward, S. (1995). The *Caenorhabditis elegans* spe-26 gene is necessary to form spermatids and encodes a protein similar to the actin-associated proteins kelch and scruin. Genes Dev **9,** 1074-1086.

Vieira, J., Fonseca, N., and Vieira, C. (2009). RNase-Based Gametophytic Self-Incompatibility Evolution: Questioning the Hypothesis of Multiple Independent Recruitments of the S Pollen Gene. Journal of Molecular Evolution **69,** 32-41.

Vierstra, R.D. (2003). The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. Trends Plant Sci **8,** 135-142.

Vierstra, R.D. (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. Nat Rev Mol Cell Biol **10,** 385-397.

Villar-Pique, A., Sabate, R., Lopera, O., Gibert, J., Torne, J.M., Santos, M., and Ventura, S. (2010). Amyloid-like protein inclusions in tobacco transgenic plants. PLoS One **5,** e13625.

Vissers, A., Dodds, P., Golz, J.F., and Clarke, A.E. (1995). Cloning and nucleotide sequence of the S7-RNase from *Nicotiana alata* Link and Otto. Plant Physiol **108,** 427-428.

Ventura, S., and Villaverde, A. (2006). Protein quality in bacterial inclusion bodies. Trends in Biotechnology **24,** 179-185.

Wang, C.-L., Xu, G.-H., Jiang, X.-T., Chen, G., Wu, J., Wu, H.-Q., and Zhang, S.-L. (2009). S-RNase triggers mitochondrial alteration and DNA degradation in the incompatible pollen tube of *Pyrus pyrifolia* in vitro. The Plant Journal **57,** 220-229.

Wang, C.-L., Wu, J., Xu, G.-H., Gao, Y.-b., Chen, G., Wu, J.-Y., Wu, H.-q., and Zhang, S.-L. (2010). S-RNase disrupts tip-localized reactive oxygen species and induces nuclear DNA degradation in incompatible pollen tubes of *Pyrus pyrifolia*. J Cell Sci **123,** 4301-4309.

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 Mol Biol **53,** 565-580.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet **10,** 57-63.

Wei, D., and Sun, Y. (2010). Small RING Finger Proteins RBX1 and RBX2 of SCF E3 Ubiquitin Ligases: The Role in Cancer and as Cancer Targets. Genes & Cancer **1,** 700-707.

Wheeler, D., and Newbigin, E. (2007). Expression of 10 S-class SLF-like genes in *Nicotiana alata* pollen and its implications for understanding the pollen factor of the S locus. Genetics **177,** 2171-2180.

Wheeler, M.J., Vatovec, S., and Franklin-Tong, V.E. (2010). The pollen S determinant in *Papaver*: comparisons with known plant receptors and protein ligand partners. J Exp Bot **61,** 2015-2025.

Wheeler, M.J., de Graaf, B.H., Hadjiosif, N., Perry, R.M., Poulter, N.S., Osman, K., Vatovec, S., Harper, A., Franklin, F.C., and Franklin-Tong, V.E. (2009). Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. Nature **459,** 992-995.

Whitehouse, H.L.K. (1950). Multiple-allelomorph Incompatibility of Pollen and Style in the Evolution of the Angiosperms. Annals of Botany **14,** 199-216.

Wiersma, P. A., Wu, Z., Zhou, L., Hampson, C., & Kappel, F. (2001). Identification of new self-incompatibility alleles in sweet cherry (*prunus avium L*.) and clarification of incompatibility groups by PCR and sequencing analysis. Theoretical and Applied Genetics, 102(5), 700-708.

Williams, J.S., Der, J.P., dePamphilis, C.W., and Kao, T.H. (2014a). 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.

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

Wissmueller, S., Font, J., Liew, C.W., Cram, E., Schroeder, T., Turner, J., Crossley, M., Mackay, J.P., and Matthews, J.M. (2011). Protein–protein interactions: Analysis of a false positive GST pulldown result. Proteins: Structure, Function, and Bioinformatics **79,** 2365-2371.

Woodward, J.R., Craik, D., Dell, A., Khoo, K.H., Munro, S.L., Clarke, A.E., and Bacic, A. (1992). Structural analysis of the N-linked glycan chains from a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. Glycobiology **2,** 241-250.

Xu, C., Li, M., Wu, J., Guo, H., Li, Q., Zhang, Y., Chai, J., Li, T., and Xue, Y. (2013). Identification of a canonical SCF(SLF) complex involved in S-RNase-based self-incompatibility of *Pyrus* (Rosaceae). Plant Mol Biol **81,** 245-257.

Xue, Y., Carpenter, R., Dickinson, H.G., and Coen, E.S. (1996). Origin of allelic diversity in *Antirrhinum* S locus RNases. Plant Cell **8,** 805-814.

Yamane, H., Ikeda, K., Ushijima, K., Sassa, H., and Tao, R. (2003b). A pollen-expressed gene for a novel protein with an F-box motif that is very tightly linked to a gene for S-RNase in two species of cherry, *Prunus cerasus* and *P. avium*. Plant Cell Physiol **44,** 764-769.

Yamane, H., Ushijima, K., Sassa, H., and Tao, R. (2003a). The use of the S haplotype-specific F-box protein gene, SFB, as a molecular marker for S haplotypes and self-compatibility in Japanese apricot (*Prunus mume*). Theor Appl Genet **107,** 1357-1361. Epub 2003 Aug 1315.

Yamasaki, S., Ivanov, P., Hu, G.-f., and Anderson, P. (2009). Angiogenin cleaves tRNA and promotes stress-induced translational repression. The Journal of Cell Biology **185,** 35-42.

Yasui, Y., Mori, M., Aii, J., Abe, T., Matsumoto, D., Sato, S., Hayashi, Y., Ohnishi, O., and Ota, T. (2012). S-LOCUS EARLY FLOWERING 3 is exclusively present in the genomes of short-styled buckwheat plants that exhibit heteromorphic self-incompatibility. PLoS One **7,** e31264.

Yamane, H., Ushijima, K., Sassa, H., and Tao, R. (2003a). The use of the S haplotype-specific F-box protein gene, SFB, as a molecular marker for S haplotypes and self-compatibility in Japanese apricot (*Prunus mume*). Theor Appl Genet **107,** 1357-1361. Epub 2003 Aug 1315.

Yuan, H., Meng, D., Gu, Z., Li, W., Wang, A., Yang, Q., Zhu, Y., and Li, T. (2014). A novel gene, MdSSK1, as a component of the SCF complex rather than MdSBP1 can mediate the ubiquitination of S-RNase in apple. J Exp Bot **65,** 3121-3131.

Zerbino, D.R., and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research **18,** 821-829.

Zhang, S., Sun, L., and Kragler, F. (2009). The Phloem-Delivered RNA Pool Contains Small Noncoding RNAs and Interferes with Translation. Plant Physiol. **150,** 378-387.

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 J **62,** 52-63.

Zheng, N., Schulman, B.A., Song, L., Miller, J.J., Jeffrey, P.D., Wang, P., Chu, C., Koepp, D.M., Elledge, S.J., Pagano, M., Conaway, R.C., Conaway, J.W., Harper, J.W., and Pavletich, N.P.

(2002). Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. Nature **416,** 703-709.

Zhou, J., Wang, F., Ma, W., Zhang, Y., Han, B., and Xue, Y. (2003). Structural and transcriptional analysis of S locus F-box genes in *Antirrhinum*. Sexual Plant Reproduction *16*, 165-17.